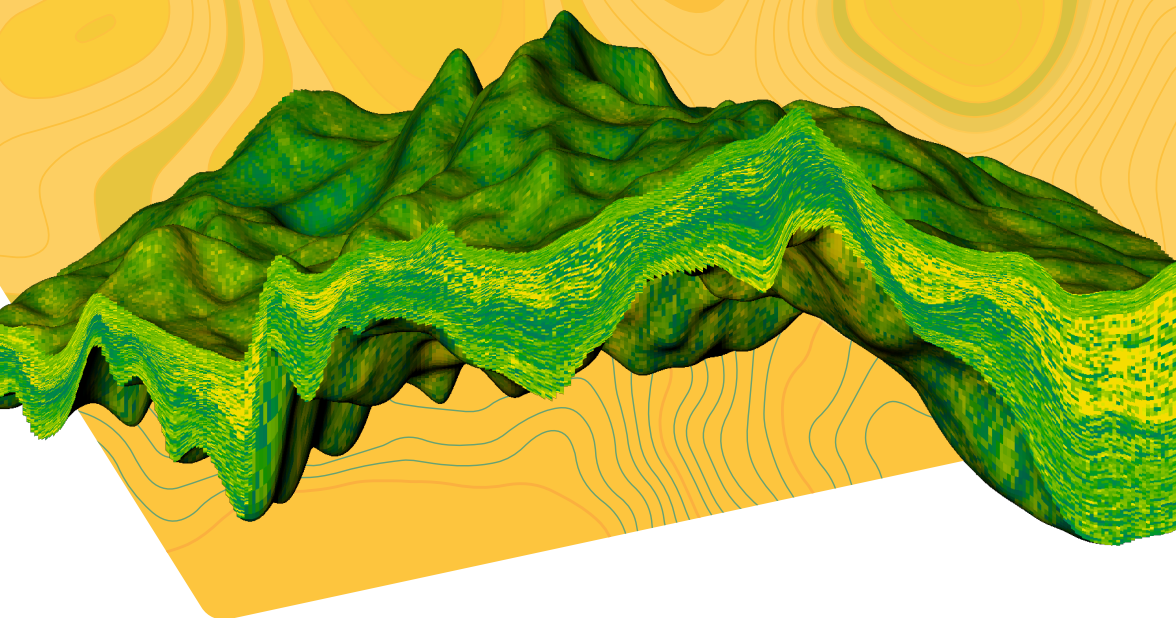




Ensemble Data Assimilation Applied to Geological Reservoir Models

Alexandre A. Emerick

Ensemble Data Assimilation Applied to Geological Reservoir Models

Alexandre A. Emerick

On the cover

Alexandre A. Emerick

Cover design

María Julia Gouffier – Firjan-SENAI

First edition | 2025



© 2025 Alexandre A. Emerick

This work is licensed under a Creative Commons Attribution–NonCommercial 4.0 International License (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0>). You are free to share (copy and redistribute the material in any medium or format) and adapt (transform and build upon the material), provided that you give appropriate credit, include a link to the license, and indicate if changes were made. Commercial use of the material is not permitted. Commercial reproduction of this work, including printed editions for sale, is reserved by the author. For commercial uses or permissions beyond the scope of this license, please contact: emerick@petrobras.com.br.

Calatogação na fonte

Esther Rocha de Souza – CBR-7 6909

E53 Emerick, Alexandre A.

Ensemble data assimilation applied to geological reservoir models / Alexandre A. Emerick – Rio de Janeiro: PETROBRAS, 2025

420 p. : il. color. [recurso eletrônico].

Inclui bibliografias.

ISBN 978-65-88763-29-2.

1. Modelos de reservatórios geológicos. 2. Assimilação de dados. 3. Dados dinâmicos. I. Emerick, Alexandre A. II. Título

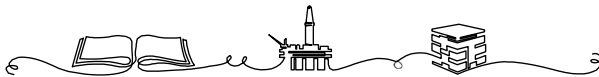
CDD 550.285

After years of business and technological achievements, Petrobras has accumulated invaluable and unique knowledge. Part of this knowledge has been generated, transformed, and transmitted by our employees, who, throughout their professional careers, acquire specific and valuable technical expertise.

Recognizing the contributions and unique skills of our employees is essential. Retaining and disseminating acquired knowledge is equally important. This is the inspiration behind the Textbook Publishing Program (*Programa de Editoração de Livros Didáticos* – PELD), which creates value for new employees, the market, and the academic community. We have published several titles, including both physical books and e-books.

Through PELD, Universidade Petrobras sponsors the entire publishing process, including printing a limited number of copies. This ensures that the published work can be used internally in training programs for new employees and continuous education while also allowing the author to distribute it in the publishing market. We are confident that this initiative enhances employee leadership and strengthens Petrobras as a generator and promoter of knowledge.

We hope that the publication in your hands, a product of PELD, contributes to your learning and continues driving the energy industry forward, delivering even more value to society. Join us on this journey!



Ensemble Data Assimilation Applied to Geological Reservoir Models

offers an in-depth examination of ensemble-based data assimilation methods for numerical modeling of geological reservoirs. Grounded in a Bayesian framework, the book emphasizes the integration of prior geological knowledge with dynamic data to enhance model predictions. A central focus is on iterative ensemble smoothers, particularly the Ensemble Smoother with Multiple Data Assimilation (ES-MDA), which has demonstrated effectiveness in real-world applications. The methods are presented with a balance of theoretical rigor and practical insights, featuring pseudo-codes and case studies to bridge theory and practice. This book is intended for researchers, students, and practitioners involved in reservoir characterization and management in diverse applications, including hydrocarbon production, groundwater hydrology, carbon capture and storage, and geothermal energy.



Alexandre A. Emerick is a Senior Advisor at the Petrobras Research Center (CENPES) in Rio de Janeiro, with over 20 years of experience in applied research in reservoir engineering. His expertise includes reservoir simulation, data assimilation, uncertainty quantification, and optimization. Dr. Emerick has extensive experience in research, software development, training, and the application of ensemble data assimilation to petroleum reservoirs. He has authored numerous peer-reviewed papers on data assimilation and holds a Ph.D. in Petroleum Engineering from The University of Tulsa.

*To Tatiana, my loving wife, whose
unwavering support and belief in me have
been my greatest source of strength.*

*To Rafael and Gabriel, my dear sons,
whose curiosity and joy inspire me every
day.*

This book is for you.

Foreword

Approximately fifteen years ago, Albert Reynolds, Ning Liu, and I published a book on the application of Bayesian inverse theory to history matching. We emphasized the benefits of highly parameterized reservoir models to reduce the effect of model error and better quantify uncertainty. As a consequence of the large number of model variables, minimization of the data mismatch was relatively difficult and typically required the solution of an adjoint system for gradient computation.

The book that Alex Emerick has written has the same general objective of parameter estimation and uncertainty quantification, but the methodology is generally much more computationally efficient and easier to implement. So, while much of the background material in the earlier book is still relevant today, ensemble-Kalman methods have revolutionized the practice of history matching, allowing practitioners to use highly parameterized models without the need for computing gradients. Dr Emerick describes in detail the evolution of the ensemble methods from the early attempts at direct application of the ensemble Kalman filter through the development of ensemble smoothers and iterative ensemble smoothers.

Despite the power of the methodology, there are many reasons that an application of an ensemble-based data assimilation to real-field history matching might fail to properly match data or quantify uncertainty, including omission of important parameters, inappropriate characterization of prior uncertainty, and failure to localize the updates. For practitioners whose only experience is with types of history matching that emphasize the need for parsimony, the importance of getting these aspects right may not be obvious. In this book, Dr Emerick explains thoroughly the “why” and then illustrates the “how” of each critical technique. By “how” I refer to things like the choice of important history-matching variables and the region of localization for updates. Making appropriate choices requires some insight into the physics of the problem, an understanding of the algorithms, and an understanding of the relationship between observations and reservoir model parameters. As the developer of the most popular algorithm for ensemble-based history

matching and his years of experience as a reservoir engineer, Dr Emerick is perhaps uniquely qualified to write a book providing both practical advice and understanding. It could serve well as a textbook on history matching for graduate students, as a reference book for experienced researchers, or as a guide to practical history matching for reservoir engineers.

Bergen, Norway, November 2024

Dean S. Oliver

Preface

I began my journey with ensemble methods during my PhD program at the University of Tulsa in 2008. I was fortunate to encounter these methods at a time when reservoir applications were still in their infancy. It was a period of intensive investigation and testing, during which the strengths and weaknesses of these approaches were gradually being understood. I'm proud to have contributed to the field and even prouder to have helped bridge the gap between academia and operational applications upon returning to Petrobras after completing my studies. At Petrobras, I had the unique opportunity to write ensemble-based codes, conduct initial field applications, deliver numerous training courses, and contribute to the gradual adoption of these methods within the company. After a decade of hands-on experience—working with reservoir engineers, geoscientists, graduate students, data scientists, and data assimilation experts, while also contributing through technical publications—I realized the need to document this journey, which ultimately led to this book.

Writing this book has been one of the most challenging yet rewarding experiences of my career. We often hear the saying, “*If you love what you do, you'll never work a day in your life.*” I'm not so sure about that. While I genuinely enjoyed writing this book, I can certainly attest that it required considerable effort! I've always aspired to create something that would benefit students, researchers, and practitioners. My goal was to strike a balance between theory and practice, ensuring the content is academically rigorous and applicable to real-world scenarios. I'm not sure I fully achieved that, but I certainly gave it my best effort!

It took me a few years to complete this book. I began by writing Chapters 4 and 5, which cover the EnKF and ES-MDA methods. While the EnKF is the most influential ensemble data assimilation technique, ES-MDA was specifically designed in the context of reservoir problems. After completing these chapters, I moved on to Chapter 7, which focuses on localization—a crucial component for ensuring the process works.

After completing the initial chapters, I began working on multiple sections simultaneously, trying to fill gaps and give the book a sense of completeness. Chapter 2, which provides the essential background needed to understand the developments throughout the book, was one of the last chapters I wrote, yet it became fundamental. Chapter 3 departs somewhat from ensemble methods, focusing on more traditional optimization-based data assimilation techniques. Initially, this chapter was intended to be an appendix, solely discussing classical methods based on derivative information. My goal in this discussion was to connect ensemble methods with Gauss-Newton. However, as the chapter expanded, I decided to include additional topics, such as proxy models and reduced parameterizations. Overall, I believe it nicely complements the scope of the book. Chapter 6 discusses other iterative ensemble smoothers. Along with ES-MDA, these methods represent the state of the art in reservoir data assimilation. Chapter 9 was particularly challenging to write. My goal was to discuss practical aspects and provide guidelines for the effective use of ensemble smoothers, something I believe is lacking in the current literature.

Additionally, the book includes four appendices that provide review material, making it easier for readers to follow the main text.

The book focuses on reservoir data assimilation, but I have aimed to keep the discussion broad enough to be useful for readers interested in ensemble methods beyond their application to geological reservoirs. That said, the examples throughout the book are drawn primarily from petroleum reservoirs, as this is my area of expertise. In other words, these are the examples I could provide most effectively based on my background.

This book is deeply influenced by the works of Professors Dean Oliver and Albert Reynolds. I have extensively studied Professor Oliver's publications and regard him as the foremost authority in reservoir data assimilation. His thoughtful review of this book, along with his invaluable recommendations for improvement, significantly refined its content. I'm profoundly honored that he also wrote the foreword. Professor Reynolds, my PhD advisor, played a pivotal role in shaping my understanding of data assimilation while instilling the principles of effective research. Working with him was not only a distinct honor and privilege, but also an immense amount of fun.

I must also acknowledge the significant influence of Dr. Geir Evensen, the inventor of the EnKF, whose contributions have profoundly shaped the field. Additionally, I would like to recognize my colleagues from NORCE—Andreas Stordal, Geir Nævdal, Kristian Fossum, Patrick Raanes, Randi Valestrand, Rolf Lorentzen, Trond Mannseth, Xiaodong Luo, Yan Chen (TotalEnergies), and Remus Hanea (Equinor)—whose contributions and insights have greatly influenced my understanding of ensemble methods.

I would also like to acknowledge the invaluable contributions of my colleagues at Petrobras. Given the many individuals involved, it would be difficult to name everyone without risking unintentional omissions. However, I must express my deep gratitude to four dear friends: Euripedes Barsanulpho, Henrique Cotrim, Vinícius Silva, and Gilson Neto (now at Equinor).

Euripedes, Henrique, and Vinícius were instrumental in introducing ensemble methods at Petrobras. They not only supported me during the early applications but also believed in the potential of these methods at a time when skepticism prevailed. Gilson, one of the brightest individuals I have worked with, made critical contributions, particularly in the integration of 4D seismic data. I sincerely thank Daniel Miranda, Henrique Cotrim, José Roberto Rodrigues, and Marcos Vitor Machado for reviewing the book and granting their recommendation for its publication.

Finally, I must thank Petrobras for investing in my education and granting me permission to publish this work. I'm truly grateful to work at an organization that values both knowledge and its people.

Rio de Janeiro, Brazil, December 2024

Alexandre A. Emerick

Contents

1	Introduction	1
1.1	Geological Reservoirs	1
1.1.1	Reservoir Models	2
1.2	Uncertainty	4
1.3	Forward Problem	5
1.4	Inverse Problem	8
1.4.1	Data Assimilation	10
1.4.2	Geostatistics	11
1.5	Bayesian Viewpoint	13
1.6	Ensembles	14
1.6.1	Ensemble-Based Methods	15
1.7	Book Organization	16
2	Bayesian Formulation of the Data Assimilation Problem	19
2.1	Introduction	19
2.2	Model Likelihood	20
2.2.1	Maximum Likelihood Estimate	21
2.2.2	Data and Model Errors I	22
2.2.3	Linear Case	25
2.3	Maximum a Posteriori	26
2.3.1	Linear Case	28
2.3.2	Nonlinear Case	35
2.3.3	Data and Model Errors II	37
2.4	Data Assimilation as a Sampling Problem	39
2.4.1	Sampling in the Linear Case	40
2.4.2	Sampling in the Nonlinear Case	47
2.4.3	Markov Chain Monte Carlo	48
2.4.4	Randomized Maximum Likelihood	50
2.4.5	Examples	51
2.5	Normalized Objective Function	56

3	Optimization Methods for Data Assimilation	61
3.1	Introduction	61
3.2	Derivative-Based Methods	62
3.2.1	Steepest Descent Method	62
3.2.2	Nonlinear Conjugate Gradient Method	64
3.2.3	Newton's Method	65
3.2.4	Gauss-Newton	66
3.2.5	Levenberg-Marquardt	67
3.2.6	Quasi-Newton	68
3.2.7	Scaling	70
3.2.8	Comments on Line Search	71
3.2.9	Comments on Gradient and Sensitivity Calculation	71
3.3	Stochastic Optimization Methods	73
3.3.1	Evolutionary Algorithms	74
3.3.2	Particle Swarm Optimization	76
3.3.3	Simulated Annealing	77
3.4	Other Derivative-Free Optimization Methods	78
3.5	Methods Based on Proxy Modeling	79
3.5.1	Design of Experiments	79
3.5.2	Standard Proxy Models	80
3.5.3	Other Frequently-Adopted Proxy Models	86
3.6	Sensitivity Analysis	87
3.6.1	Sensitivity Analysis Based on Monte Carlo Simulations	88
3.7	Parametrization	91
3.7.1	Zonation	93
3.7.2	Pilot Point	94
3.7.3	Gradual Deformation	95
3.7.4	Parametrizations Based on PCA	96
3.7.5	Parametrizations Based on Data Sensitivity	99
3.7.6	Parametrizations for Facies	100
3.7.7	Parametrizations for Complex Geological Models	101
3.7.8	Big-Loop Parametrization	101
4	Ensemble Kalman Filter	103
4.1	Introduction	103
4.2	Kalman Filter	104
4.2.1	Kalman Filter as a Minimum Variance Estimator	106
4.2.2	Kalman Filter as a Bayesian Estimator	114
4.2.3	Sequential Data Assimilation	119
4.3	Ensemble Kalman Filter	121
4.3.1	Perturbed Observations	125
4.3.2	Ensemble-Based Covariance	127
4.3.3	Parameters-State Consistency	131
4.3.4	Coupling EnKF with a Reservoir Simulator	132
4.3.5	Filter Divergence	135
4.3.6	Ensemble Square Root Filters	136

5	Ensemble Smoother with Multiple Data Assimilation . . .	139
5.1	Introduction	139
5.2	Ensemble Smoother	140
5.3	Ensemble Smoother with Multiple Data Assimilation	142
5.3.1	MDA for the Linear Case	143
5.3.2	MDA from Bayes' Rule	147
5.3.3	MDA from Regularized Nonlinear Least-Squares	150
5.3.4	Examples	152
5.3.5	How Many Data Assimilations?	156
5.3.6	Deterministic ES-MDA	168
6	Iterative Ensemble Smoothers	175
6.1	Introduction	175
6.2	Ensemble Randomized Maximum Likelihood	175
6.2.1	LM-EnRML	178
6.3	Subspace Iterative Ensemble Smoother	182
7	Sampling Errors and Rank Deficiency	187
7.1	Introduction	187
7.2	Rank Deficiency	188
7.2.1	Example: Porosity and Permeability of a Rock Sample	189
7.3	Spurious Correlations	190
7.3.1	Example: Porosity and Permeability of a Rock Sample	190
7.3.2	Example: Porosity Distribution in a Core Sample	194
7.4	Covariance Inflation	194
7.5	Localization	196
7.5.1	Spurious Correlation	197
7.5.2	Degrees of Freedom	198
7.5.3	Distance-Based Localization	200
7.5.4	Non-Distance Dependent Localization	207
7.5.5	Domain Localization	214
8	Computational Implementation of the Analysis	217
8.1	Introduction	217
8.2	Analysis in Matrix Form	218
8.3	Inversion	219
8.3.1	Pseudoinverse	220
8.3.2	Subspace Inversion	221
8.3.3	Sherman-Morrison-Woodbury Pseudoinverse	224
8.3.4	Truncation of Singular Values	225
8.4	Matrix Multiplications	227
8.4.1	Example: Parallelization of the Analysis	229

9	Practical Aspects and Field Examples	231
9.1	Introduction	231
9.2	Parametrization	231
9.2.1	Realizations versus Scenarios to Describe Uncertainty	232
9.2.2	Integrated Workflows	234
9.2.3	Main Types of Reservoir Parameters	234
9.2.4	Structural Parameters	243
9.3	Observed Data	243
9.3.1	Production Data	243
9.3.2	4D Seismic Data	252
9.3.3	Other Data	255
9.4	Results Evaluation	259
9.4.1	Example: Non-Representative Priors	260
9.4.2	Prior Observation Coverage and Mean Squared Error	261
9.4.3	Plausibility of the Posterior Realizations	265
9.4.4	Data-Mismatch Objective Function	269
9.5	Field Examples	270
9.5.1	Field 1	271
9.5.2	Field 2	275
9.5.3	Field 3	282
9.6	Representative Models	287
A	Elements of Linear Algebra	289
A.1	Preliminaries	289
A.2	Vector Spaces	290
A.3	Vector and Matrix Calculus	294
A.4	Singular Value Decomposition	297
A.5	Pseudoinverse	298
A.5.1	Range, Null Space and Rank	299
A.5.2	Poor Conditioning	300
A.6	Square Root of a Matrix	300
B	Elements of Linear Inverse Problems	303
B.1	Discrete Linear Problems	303
B.2	Ill-Posed Problems	304
B.3	Classification of Linear Inverse Problems	304
B.4	Solutions for Discrete Linear Inverse Problems	305
B.4.1	Least Squares	305
B.4.2	Weighted Least Squares	306
B.4.3	Constrained Least Squares	307
B.4.4	Regularized Least Squares	308
B.4.5	Tikhonov Regularization	309
B.4.6	SVD Solution	310

C	Elements of Probability and Geostatistics	315
C.1	Basic Concepts	315
C.1.1	Random Variables	315
C.1.2	Probability Mass Function	315
C.1.3	Probability Density Function	316
C.1.4	Cumulative Density Function	316
C.1.5	Joint Probability	316
C.1.6	Marginal PDF	317
C.1.7	Conditional PDF	317
C.1.8	The Chain Rule of Conditional Probabilities	317
C.1.9	Expected Value	318
C.1.10	Independent Random Variables	318
C.1.11	Variance	319
C.1.12	Covariance	320
C.1.13	Correlation Coefficient	320
C.1.14	Covariance Matrix	320
C.1.15	Ensemble Estimators	321
C.1.16	Bias	322
C.1.17	Mean Square Error	322
C.1.18	Typical Distributions	323
C.2	Likelihood	329
C.2.1	Maximum Likelihood Estimation	329
C.3	Bayes' Rule	329
C.4	Geostatistics	331
C.4.1	Stationarity	331
C.4.2	Transformation of Variables	331
C.4.3	Variogram	333
C.4.4	Kriging	334
C.4.5	Geostatistical Simulation	335
C.4.6	Geostatistical Simulation of Facies	335
C.4.7	Beyond Covariance-Based Models	336
D	Brief Literature Review	339
D.1	Introduction	339
D.2	EnKF and Its Variants	339
D.3	From EnKF to Iterative ES	341
D.4	Localization	343
D.5	Model Errors	345
D.6	Facies	346
D.7	4D Seismic Data Assimilation	347
D.8	Data-Space Inversion	348
D.9	Ensemble Kalman Inversion	349
D.10	Ensemble-Based Optimization	349
D.11	Ensemble-Based Methods in Multimodal Distributions	350
D.12	Reservoir Applications	351
D.12.1	Petroleum Reservoirs	351

Contents

D.12.2 Groundwater Hydrology	352
D.12.3 Geological Carbon Storage	354
D.12.4 Geothermal Energy	354
D.12.5 Seismic Inversion	355
References	357
Index	397

Symbols

Integers

N_d	Number of data points
N_n	Number of data points at time step t_n
N_m	Number of model parameters
$N_{m,s}$	Number of scalar model parameters
N_e	Number of ensemble members (ensemble size)
N_r	Number of retained singular values
N_s	Number of samples
N_t	Number of time steps

Scalars

c	Compressibility, or normalizing constant, or covariance value
c_1	Coefficient of the wetting phase for capillary pressure calculation
c_2	Coefficient of the non-wetting phase for capillary pressure calculation
e_i	Corey exponent for relative permeability calculation of phase i
e_{p1}	Wetting phase exponent for capillary pressure calculation
e_{p2}	Non-wetting phase exponent for capillary pressure calculation
E_i	LET parameter for relative permeability calculation of phase i
g	Acceleration due to gravity
h	Distance
L	Correlation or critical length
L_i	LET parameter for relative permeability calculation of phase i
t	Time
p	Pressure
p_c	Capillary pressure

q	Source or sink term of mass
r	Localization coefficient
s	Fluid saturation
s_{ni}	Normalized saturation for phase i
$s_{w,con}$	Connate water saturation
$s_{w,cr}$	Critical water saturation
$s_{o,rw}$	Residual oil saturation at oil-water table
T	Temperature parameter in simulated annealing
T_i	LET parameter for relative permeability calculation of phase i
w	Weight
α	ES-MDA inflation coefficient or regularization coefficient in an inverse problem
β	Step size in derivative-based methods, or sum of the inverse inflation coefficients in ES-MDA
γ	Correction coefficient (varies depending on the context)
κ	Permeability
κ_{ri}	Relative permeability of phase i
$\kappa_{ri,max}$	Maximum relative permeability of phase i
λ	Eigenvalue, Levenberg-Marquadt parameter, or Lagrange multipliers
μ	Viscosity
ν	Gradual deformation parameter
ξ	Retained singular value energy
ρ	Fluid density or correlation coefficient
σ	Singular value or standard deviation
σ_e	Data-error standard deviation
σ_{pr}	Prior standard deviation
σ_{post}	Posterior standard deviation
τ	Localization threshold
τ_*	Optimal hard threshold of singular values
ν	BFGS coefficient
ϕ	Porosity
ω	Coefficient to compute the optimal hard threshold of singular values

Vectors

$\mathbf{d}$	Predicted data (dimension N_d)
$\mathbf{d}_{obs}$	Observed data (dimension N_d)
$\mathbf{d}_{true}$	Noiseless data (dimension N_d)
$\mathbf{e}$	Data error (dimension N_d)
$\mathbf{e}_d$	Measurement error (dimension N_d)
$\mathbf{e}_g$	Model error (dimension N_d)

$\mathbf{e}_y$	Model error in the state (dimension N_y)
$\mathbf{m}$	Model parameters (dimension N_m)
$\mathbf{m}_c$	Conditional (posterior) model parameters (dimension N_m)
$\mathbf{m}_{ml}$	Maximum likelihood estimate (dimension N_m)
$\mathbf{m}_{map}$	MAP estimate (dimension N_m)
$\mathbf{m}_{pr}$	Prior estimate (dimension N_m)
$\mathbf{m}_{true}$	Ground truth for the vector of the model parameters (dimension N_m)
$\mathbf{p}$	Gradient change for BFGS (dimension N_m)
$\mathbf{q}$	Model change for BFGS (dimension N_m)
$\mathbf{s}$	Search direction (derivative-based methods)
$\mathbf{u}$	Model controls, left singular vector, or coordinate position (in Geostatistics)
$\mathbf{u}_h$	Historical (past) controls
$\mathbf{u}^*$	Optimal set of controls
$\mathbf{v}$	Phase velocity or right singular vector
$\mathbf{w}$	Weights for LS-SVR or SIES methods
$\mathbf{x}$	State vector for Kalman filter and EnKF
$\mathbf{y}$	Combined parameter-state vector for Kalman filter and EnKF (dimension N_y)
$\mathbf{z}$	Random normal deviates or model reparametrization vector (dimension N_z)
$\delta\mathbf{d}$	Innovation vector (dimension N_d)
$\delta\mathbf{y}$	State vector error (difference between true and predicted state) (dimension N_y)
$\delta\mathbf{z}$	Random normal deviates with reduced variance
$\boldsymbol{\eta}$	Data mismatch tolerances (dimension N_d)
$\mathbf{0}$	Zero or null vector (all entries equals to zero)
$\mathbf{1}$	Unity vector (all entries equals to one)

Matrices

$\mathbf{A}$	Centering matrix (dimension $N_e \times N_e$)
$\mathbf{B}$	Hessian matrix (dimension $N_m \times N_m$)
$\mathbf{C}_e$	Data-error covariance matrix (dimension $N_d \times N_d$)
$\mathbf{C}_{e_d}$	Covariance matrix of measurement errors (dimension $N_d \times N_d$)
$\mathbf{C}_{e_g}$	Covariance matrix of model errors (dimension $N_d \times N_d$)
$\mathbf{C}_m$	Prior covariance of model parameters (dimension $N_m \times N_m$)
$\mathbf{C}_{m_c}$	Posterior covariance of model parameters (dimension $N_m \times N_m$)
$\mathbf{C}_{md}$	Covariance matrix between model parameters and predicted data (dimension $N_m \times N_d$)
$\mathbf{C}_{dd}$	Covariance matrix of predicted data (dimension $N_d \times N_d$)

$\mathbf{C}_y^f$	Forecast covariance matrix of the state vector (dimension $N_y \times N_y$)
$\mathbf{C}_y^a$	Analysed covariance matrix of the state vector (dimension $N_y \times N_y$)
$\mathbf{D}$	Ensemble of predicted data (dimension $N_d \times N_e$)
$\mathbf{E}$	Ensemble data errors (dimension $N_d \times N_e$)
$\mathbf{F}$	Linear forward model in Kalman filter (dimension $N_y \times N_y$)
$\mathbf{G}$	Jacobian or sensitivity matrix (dimension $N_d \times N_m$)
$\mathbf{G}_d$	Dimensionless sensitivity matrix (dimension $N_d \times N_m$)
$\mathbf{H}$	Observation matrix in Kalman filter and EnKF (dimension $N_n \times N_y$)
$\mathbf{I}$	Identity matrix
$\mathbf{J}$	Innovation matrix (dimension $N_d \times N_e$)
$\mathbf{K}$	Kalman gain (dimension $N_m \times N_d$ or $N_y \times N_n$)
$\mathbf{K}$	Permeability tensor (dimension 3×3)
$\mathbf{L}$	Lower triangular matrix obtained with the Cholesky decomposition, or Tikhonov matrix
$\mathbf{M}$	Ensemble of model parameters (dimension $N_m \times N_e$)
$\mathbf{O}$	Zero or null matrix (all entries equals to zero)
$\mathbf{P}$	Reparameterization matrix (dimension $N_m \times N_z$)
$\mathbf{Q}$	BFGS matrix (dimension $N_m \times N_m$)
$\mathbf{R}$	Localization matrix
$\mathbf{R}_{\text{md}}$	Localization matrix for the Kalman gain or for the matrix $\mathbf{C}_{\text{md}}$ (dimension $N_m \times N_d$)
$\mathbf{R}_{\text{dd}}$	Localization matrix for $\mathbf{C}_{\text{dd}}$ (dimension $N_d \times N_d$)
$\mathbf{R}_1$	Random diagonal matrix with element drawn from $\mathcal{U}[0, 1]$ for PSO method
$\mathbf{R}_2$	Random diagonal matrix with element drawn from $\mathcal{U}[0, 1]$ for PSO method
$\mathbf{S}$	Diagonal matrix with data-error standard deviations (dimension $N_d \times N_d$)
$\mathbf{T}_R$	Right-square root transform matrix (dimension $N_e \times N_e$)
$\mathbf{T}_L$	Left-square root transform matrix (dimension $N_y \times N_y$)
$\mathbf{U}$	Left singular vectors
$\mathbf{V}$	Right singular vectors
$\mathbf{W}$	Ensemble of weight parameters for SIES method (dimension $N_e \times N_e$)
$\mathbf{X}$	Ensemble of model states (dimension $N_x \times N_e$)
$\mathbf{Y}$	Ensemble of combined parameters-states (dimension $N_y \times N_e$)
$\Delta \mathbf{M}$	Ensemble anomalies of model parameters (dimension $N_m \times N_e$)
$\Delta \mathbf{D}$	Ensemble anomalies of predicted data (dimension $N_d \times N_e$)
$\Delta \mathbf{X}$	Ensemble anomalies of model states (dimension $N_x \times N_e$)
$\Delta \mathbf{Y}$	Ensemble anomalies of combined parameters-states (dimension $N_x \times N_e$)
Σ	Singular values
Λ	Eigenvalue values

Θ	Auxiliary matrix used in subspace inversion (dimension $N_r \times N_r$)
Γ	Matrix with eigenvalues of Θ used in subspace inversion (dimension $N_r \times N_r$)
Ψ	Matrix with eigenvectors of Θ used in subspace inversion (dimension $N_r \times N_r$), or feature matrix for LS-SVR (dimension $N_s \times N_s$)
Φ	Auxiliary matrix for implementation of the analysis (dimension $N_d \times N_r$)
Υ	Auxiliary matrix for implementation of the analysis (dimension $N_r \times N_r$)
Ξ	Auxiliary matrix for implementation of the analysis (dimension $N_r \times N_r$)
Ω	Auxiliary matrix for implementation of the analysis (dimension $N_d \times N_r$)
Ω_1	Auxiliary matrix for implementation of the analysis (dimension $N_r \times N_d$)
Ω_2	Auxiliary matrix for implementation of the analysis (dimension $N_r \times N_d$)
Ω_3	Auxiliary matrix for implementation of the analysis (dimension $N_d \times N_e$)
Ω_4	Auxiliary matrix for implementation of the analysis (dimension $N_e \times N_e$)
Ω_5	Auxiliary matrix for implementation of the analysis (dimension $N_m \times N_d$)
Ω_6	Auxiliary matrix for implementation of the analysis (dimension $N_e \times N_d$)
Ω_7	Auxiliary matrix for implementation of the analysis (dimension $N_e \times N_d$)
Ω_8	Auxiliary matrix for implementation of the analysis (dimension $N_m \times N_e$)

Functions

$g(\cdot)$	Predicted datum from a forward model
$p(\cdot)$	Probability density function
$P(\cdot)$	Cumulative density function
$p^*(\cdot)$	Target probability density function
$q(\cdot)$	Proposal probability density function
$r(\cdot)$	Correlation function for localization
$OF(\cdot)$	Weighted least-squares data mismatch objective function
$\mathcal{B}(\cdot, \cdot)$	Beta function
$\mathcal{C}(\cdot)$	Covariance function
$\mathcal{D}(\cdot)$	Mahalanobis distance
$\mathcal{I}_z(\cdot, \cdot)$	Incomplete beta function

$\mathcal{O}_d(\cdot)$	Data-mismatch or likelihood objective function
$\mathcal{O}_m(\cdot)$	Model-mismatch or prior objective function
$\mathcal{O}(\cdot)$	Total objective function
$\mathcal{O}_r(\cdot)$	RML objective function
$\mathcal{O}_N(\cdot)$	Normalized objective function
$\mathcal{O}_{N,d}(\cdot)$	Normalized data-mismatch objective function
$\mathcal{L}(\cdot)$	Likelihood function
$\Gamma(\cdot)$	Gamma function
$\gamma(\cdot, \cdot)$	Incomplete gamma function
$\mathcal{Q}(\cdot)$	Quadratic function
$\mathcal{F}(\cdot)$	Quantity of interest (proxy modeling)
$\mathcal{Z}(\cdot)$	Kriging model
$\mathcal{K}(\cdot, \cdot)$	Kernel function

Vector functions

$\mathbf{f}(\cdot, \cdot)$	Forward model
$\mathbf{g}(\cdot)$	Predicted data from a forward model (dimension N_d)
$\mathbf{h}(\cdot)$	Nonlinear relationship between state and data (dimension N_d)
$\varphi(\cdot)$	Nonlinear mapping to a feature space
$\mathfrak{L}(\cdot)$	Augmented Lagrangian function

Vector spaces

$\mathbf{V}$	Vector space
$\mathbb{R}^N$	N -dimensional real space
$\mathfrak{N}(\cdot)$	Null space
$\mathfrak{R}(\cdot)$	Range

Statistical operators

$\mathbb{B}[\cdot]$	Bias
$\mathbb{E}[\cdot]$	Expected value
$\mathbb{V}[\cdot]$	Variance
$\mathbb{C}[\cdot]$	Covariance

Probability distributions

$\chi^2(\cdot)$	Chi-squared
$\mathcal{N}(\cdot, \cdot)$	Normal (Gaussian)
$\mathcal{U}(\cdot, \cdot)$	Uniform

Acronyms

AHM	Assisted History Matching
AVO	Amplitude Variation with Offset
BLUE	Best Linear Unbiased Estimate
BFGS	Broyden-Fletcher-Goldfarb-Shanno
CCS	Carbon Capture and Storage
CCUS	Carbon Capture, Utilization, and Storage
CDF	Cumulative Density Function
CMA-ES	Covariance Matrix Adaptation Evolution Strategy
CMC	Canadian Meteorological Centre
CPU	Central Processing Unit
CS	Cosine Similarity
CUDA	Compute Unified Device Architecture
DES-MDA	Deterministic Ensemble Smoother with Multiple Data Assimilation
DFO	Derivative-Free Optimization
DNAPL	Dense Nonaqueous Phase Liquid
DSI	Data-Space inversion
EA	Evolutionary Algorithms
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
EKI	Ensemble Kalman Inversion
EnOpt	Ensemble-Based Optimization
EnRML	Ensemble Randomized Maximum Likelihood
EnSRF	Ensemble Square Root Filters
ERT	Ensemble Reservoir Tool
ES	Ensemble Smoother
ES-MDA	Ensemble Smoother with Multiple Data Assimilation
FOC	Facies Overlap Coefficient
FMU	Fast Model Update
FWI	Full Waveform Inversion

FZI	Flow Zone Indicator
GOR	Gas-Oil Ratio
GN	Gauss-Newton
GA	Genetic Algorithms
GCS	Geological Carbon Storage
GPL	General Public License
GPU	Graphics Processing Unit
GRACE	Gravity Recovery and Climate Experiment
IAGS	Iterative Adaptive Gaussian Mixture Smoother
ILUES	Iterative Local Updating Ensemble Smoother
KF	Kalman Filter
LETKF	Local Ensemble Transform Kalman Filter
LHS	Latin Hypercube Sampling
LM	Levenberg-Marquadt
LM-EnRML	Levenberg-Marquadt Ensemble Randomized Maximum Likelihood
LS-SVR	Least-Squares Support Vector Regression
LVA	Locally Varying Anisotropy
MAD	Median Absolute Deviation
MAP	Maximum a Posteriori
MCMC	Markov Chain Monte Carlo
MDA	Multiple Data Assimilation
MDP	Morozov's Discrepancy Principle
ML	Maximum Likelihood
MSE	Mean Squared Error
NERSC	Nansen Environmental and Remote Sensing Centre
NCG	Nonlinear Conjugate Gradient
NICE	Noise-Informed Covariance Estimation
NMC	Normalized Model Change
NTG	Net-to-Gross Ratio
NV	Normalized Variance
NPV	Net Present Value
NRMS	Normalized Root Mean Squared
NWP	Numerical Weather Prediction
OAS	Orthogonal Array Sampling
OBN	Ocean Bottom Nodes
OC	Observation Coverage
OHT	Optimal Hard Thresholding
OIP	Oil in Place
PCA	Principal Component Analysis
PDF	Probability Density Function
PDG	Permanent Downhole Gauges
PEM	Petroelastic Model
PLT	Production-Logging Tool
PMF	Probability Mass Function
PSO	Particle Swarm Optimization

PTA	Pressure Transient Analysis
RFT	Repeat Formation Testers
RLM-MAC	Regularized Levenberg-Marquardt for Minimum-Average-Cost
RML	Randomized Maximum Likelihood
SA	Sensitivity Analysis
SDA	Sequential Data Assimilation
SGS	Sequential Gaussian Simulation
StoSAG	Stochastic Simplex Approximate Gradient
SIES	Subspace Iterative Ensemble Smoother
SIS	Sequential Indicator Simulation
SVR	Support Vector Regression
SVD	Singular Value Decomposition
TPG	Truncated Plurigaussian
TSVD	Truncated Singular Value Decomposition
WAG	Water-Alternating-Gas

Introduction

Abstract: *This chapter reviews key concepts explored throughout the book to contextualize the reservoir data assimilation problem. It briefly introduces elements such as geological reservoirs, reservoir models, uncertainty, forward and inverse problems, data assimilation, Bayesian formulation, and ensembles. Additionally, this chapter outlines the organization of the book.*

1.1 Geological Reservoirs

Geological reservoirs are subsurface formations capable of storing and transmitting fluids, such as hydrocarbons (oil and gas), water, hydrogen (H_2), carbon dioxide (CO_2), and waste materials from industrial processes. These reservoirs comprise porous and permeable rocks, such as sandstones, limestones, or fractured igneous and metamorphic rocks, allowing for fluid accumulation and movement within their pore spaces and fractures.

Geological reservoirs result from complex natural processes with varied types and characteristics. However, three key components are always present. First, there must be a rock with sufficient porosity and permeability. Porosity refers to the percentage of the rock's volume made up of pore spaces that can store fluids, with higher porosity indicating greater storage capacity. Permeability describes the ability of the rock to transmit fluids through its pore network, with higher permeability allowing fluids to flow more quickly through the rock, making extraction and injection processes more efficient.

The second component is the presence of a trap and seal. Geological reservoirs are associated with structural or stratigraphic traps that prevent fluids from escaping. A seal, typically a layer of impermeable rock such as shale or salt, overlies the formation and ensures the containment of the fluids within the reservoir.

The third component is the fluid itself, which can be extracted for use, injected for permanent disposal, or stored temporarily. The type, composi-

tion, and thermodynamic conditions of the fluid significantly influence the behavior and management of the reservoir.

Geological reservoirs serve several critical functions across various industries, with their primary applications encompassing the extraction and storage of natural resources. The main applications of geological reservoirs include:

1. **Hydrocarbon production:** This is the most known application. Geological reservoirs store and allow the extraction of hydrocarbons like oil and gas.
2. **Natural gas storage:** Geological reservoirs, particularly salt caverns, saline aquifers, and depleted oil and gas fields, can be used for temporary storage of natural gas, helping to balance the supply and demand of energy resources.
3. **Carbon capture and storage (CCS):** Geological reservoirs are increasingly being utilized for the storage of CO_2 to mitigate the impacts of climate change. In CCS processes, CO_2 is captured from industrial sources or directly from the air and injected into deep geological formations for permanent storage.
4. **Geothermal energy production:** Some geological reservoirs contain hot water or steam that can be explored for geothermal energy production, providing a low-carbon alternative to fossil fuels.
5. **Water supply and management:** Aquifers are essential freshwater sources. They provide water for drinking, irrigation, and industrial use.
6. **Underground waste disposal:** Certain geological reservoirs are used to dispose of hazardous and non-hazardous waste from industrial processes, particularly those related to energy production. Deep geological formations provide a secure environment for the long-term containment of waste materials, reducing the risk of surface contamination and environmental impact.
7. **Underground hydrogen storage:** Geological reservoirs present a promising solution for large-scale hydrogen storage, offering a cost-effective alternative to traditional methods such as compressed or liquefied hydrogen storage. Although the technology is still in the developmental phase, ongoing research and advancements continue to enhance its feasibility and potential for widespread adoption.

1.1.1 Reservoir Models

Understanding and managing geological reservoirs are essential for optimizing resource extraction while maintaining economic and environmental sustainability. Effective development and management involve constructing detailed numerical models that provide insights into fluid flow dynamics and simulate reservoir performance under various operational scenarios. These

models are essential for predicting reservoir behavior, making informed decisions, and enhancing overall efficiency and sustainability in resource utilization.

The construction of a reservoir model is a multidisciplinary process that integrates expertise from various fields to create a comprehensive representation of the subsurface. Fig. 1.1 provides a schematic representation of this process, highlighting some of its main components. In this process, geophysicists use seismic data to map the reservoir's geometry and identify key subsurface features, while geologists interpret the depositional environment and characterize the reservoir's structural and stratigraphic framework. Petrophysicists assess rock properties, such as porosity and permeability, using well-log data and rock samples, and study rock-fluid interactions through laboratory experiments. Geologists and geostatisticians work together to integrate diverse data sources, generating plausible distributions of rock properties. Engineers—including those specializing in reservoirs, production, and drilling—provide essential data on fluid dynamics, well performance, and operational constraints. Reservoir engineers are responsible for building reservoir simulation models and assimilating data acquired during field operations to reduce uncertainties and improve model accuracy.

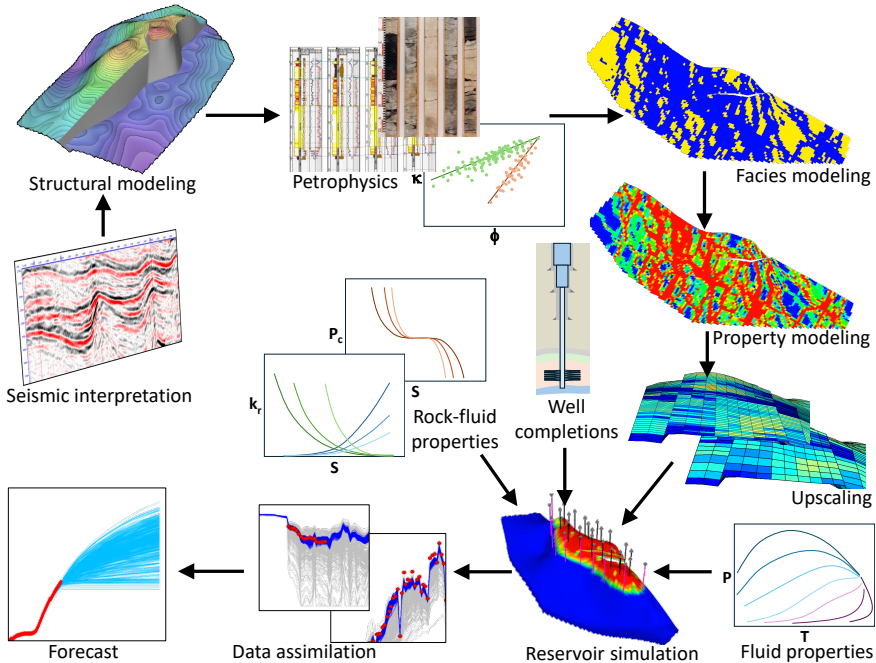


Fig. 1.1: Schematic representation of the reservoir model construction process, highlighting the main steps involved, from geological characterization to simulation and data assimilation.

Reservoir models are inherently complex and uncertain due to several interrelated factors. The heterogeneity of reservoir rocks is a primary challenge, as geological reservoirs consist of various rock types with differing properties such as porosity, permeability, and mineral composition. These properties can vary significantly over short distances, resulting in intricate spatial distributions that are difficult to predict accurately.

Parameter uncertainty further complicates reservoir modeling. Many input parameters, including multiphase relative permeability, capillary pressure, and fluid properties, are not directly measurable and must be estimated, often with substantial uncertainty. This uncertainty is compounded by the reservoirs' dynamic nature, where fluid flow, pressure changes, and chemical interactions evolve over time. Modeling these dynamic processes requires solving complex partial differential equations that account for multiphase flow, heat transfer, geomechanics, and geochemical reactions, each adding layers of complexity and potential sources of uncertainty.

The initial conditions of a reservoir also introduce uncertainty. The initial distribution and composition of fluids, as well as pressure conditions, are often imprecisely known. This uncertainty in initial conditions propagates through the model, affecting predictions of future reservoir behavior. Additionally, the data available for reservoir models are often limited and sparse, typically derived from well logs, core samples, seismic surveys, and dynamic measurements such as flow rates and pressure readings. These data provide only indirect and incomplete information about the reservoir, and the sparse coverage compared to the reservoir's size requires interpolation or extrapolation, introducing further uncertainty.

Measurement errors are another significant source of uncertainty. All measurements used to construct and calibrate reservoir models are subject to some degree of error, stemming from limitations in measurement techniques, instrument precision, and data interpretation. The challenge of integrating data collected at various scales—from microscopic pore-scale observations to large-scale seismic surveys—adds to the complexity of creating a coherent model. Finally, despite advances in computational power, there are still limitations in the resolution and complexity that can be feasibly modeled. Simplifications and assumptions are always necessary, which introduce additional uncertainty.

1.2 Uncertainty

Nearly all problems involving physical systems require decision-making under uncertainty. Uncertainty sources can be classified into three primary types:

1. **Stochasticity in the dynamical system:** Some systems exhibit instability, where small changes in initial conditions lead to divergent solutions

and chaotic dynamics. Consequently, models for these systems have limited predictive capabilities over extended time horizons. A classic example is numerical weather forecast.

2. **Incomplete observability:** The system's outcomes may appear stochastic to the modeler because they cannot observe or accurately determine all the parameters that influence the system's behavior.
3. **Incomplete modeling:** The system's outcomes may also seem stochastic due to the model's inability to capture all the relevant physical processes present in the real system.

Unlike atmospheric models, reservoir models are generally stable, making stochasticity in the dynamical system a less significant source of uncertainty. However, the other two sources—incomplete observability and modeling—are always present.

Ultimately, the goal of a reservoir model is to support informed decision-making in investments and management of natural resources. However, due to inherent uncertainties in model construction, predictions are also uncertain. Therefore, it is essential to quantify this uncertainty in the model's outcomes to manage the potential consequences of different actions.

In this context, uncertainty quantification can be approached from two perspectives: forward propagation of uncertainty and inverse assessment of model uncertainty. Forward propagation involves carrying various sources of uncertainty through the forward model to predict the overall uncertainty in the system's response. In contrast, inverse assessment involves estimating the uncertainty in model input parameters based on measurements from the reservoir. The first perspective—forward modeling—represents the primary goal of reservoir simulation: predicting the future. The second perspective—inverse modeling—is generally more challenging but of great relevance. During inverse modeling, data assimilation occurs, reducing uncertainty in the model parameters and, consequently, improving confidence in the model forecasts.

1.3 Forward Problem

This book centers on data assimilation, particularly emphasizing the inverse modeling perspective of uncertainty quantification. However, before delving into inverse problems, it is helpful first to outline the forward problem addressed in this book.

The forward problem discussed in this book involves a mathematical model governed by partial differential equations with initial and boundary conditions. These equations govern heat and mass transfer through porous media and may also include additional physical processes, such as geomechanical effects and geochemical reactions.

If the mathematical model is sufficiently simple, it may have an analytical solution. For instance, a one-dimensional, two-phase incompressible flow in a homogeneous reservoir, describing an immiscible displacement process such as oil being displaced by water, has a well-known analytical solution derived by Buckley and Leverett [50]. However, for most reservoir problems, analytical solutions are not available, necessitating the use of numerical models, commonly referred to as reservoir simulators.

For example, the mathematical equations that describe an isothermal two-phase flow represent one of the simplest yet common problems in reservoir simulation. In this scenario, fluid flow can be described by the following set of equations:

- Conservation of mass:

$$\nabla \cdot (\rho_i \mathbf{v}_i) + \frac{\partial (\rho_i s_i \phi)}{\partial t} = q_i. \quad (1.1)$$

- Darcy's law:

$$\mathbf{v}_i = -\frac{\kappa_{r,i}}{\mu_i} \mathbf{K} (\nabla p_i - \rho_i g \nabla z). \quad (1.2)$$

- Equation of state:

$$c_i = \frac{1}{\rho_i} \left. \frac{\partial \rho_i}{\partial p_i} \right|_{T=\text{const}}. \quad (1.3)$$

- Saturation:

$$s_1 + s_2 = 1. \quad (1.4)$$

- Capillary pressure:

$$p_2 - p_1 = p_c(s_1). \quad (1.5)$$

In the equations above, the subscript i denotes the phase number, for example, water and oil. Eq. (1.1) represents the conservation of mass, stating that the difference between the mass flux entering and exiting a medium equals the change in the amount accumulated within that medium. The divergence term $\nabla \cdot (\rho_i \mathbf{v}_i)$ represents the net flux within the medium, where ρ_i is the phase density and $\mathbf{v}_i$ is the phase velocity. $\frac{\partial (\rho_i s_i \phi)}{\partial t}$ represents the rate of accumulation of mass, where s_i is the saturation of phase i and ϕ is the porosity of the medium. q_i is a source (or sink) term of mass.

Eq. (1.2) represents Darcy's law, which is a statement of momentum conservation and relates phase velocity to the pressure gradient ∇p_i . Here, $\mathbf{K}$ is the permeability tensor, and $\kappa_{r,i} = \kappa_{r,i}(s_i)$ denotes the relative permeability of phase i , modeled as a function of s_i . $\mu_i = \mu_i(p_i)$ represents the phase viscosity, typically a function of pressure, and $\rho_i g \nabla z$ accounts for gravitational effects, where g is the acceleration due to gravity and ∇z is the depth gradient.

Eq. (1.3) represents an isothermal equation of state that relates fluid density variation to changes in pressure. It provides a straightforward description of phase compressibility, c_i , and assumes no mass transfer between

phases. Eq. (1.4) states that the medium is fully saturated with the two phases, while Eq. (1.5) writes the capillary pressure, p_c , as the difference between the pressure in the non-wetting phase, p_2 , and the pressure in the wetting phase, p_1 . p_c is modeled as a function of the saturation of the wetting phase, s_1 .

Typically, reservoir simulators solve the fluid flow equations using finite difference or finite volume methods. In these approaches, the reservoir is discretized into gridblocks, resulting in a system of nonlinear equations that are solved using an iterative process for the primary variables (dynamical state) of the reservoir at a sequence of times. In the example of two-phase flow described by Eqs. (1.1)–(1.5), the primary variables are pressure and phase saturations. This model has several parameters representing physical properties that must be specified in advance. In the two-phase flow example, ϕ , $\mathbf{K}$, $\kappa_{ri}(s_i)$, $\rho_i(p_i)$, $\mu_i(p_i)$, c_i and $p_c(s_2)$ are model parameters. Following the notation used in Tarantola [428] and Oliver et al. [334], in this book, we reserve $\mathbf{m}$ to denote the vector containing the model parameters. Fig. 1.2 shows an example of a reservoir model of an oilfield in the Campos Basin, illustrating the values of permeability, which represent the model parameters and the corresponding predicted water saturation at a specific time, representing the dynamical state of the system.

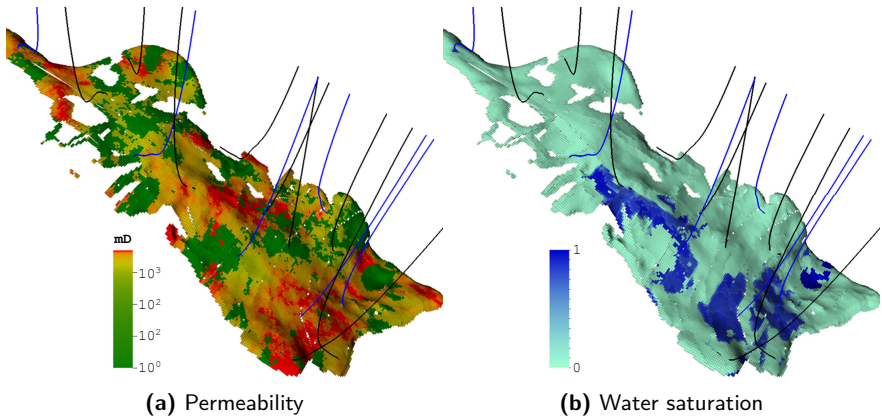


Fig. 1.2: Reservoir simulation model of an oilfield in the Campos Basin. Panel (a) illustrates the permeability distribution, while panel (b) shows the predicted water saturation after 15 years of field operations. The black lines represent oil-producing wells, and the blue lines indicate water injection wells.

The simulation model also requires initial and boundary conditions. In this example, the initial conditions represent the values of p_i and s_i at the beginning of the simulation, typically corresponding to an equilibrium state of the reservoir before operations commence. The boundary conditions define the geometry of the reservoir, often represented as no-flow boundaries, but

they also include possible external sources and sinks of mass, such as aquifers and wells.

Wells play a crucial role in this process, serving as the communication link between the reservoir and the surface. However, they introduce complexities to the overall modeling because the flow within the wellbore differs fundamentally from the flow through porous media. To address this, coupling equations are necessary to model the interaction between these elements accurately. Additionally, wells function as control components in the dynamical system, acting as boundary conditions that must be specified during simulations. Consequently, reservoir simulation predictions are functions of both the model parameters and well controls. We use $\mathbf{g}(\mathbf{m}, \mathbf{u})$ to denote a prediction of a reservoir model, where $\mathbf{u}$ represents a set of controls of the system. For example, $\mathbf{u}$ may include the schedule of fluid injection rates, operational pressure constraints for individual wells, and total fluid processing capacity.

The primary objectives of reservoir simulation are to predict the behavior and performance of a reservoir under various development and operational scenarios. These simulations help optimize recovery strategies and evaluate the economic viability of various operational approaches. In this context, reservoir simulation can be used to predict the optimal set of controls, $\mathbf{u}^*$, while accounting for uncertainties in the parameters, $\mathbf{m}$.

Numerous commercial and open-source reservoir simulators are available, offering a diverse array of formulations and numerical schemes. Prominent open-source examples include MRST [261], OPM-Flow [359], open-DARTS [455], PFLOTRAN-OGS [337], JutulDarcy.jl [311], and GEOS [177], among others. A detailed discussion of the formulations and methods employed by these simulators is beyond the scope of this book. The literature on this topic is extensive and continually evolving; for further reading, see, e.g., [21, 84, 30, 261].

1.4 Inverse Problem

In the forward problem, we are interested in predicting

$$\mathbf{d} = \mathbf{g}(\mathbf{m}, \mathbf{u}) \tag{1.6}$$

for different combinations of $(\mathbf{m}, \mathbf{u})$. In contrast, the inverse problem involves solving

$$\mathbf{g}(\mathbf{m}|\mathbf{u}_h) = \mathbf{d}_{\text{obs}} \tag{1.7}$$

for $\mathbf{m}$, given a set of measurements $\mathbf{d}_{\text{obs}}$. Here, $\mathbf{u}_h$ denotes the historical (past) controls used in the field. Since these controls are fixed during the

inversion¹, we simplify the notation and refer to the forward model as $\mathbf{g}(\mathbf{m})$ from here on.

However, a fundamental flaw exists in attempting to find a solution for (1.7): data contain errors, and models are imperfect representations of reality. Therefore, finding $\mathbf{m}$ such that $\mathbf{g}(\mathbf{m}) = \mathbf{d}_{\text{obs}}$ is exactly satisfied may not be possible nor desirable. The solution must account for data uncertainties and limitations in the forward modeling.

One naive attempt to circumvent this issue would be to search for a model such that the distance between the vectors $\mathbf{g}(\mathbf{m})$ and $\mathbf{d}_{\text{obs}}$ is within some tolerance. For example, one might set an algorithm to look for a model such that

$$\|\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}\| \leq \|\boldsymbol{\eta}\|, \quad (1.8)$$

where $\boldsymbol{\eta}$ represents a vector of data mismatch tolerances. However, several challenges remain. First, although condition (1.8) is less stringent than Eq. (1.7), it may still lead to model overfitting data. Hence, it is necessary to formulate the problem in a more robust manner. Second, since the number of unknown model parameters is usually much larger than the number of observations, multiple or even infinite combinations of model parameters can fit the data within some tolerance. In the context of inverse problems, this constitutes an ill-posed problem². Third, there are always additional sources of information about the model that are not captured by the vector $\mathbf{d}_{\text{obs}}$. This *a priori* information must be combined with the data contained in $\mathbf{d}_{\text{obs}}$ to better constrain the solution of the inverse problem. Eq. (1.8) does not specify a procedure for integrating different sources of information.

The *a priori* information is a critical aspect of inverse problems in reservoir modeling. Reservoir models result from geological modeling efforts that synthesize various sources of information, observations, and hypotheses. For example, geologists often study the depositional environment that formed the reservoir to build a prior expectation about the structures and values of the rock properties.

Fig. 1.3 illustrates a simplified version of this process, where a conceptual model of channelized turbidite deposits is used to classify the expected distribution of rock properties in a reservoir. Fig. 1.3b shows a schematic cross-section with stacked turbidite channels. The yellow areas represent regions rich in sand sediments associated with high reservoir quality, while the green areas indicate regions rich in mud sediments. As one moves away from the central axis of the turbidite channels, a decline in reservoir quality is expected. This conceptual model was used to generate a reservoir model for an oilfield in the Campos Basin, classifying the spatial distribution of

¹ Even past controls can have some level of uncertainty, which might need consideration during the solution of the inverse problem.

² Practical inverse problems are often ill-posed, meaning they typically fail to meet at least one of the following conditions: existence of a solution, uniqueness of the solution, or stability; see Appendix B (Section B.2).

rocks into four categories (facies) [300]. Each category is expected to have different values of effective porosity and permeability.

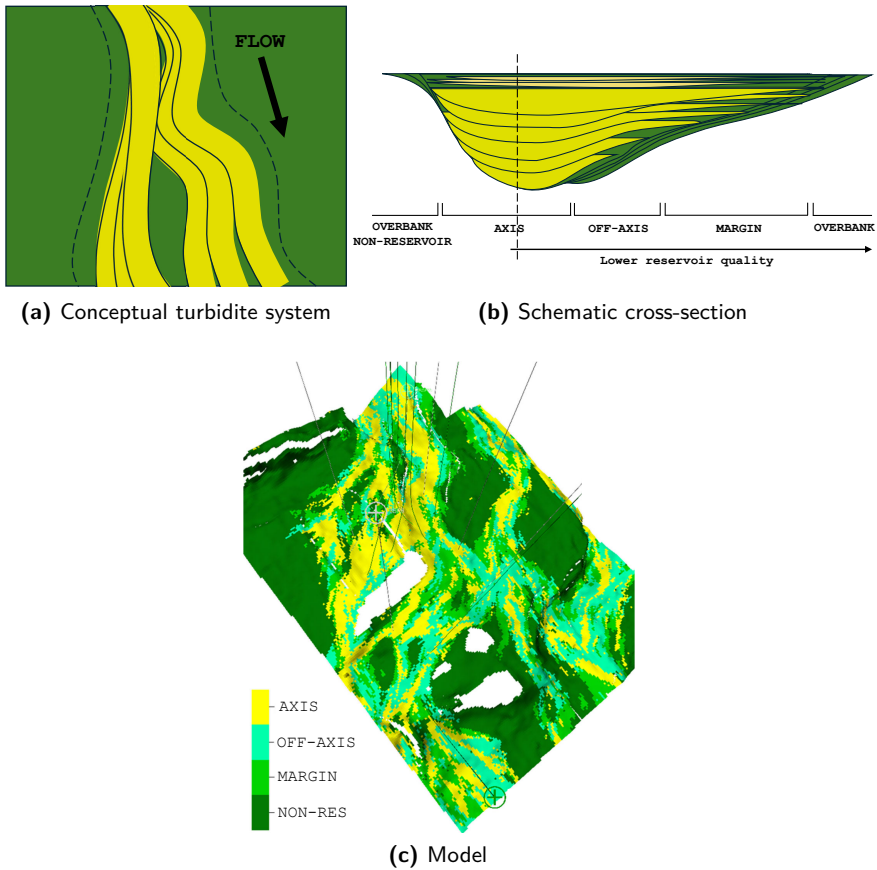


Fig. 1.3: Illustration of prior information used to develop a model for a turbidite reservoir. Panel (a) depicts the conceptual model of amalgamated turbidite channels. Panel (b) presents a schematic cross-section showing the increasing intercalation of sand and mud deposits. Panel (c) displays the model of an actual oilfield, where the rock type distribution follows the classification outlined in (b).

1.4.1 Data Assimilation

Data assimilation involves integrating observed data into numerical models to improve their predictive capabilities. Therefore, it can be viewed as a type of inverse problem. In the context of geological reservoirs, data assimilation

combines various types of data—such as geological, geophysical, petrophysical, and production data—with sophisticated reservoir simulation models. This integration allows for continuous updating and refinement of the models, ensuring they more accurately reflect the system’s current behavior.

In the literature, data assimilation is known by many names, including data integration, model calibration, parameter or state estimation, and history matching. The term “history matching” is particularly prevalent in the petroleum and groundwater hydrology fields. However, this book adopts the term “reservoir data assimilation.”

This book focuses on the assimilation of dynamic data collected during reservoir operations through drilled wells or geophysical techniques such as seismic and electromagnetic methods. These data include flow rate measurements, which indicate the volume of fluid produced or injected; pressure measurements, which reflect reservoir pressure distribution and temporal variations; rock property measurements from well logging and core sampling; fluid composition data obtained from downhole sampling tools, production logging instruments, and surface analysis techniques; temperature measurements from downhole sensors, providing insights into water or gas influx and geothermal gradients; tracer injection data, which track injected tracers to understand fluid movement within the reservoir; and 4D (time-lapse) seismic data, which capture changes in subsurface properties over time.

Fig. 1.4 illustrates some of these data alongside model predictions for an oilfield in the Campos Basin, highlighting discrepancies between observed data and model outputs. The objective of data assimilation is to minimize these discrepancies while accounting for data noise and model limitations.

1.4.2 Geostatistics

At this point, it is convenient to introduce a fundamental element of reservoir modeling: geostatistics. Geostatistics is a specialized branch of statistics that focuses on the analysis and interpretation of spatially or spatiotemporally correlated data. It has become a crucial tool in various fields for predicting the spatial distribution of natural resources and environmental variables.

Originally, geostatistics was closely associated with interpolation methods, particularly with the development of kriging. However, modern geostatistics extends far beyond simple interpolation problems. It has revolutionized the understanding and modeling of geological reservoirs and their uncertainties. This transformation has been driven by the development of advanced algorithms for estimating and simulating reservoir models. Instead of relying on a single “best” estimate, geologists use geostatistics to generate multiple model realizations, embracing a much greater degree of heterogeneity representation and assessment of uncertainty.

Geostatistics is closely tied to inverse problems and can be considered a form of data assimilation, as it seeks to infer the underlying properties of

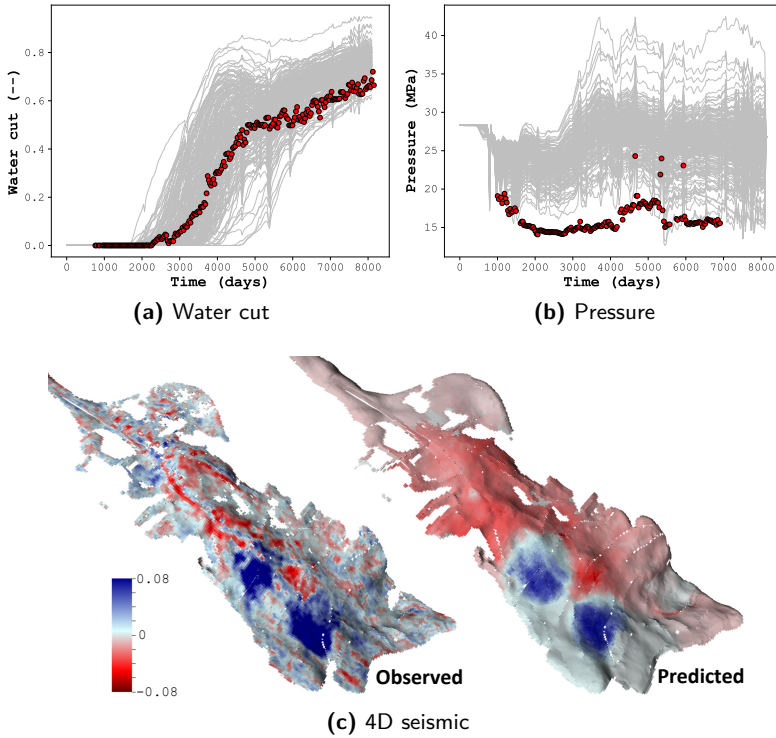


Fig. 1.4: Examples of actual data and predictions from a model of an oilfield in the Campos Basin. Panels (a) and (b) show the observed values of water cut and pressure (red dots), while the gray lines represent predictions from 200 different versions (realizations) of the model. Panel (c) shows the observed and predicted 4D seismic data for this field. In this image, blue colors indicate a relative increase in impedance in the reservoir, typically associated with the injection of water, which is denser than the original oil. The red regions indicate a decrease in impedance, generally associated with an increase in pore pressure.

a system from available data. In practice, however, geostatistics and data assimilation are often treated as complementary components of the process. This distinction partly arises because, in data assimilation, the forward simulations that connect model parameters to data are often nonlinear and computationally intensive. In contrast, geostatistics typically involves data linearly related to model parameters.

Fig. 1.5 illustrates this concept. In this example, porosity measurements from well logs (Fig. 1.5a) and an image of acoustic impedance from seismic data (Fig. 1.5b) were used in a geostatistical method to estimate the porosity distribution for a petroleum reservoir in the Campos Basin (Fig. 1.5c). Here, both data sources—well logs and seismic data—are assumed to be linearly related to the property of interest, which is porosity.

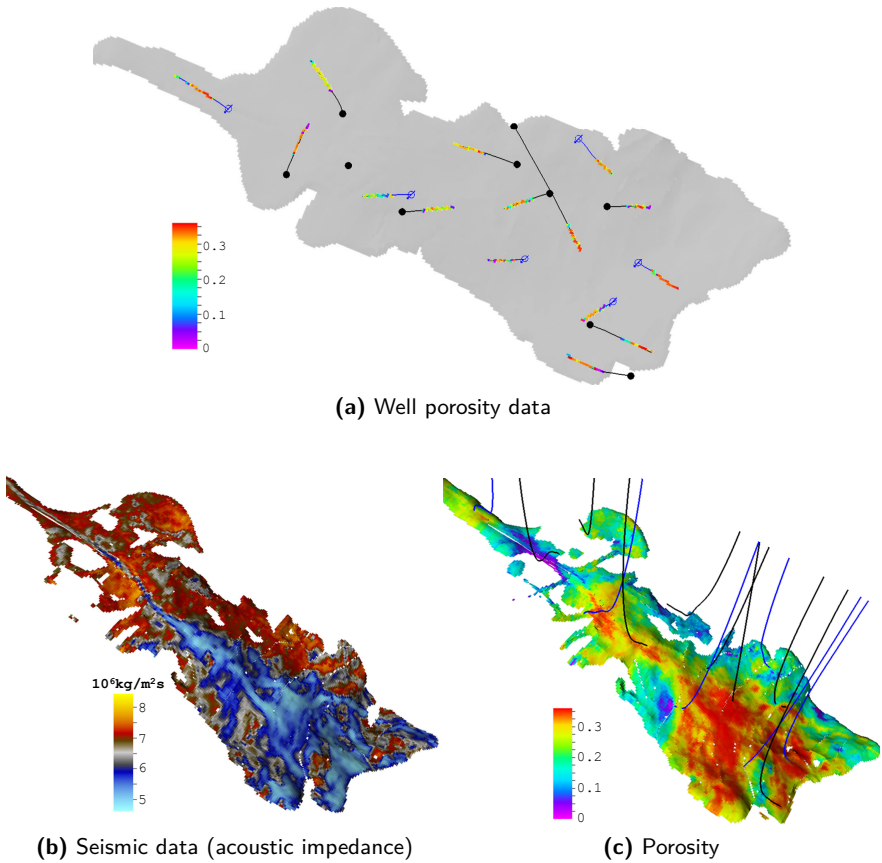


Fig. 1.5: Porosity model of an oilfield in the Campos Basin. Panel (a) illustrates the porosity data at well locations, (b) shows a 3D seismic image (acoustic impedance) used as correlated data to build the porosity model presented in (c).

Geostatistics, however, has specialized in generating complex (non-Gaussian) distributions of model properties by integrating multiple data sources with prior information. In reservoir modeling, and particularly in the data assimilation problems discussed in this book, geostatistics is almost always employed to construct the prior model realizations.

1.5 Bayesian Viewpoint

A practical consequence of inverse problems in reservoir modeling is that we often have multiple solutions consistent with the data, requiring us to

incorporate prior information. Therefore, it is more meaningful to describe the solution in probabilistic terms.

Probability theory has its roots in the analysis of the frequency of events, a concept most intuitive when applied to repeatable events. When we say that an event has a probability P of occurring, it means that if we were to repeat the event an infinite number of times, P would represent the proportion of occurrences. For example, the probability of rolling an odd number with a single die is 50%. With enough rolls, the frequency of odd numbers will eventually converge to 50%.

However, the frequency interpretation of probability does not apply directly to non-repeatable events. For instance, if a doctor says that a patient has a 90% chance of having a particular disease, he or she is not dealing with an infinite number of identical patients with the same symptoms but different underlying conditions. In this case, probability represents a degree of belief rather than a frequency of occurrence.

This use of probability to express relative levels of uncertainty is known as Bayesian probability. An interesting (and sometimes controversial³) consequence is that we assign probabilities to events that are not inherently random. For example, we use probabilities to describe the distribution of rock properties in a reservoir, even though the physical laws governing these properties are deterministic. The reservoir, a result of complex geological processes, is modeled as a random event drawn from a probability distribution.

Bayesian probability theory provides mathematical tools for integrating *a priori* information with observed data. In this framework, the solution to an inverse problem is represented by the *a posteriori* probability distribution⁴.

1.6 Ensembles

Bayesian probability is one of the foundations for the methods discussed in this book. Bayes' rule describes how to update probability density functions (PDFs) in light of new data. However, the actual PDFs are of little practical interest in reservoir modeling. First, these PDFs can be extremely high-

³ Sharon McGrayne's book, *The Theory that Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy* [299], offers a compelling account of the evolution of Bayesian statistics, highlighting its tumultuous journey before ultimately gaining widespread acceptance.

⁴ It is important to note that the terms *a priori* and *a posteriori* should not be interpreted literally as "before" and "after." They do not imply any temporal causality. Instead, these terms refer to the state of understanding of the system. The prior represents our initial knowledge or assumptions about the structure of the model parameters we want to estimate. The combination of this prior knowledge with field measurements results in the posterior.

dimensional, with the number of model parameters ranging from 10^0 to 10^7 , while the number of observations typically ranges from 10^0 to 10^5 . Second, practitioners are more interested in generating alternative models to evaluate the reservoir performance under various operational conditions.

This is where the ensemble approach comes into play. The ensemble represents the Monte Carlo component of the process and complements the solution in two key ways. First, instead of representing probability distributions directly, we use samples. This means that the solution to the inverse problem is a collection of samples—an ensemble of models—that are consistent with both the data and prior information. Second, ensemble methods provide the mechanisms to update these models, with the goal of “moving” prior samples toward regions of high posterior probability.

A significant complication in this process is that executing the forward model can be computationally expensive and time-consuming, limiting the number of samples that can be evaluated. As a practical consequence, naive trial-and-error strategies or even rigorous sampling methods like Markov chain Monte Carlo can become computationally prohibitive despite the impressive advances in computational power over the last decades. This means we are constrained to approximate solutions, as we are particularly interested in methods that scale well with the number of parameters and data points.

1.6.1 Ensemble-Based Methods

Ensemble-based methods encompass a range of data assimilation techniques grounded in Monte Carlo formulations of the Kalman filter [228, 229]. The ensemble Kalman filter (EnKF), introduced by Evensen [136], was the pioneering approach in this field. EnKF and its numerous variations have achieved notable success across various domains, including oceanography [34], atmospheric modeling [464], and numerical weather prediction [201, 338].

The introduction of EnKF as a reservoir data assimilation method by Nævdal et al. [313] sparked significant research activity, leading to the publication of numerous studies in subsequent years; see, e.g., [3, 331] and references therein. In its standard form, EnKF uses an ensemble of states of the dynamical system to represent the mean and covariance, which are updated sequentially over time.

While the sequential data assimilation scheme is a key advantage in oceanography and weather prediction applications, it presents challenges when applied to reservoir data assimilation. Specifically, frequent restarts of the reservoir simulations are required, substantially increasing the overall computational cost of the process.

An alternative approach is to apply the EnKF equations to update the entire history of observations simultaneously, a process known as the ensemble

smoother (ES) [448, 400]. The batch update scheme of ES makes the method more appealing for reservoir applications, though it requires iterative strategies to ensure a reasonable match of the observations. This fact led to the development of several iterative forms of ES [78, 79, 132, 413, 279, 354]. Currently, iterative forms of ES are among the most powerful and successful reservoir data assimilation methods available. The number of field applications has grown significantly, and these methods are now in operational use in the industry.

This book focuses on iterative ensemble smoothers, with particular emphasis on the ensemble smoother with multiple data assimilation (ES-MDA) [132], which stands out as one of the most popular in the reservoir data assimilation literature.

1.7 Book Organization

The book is structured into nine chapters and four appendices, designed to provide a comprehensive understanding of ensemble data assimilation techniques and their applications to geological reservoir models. The methods are presented looking for a balance between mathematical formulation and practical applicability. Each method is accompanied by “pseudo-codes,” which are not intended for direct computational implementation but rather offer a concise overview of the procedural steps involved. Additionally, the book includes practical examples of reservoir data assimilation, primarily drawn from petroleum reservoirs, the author’s area of expertise.

The book is organized as follows: Chapter 2 delves into the Bayesian formulation of the data assimilation problem, offering a statistical framework for updating reservoir models using observational data. Chapter 3 covers optimization methods employed in reservoir data assimilation and includes a discussion of parametrization strategies commonly used in this field. Chapter 4 focuses on the EnKF method, beginning with the derivation of the standard Kalman filter, followed by the introduction of the ensemble component. This chapter also covers topics such as sequential data assimilation, ensemble covariances, parameter-state consistency, and square root formulations. Chapter 5 introduces the ES and ES-MDA methods, emphasizing their advantages in reservoir modeling. Chapter 6 reviews two widely used iterative versions of the ES: ensemble randomized maximum likelihood (EnRML) and subspace iterative ensemble smoother (SIES). Chapter 7 addresses the challenges posed by limited-sized ensembles, namely sampling errors and limited degrees of freedom, and explores strategies to mitigate their negative effects, with a focus on the method known as localization. Chapter 8 provides an overview of the computational implementation of ES-MDA, focusing on the matrix operations essential for efficiently performing the analysis step. Chapter 9 discusses practical aspects and field examples, showcasing real-

world applications of ensemble data assimilation. The book is supplemented by four appendices: A *Elements of Linear Algebra*, B *Elements of Linear Inverse Problems*, C *Elements of Probability and Geostatistics*, and D *Brief Literature Review*. The first three appendices review essential concepts for understanding the methods discussed in the book, while the last appendix offers a brief literature review, highlighting influential papers and current applications of ensemble methods in reservoir modeling.

This page is intentionally left blank. (Paradox alert!)

Bayesian Formulation of the Data Assimilation Problem

Abstract: *This chapter frames data assimilation within the framework of Bayesian inference. It establishes both the likelihood and prior distribution and leverages Bayes' rule to derive the posterior distribution. The analytical solutions for the linear-Gaussian case are presented, forming the foundational basis for subsequent discussions on techniques such as randomized maximum likelihood and ensemble smoother. Additionally, this chapter introduces the primary notation components used throughout the book. Therefore, a solid understanding of the content in this chapter is essential for comprehending the remaining chapters of the book.*

2.1 Introduction

Bayesian inference and Monte Carlo sampling are the cornerstones of modern data assimilation methods. Bayes' rule establishes a mechanism to combine prior information about the reservoir with data, while Monte Carlo sampling provides a strategy for practical applications.

This chapter reviews fundamental elements of Bayesian inference, presenting both likelihood and prior distributions and leveraging Bayes' rule to derive the posterior distribution. It also discusses analytical solutions for the linear-Gaussian case, providing the foundational basis for subsequent techniques such as randomized maximum likelihood and ensemble smoother.

While there are excellent books on this subject, particularly the seminal works of Tarantola [428] and Oliver et al. [334], the objective here is to summarize the indispensable elements necessary for the development of the

remaining chapters of this book. Additionally, this chapter introduces the primary notation used throughout the book.

2.2 Model Likelihood

Let $\mathbf{m} \in \mathbb{R}^{N_m}$ be the random vector of model parameters, which means that $\mathbf{m}$ contains a stochastic realization of the uncertain parameters required to construct a plausible representation of the reservoir of interest. Let $\mathbf{g}(\mathbf{m}) \in \mathbb{R}^{N_d}$ to denote the predicted data obtained running the forward model $\mathbf{g}(\cdot)$ given the realization $\mathbf{m}$. Finally, let $\mathbf{d}_{\text{obs}} \in \mathbb{R}^{N_d}$ be the vector of observed data, which is assumed to be corrupted with an unknown additive random error such that

$$\mathbf{d}_{\text{obs}} = \mathbf{d}_{\text{true}} + \mathbf{e}_d, \quad (2.1)$$

where $\mathbf{d}_{\text{true}}$ represents the true (noiseless) data, and $\mathbf{e}_d$ denotes the data-error vector associated with measurement errors.

For now, assume that the forward model is perfect, which means that if we were able to provide the ground truth for the vector of the model parameters¹, it would return the observation true values

$$\mathbf{g}(\mathbf{m}_{\text{true}}) = \mathbf{d}_{\text{true}}. \quad (2.2)$$

From (2.1) and (2.2), we have

$$\mathbf{d}_{\text{obs}} = \mathbf{g}(\mathbf{m}_{\text{true}}) + \mathbf{e}_d. \quad (2.3)$$

Note that because $\mathbf{e}_d$ is a random vector, $\mathbf{d}_{\text{obs}}$ is also a random vector. Assume that $\mathbf{e}_d$ is a sample from a multivariate Gaussian (normal) distribution with zero mean and covariance $\mathbf{C}_{\mathbf{e}_d}$, which is represented by the notation $\mathbf{e}_d \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{e}_d})$. In this case, the probability density function (PDF) of $\mathbf{e}_d$ has the form

$$p(\mathbf{e}_d) = \frac{1}{[(2\pi)^{N_d} \det \mathbf{C}_{\mathbf{e}_d}]^{1/2}} \exp \left\{ -\frac{1}{2} \mathbf{e}_d^\top \mathbf{C}_{\mathbf{e}_d}^{-1} \mathbf{e}_d \right\}. \quad (2.4)$$

The expectation of $\mathbf{d}_{\text{obs}}$ is

$$\mathbb{E}[\mathbf{d}_{\text{obs}}] = \mathbb{E}[\mathbf{g}(\mathbf{m}_{\text{true}}) + \mathbf{e}_d] = \mathbf{g}(\mathbf{m}_{\text{true}}), \quad (2.5)$$

and the covariance is

¹ Note that $\mathbf{m}_{\text{true}}$ is an abstraction; the actual reservoir is a physical entity, not a model. Hence, it does not have parameters.

$$\begin{aligned}
 \mathbb{C}[\mathbf{d}_{\text{obs}}] &= \mathbb{E} \left[(\mathbf{d}_{\text{obs}} - \mathbb{E}[\mathbf{d}_{\text{obs}}]) (\mathbf{d}_{\text{obs}} - \mathbb{E}[\mathbf{d}_{\text{obs}}])^\top \right] \\
 &= \mathbb{E} \left[(\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}_{\text{true}})) (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}_{\text{true}}))^\top \right] \\
 &= \mathbb{E} [\mathbf{e}_d \mathbf{e}_d^\top] \\
 &= \mathbb{C}[\mathbf{e}_d] \\
 &= \mathbf{C}_{\mathbf{e}_d}.
 \end{aligned} \tag{2.6}$$

Thus, $\mathbf{d}_{\text{obs}} \sim \mathcal{N}(\mathbf{g}(\mathbf{m}_{\text{true}}), \mathbf{C}_{\mathbf{e}_d})$ and the PDF for $\mathbf{d}_{\text{obs}}$ given $\mathbf{m}_{\text{true}}$ is

$$\begin{aligned}
 p(\mathbf{d}_{\text{obs}} | \mathbf{m}_{\text{true}}) &= \frac{1}{[(2\pi)^{N_d} \det \mathbf{C}_{\mathbf{e}_d}]^{1/2}} \times \\
 &\exp \left\{ -\frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}_{\text{true}}))^\top \mathbf{C}_{\mathbf{e}_d}^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}_{\text{true}})) \right\}.
 \end{aligned} \tag{2.7}$$

Instead of $p(\mathbf{d}_{\text{obs}} | \mathbf{m}_{\text{true}})$, we can speculate about the likelihood of a model $\mathbf{m}$ for a given set of actual measurements, $\mathbf{d}_{\text{obs}}$. In this case, we write

$$\begin{aligned}
 \mathcal{L}(\mathbf{m}) &= \mathcal{L}(\mathbf{m} | \mathbf{d}_{\text{obs}}) \equiv p(\mathbf{d}_{\text{obs}} | \mathbf{m}) \\
 &= \frac{1}{[(2\pi)^{N_d} \det \mathbf{C}_{\mathbf{e}_d}]^{1/2}} \times \\
 &\exp \left\{ -\frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_{\mathbf{e}_d}^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})) \right\}.
 \end{aligned} \tag{2.8}$$

Eq. (2.8) represents the likelihood function² of $\mathbf{m}$ given $\mathbf{d}_{\text{obs}}$. This function measures how well a model explains the data [428].

2.2.1 Maximum Likelihood Estimate

The vector $\mathbf{m}$ that maximizes $\mathcal{L}(\mathbf{m})$ is called the maximum likelihood (ML) estimate. Note that maximizing $\mathcal{L}(\mathbf{m})$ defined in Eq. (2.8) is equivalent to solving the following minimization problem

$$\mathbf{m}_{\text{ml}} = \arg \min_{\mathbf{m}} \mathcal{O}_d(\mathbf{m}), \tag{2.9}$$

where

$$\mathcal{O}_d(\mathbf{m}) = \frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_{\mathbf{e}_d}^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})). \tag{2.10}$$

We refer to $\mathcal{O}_d(\mathbf{m})$ as the data-mismatch or likelihood objective function.

² Appendix C provides a brief review of fundamental elements of probability and statistics used in this book.

Remark 2.1: Objective Function

In the petroleum history matching literature a very common assumption for production data is that the measurement errors are uncorrelated in time and space. In this case, the matrix $\mathbf{C}_{\mathbf{e}_d}$ becomes diagonal

$$\mathbf{C}_{\mathbf{e}_d} = \begin{bmatrix} \sigma_{e,1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{e,2}^2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_{e,N_d}^2 \end{bmatrix}, \quad (2.11)$$

where $\sigma_{e,i}^2$ is the variance of the measurement error of the i th datum. In case of uncorrelated measurement errors, $\mathcal{O}_d(\mathbf{m})$ reduces to

$$\mathcal{O}_d(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^{N_d} \frac{(d_{\text{obs},i} - g_i(\mathbf{m}))^2}{\sigma_{e,i}^2}. \quad (2.12)$$

In commercial assisted history matching (AHM) packages, the objective function is typically defined as a sum of weighted squares, i.e.,

$$OF(\mathbf{m}) = \sum_{i=1}^{N_d} w_i (g_i(\mathbf{m}) - d_{\text{obs},i})^2, \quad (2.13)$$

where w_i is the weight of the i th datum. Comparing these objective functions, we conclude that in order to obtain the ML estimate, we should use

$$w_i = \frac{1}{2\sigma_{e,i}^2}, \quad (2.14)$$

which means that the weight of each datum in $OF(\mathbf{m})$ should reflect the confidence level we have in the particular datum.

2.2.2 Data and Model Errors I

The vector $\mathbf{e}_d$ was defined as the random vector representing the noise due to measurement errors present in the data. A critical assumption made to derive the likelihood function in Eq. (2.8) is $\mathbf{e}_d \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{e}_d})$, which implies the following:

- There are no biases in the observed data vector, meaning there are no systematic errors in the measurements.
- The noise follows a Gaussian distribution with a known covariance.
- The forward model is perfect.

Clearly, these assumptions do not hold in practice. In fact, the uncertainty attributed to the measurements strongly impacts the data assimilation as they define the accuracy required for data matching. Unfortunately, how to appraise data uncertainty is not always evident because more than just random noise may contaminate the observations. Often, there are systematic errors in the data acquisition, which are challenging to characterize and frequently unknown to practitioners. A poor representation of the uncertainty in the measurements deteriorates the data assimilation results. Underestimation leads to overly strong updates in model parameters, while overestimation hinders the utilization of all available information in the data.

Moreover, models inherently contain imperfections and deviations from reality; thus, model errors are a persistent factor in practical applications. Perhaps one of the first works to formally incorporate model errors within the Bayesian uncertainty quantification framework was presented by Kennedy and O’Hagan [235]. They define model errors as “*discrepancies between model predictions and the actual physical processes, even in the absence of uncertainty in the model input parameters.*” Model errors also appear in the literature with other names such as model inadequacy, model discrepancy, model bias, or structural uncertainty [49]. These errors typically arise from simplifications made in the formulation of the physical process of interest or numerical approximations (discretization errors) resulting from the solution of the differential equations. Another important source of model errors is the fact that it is almost inevitable that some parameters that have an influence on the predicted data are neglected during data assimilation. In this situation, the parameters used in the data assimilation will have to compensate for the missing ones. As a result, we may observe some of these parameter values outside of the expected range.

In practice, however, estimating the effects of model errors is difficult. Model errors and observation biases typically have similar consequences in data assimilation, making it challenging to separate these effects without additional information.

Remark 2.2: Model Errors and Observation Biases in Reservoir Data Assimilation

Examples of sources of model errors in reservoir data assimilation include:

- Under-parametrization (missing parameters with a significant effect on predicted data).
- Discretization errors (excessive coarseness).
- Well constraint specification (control the simulation forcing to honor noisy rate data).
- Fluid modeling (use of black-oil formulation to approximate compositional).
- Simplified well model (neglect long-term loss of performance due to local effects).

Examples of sources of observation biases:

- Allocation of production/injection rates based on infrequent well tests.
- Extrapolation of buildup pressures.
- Incorrect time-to-depth conversion for 4D seismic.

A common remedy for compensating for neglecting model errors and observation bias is inflating the data error covariance. This procedure does not resolve the problem but alleviates its negative consequences. This procedure can be justified by assuming additive model errors, denoted as $\mathbf{e}_g$,

$$\mathbf{g}(\mathbf{m}_{\text{true}}) = \mathbf{d}_{\text{true}} + \mathbf{e}_g, \quad (2.15)$$

in which case the observation vector can be written as

$$\begin{aligned} \mathbf{d}_{\text{obs}} &= \mathbf{d}_{\text{true}} + \mathbf{e}_d \\ &= \mathbf{g}(\mathbf{m}_{\text{true}}) - \mathbf{e}_g + \mathbf{e}_d. \end{aligned} \quad (2.16)$$

Assuming independence between $\mathbf{e}_g$ and $\mathbf{e}_d$ ³ and disregarding model bias leads to

$$\mathbb{E}[\mathbf{d}_{\text{obs}}] = \mathbf{g}(\mathbf{m}_{\text{true}}) \quad (2.17)$$

and

³ Independence means that $\mathbb{E}[\mathbf{e}_g \mathbf{e}_d^\top] = \mathbb{E}[\mathbf{e}_d \mathbf{e}_g^\top] = \mathbf{O}$, with $\mathbf{O}$ denoting the matrix with zeros in all entries.

$$\begin{aligned}
 \mathbb{C}[\mathbf{d}_{\text{obs}}] &= \mathbb{E} \left[(-\mathbf{e}_g + \mathbf{e}_d) (-\mathbf{e}_g + \mathbf{e}_d)^\top \right] \\
 &= \mathbb{E} [\mathbf{e}_g \mathbf{e}_g^\top] + \mathbb{E} [\mathbf{e}_d \mathbf{e}_d^\top] - \mathbb{E} [\mathbf{e}_g \mathbf{e}_d^\top] - \mathbb{E} [\mathbf{e}_d \mathbf{e}_g^\top] \\
 &= \mathbb{C}[\mathbf{e}_g] + \mathbb{C}[\mathbf{e}_d] \\
 &= \mathbf{C}_{\mathbf{e}_g} + \mathbf{C}_{\mathbf{e}_d} \\
 &\equiv \mathbf{C}_{\mathbf{e}}.
 \end{aligned} \tag{2.18}$$

This means that the likelihood expression in Eq. (2.8) remains valid, replacing $\mathbf{C}_{\mathbf{e}_d}$ by a total data-error covariance, $\mathbf{C}_{\mathbf{e}}$, which represents the sum of the covariance matrix of the measurement and model errors. Unlike measurement errors, which can often be assumed to be independent in time and space, model errors are almost always correlated.

2.2.3 Linear Case

Assume that the predicted data are linearly related to $\mathbf{m}$, i.e.,

$$\mathbf{g}(\mathbf{m}) = \mathbf{G}\mathbf{m}, \tag{2.19}$$

where $\mathbf{G}$ is an $N_d \times N_m$ matrix. The entry in the i th row and j th column of $\mathbf{G}$ is given by

$$G_{ij} = \frac{\partial g_i(\mathbf{m})}{\partial m_j}, \tag{2.20}$$

which means that G_{ij} is the sensitivity of the i th datum with respect to the j th model parameter. We refer to the matrix $\mathbf{G}$ as the Jacobian or sensitivity matrix.

For the linear case, the ML estimate can be found by minimizing

$$\mathcal{O}_d(\mathbf{m}) = \frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) \tag{2.21}$$

analytically. This can be achieved by setting the gradient of $\mathcal{O}_d(\mathbf{m})$ to zero⁴ and solving for $\mathbf{m}$,

$$\begin{aligned}
 \mathbf{0} &= \nabla \mathcal{O}_d(\mathbf{m}) = \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) \\
 &= \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs}} - \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G}\mathbf{m},
 \end{aligned} \tag{2.22}$$

which leads to

$$\mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G}\mathbf{m} = \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs}}. \tag{2.23}$$

The matrix $\mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G}$ is at least positive semidefinite, but there is no guarantee that this matrix has an inverse. In case the inverse does not exist, we may have none or an infinite number of solutions. Nevertheless, we can

⁴ See Appendix A, Section A.3, for a review of basic vector and matrix calculus.

pick a “solution” for this problem using some regularization strategy (as discussed in the Appendix B, Section B.4).

2.3 Maximum a Posteriori

The ML estimate focuses solely on maximizing the likelihood function based on the observed data without considering any prior information about the problem. Fortunately, Bayes’ rule provides a way to combine information from the data with prior beliefs about the distribution of model parameters.

Let $p(\mathbf{m}, \mathbf{d}_{\text{obs}})$ denote the joint distribution of $\mathbf{m}$ and $\mathbf{d}_{\text{obs}}$. Recall from Bayes’ rule (Appendix C, Section C.3):

$$\begin{aligned}
 p(\mathbf{m}|\mathbf{d}_{\text{obs}}) &= \frac{p(\mathbf{m}, \mathbf{d}_{\text{obs}})}{p(\mathbf{d}_{\text{obs}})} \\
 &= \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{m}) p(\mathbf{m})}{p(\mathbf{d}_{\text{obs}})} \\
 &= \frac{\mathcal{L}(\mathbf{m}) p(\mathbf{m})}{p(\mathbf{d}_{\text{obs}})} \\
 &= \frac{\mathcal{L}(\mathbf{m}) p(\mathbf{m})}{\int \mathcal{L}(\mathbf{m}) p(\mathbf{m}) d\mathbf{m}}
 \end{aligned} \tag{2.24}$$

or

$$p(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \text{const} \times \mathcal{L}(\mathbf{m}) p(\mathbf{m}). \tag{2.25}$$

Now, let the prior PDF of $\mathbf{m}$ be a multivariate Gaussian, $p(\mathbf{m}) = \mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_{\mathbf{m}})$. In this case, $p(\mathbf{m})$ assumes the form

$$p(\mathbf{m}) = \frac{1}{[(2\pi)^{N_m} \det \mathbf{C}_{\mathbf{m}}]^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}) \right\}, \tag{2.26}$$

where $\mathbf{m}_{\text{pr}}$ is the prior mean and $\mathbf{C}_{\mathbf{m}}$ is the prior covariance of $\mathbf{m}$.

Using the expressions obtained for $\mathcal{L}(\mathbf{m})$ (2.8) and $p(\mathbf{m})$ (2.26) in (2.25) results in

$$\begin{aligned}
 p(\mathbf{m}|\mathbf{d}_{\text{obs}}) &= \text{const} \times \exp \left\{ -\frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})) \right\} \times \\
 &\quad \exp \left\{ -\frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}) \right\} \\
 &= \text{const} \times \exp \left\{ -\frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})) - \right. \\
 &\quad \left. \frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}) \right\} \\
 &= \text{const} \times \exp \{ -\mathcal{O}(\mathbf{m}) \}.
 \end{aligned} \tag{2.27}$$

In this expression, $\mathcal{O}(\mathbf{m})$ is the total objective function defined as

$$\mathcal{O}(\mathbf{m}) \equiv \mathcal{O}_d(\mathbf{m}) + \mathcal{O}_m(\mathbf{m}) \tag{2.28}$$

with

$$\mathcal{O}_d(\mathbf{m}) = \frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})) \tag{2.29}$$

and

$$\mathcal{O}_m(\mathbf{m}) = \frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}). \tag{2.30}$$

The maximum a posteriori (MAP) estimate is defined as the model $\mathbf{m}$ that maximizes $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$. This is equivalent to minimizing $\mathcal{O}(\mathbf{m})$, i.e.,

$$\mathbf{m}_{\text{map}} = \arg \min_{\mathbf{m}} \mathcal{O}(\mathbf{m}). \tag{2.31}$$

Remark 2.3: Objective Function

1. The functional $\mathcal{O}(\mathbf{m})$ defined in Eq. (2.28) includes two components: $\mathcal{O}_d(\mathbf{m})$, representing the data-mismatch part, and $\mathcal{O}_m(\mathbf{m})$, representing the model-mismatch part. Therefore, minimizing $\mathcal{O}(\mathbf{m})$ seeks a compromise between adjusting $\mathbf{m}$ to match the data while maintaining its proximity to $\mathbf{m}_{\text{pr}}$. In essence, it aims for the smallest changes in $\mathbf{m}$ required to assimilate the observations.
2. In reservoir data assimilation problems, typically, the number of model parameters greatly exceeds the number of observations ($N_m \gg N_d$). This makes the problem of minimizing the mismatch between predicted and observed data ill-posed, and some regularization is often needed. Fortunately, $\mathcal{O}(\mathbf{m})$, which was obtained from a direct application of Bayes' rule, already contains a built-in regularization term: $\mathcal{O}_m(\mathbf{m})$ (see Appendix B, Section B.4 for a short discussion on regularization methods).

2.3.1 Linear Case

For the linear case, we have $\mathbf{g}(\mathbf{m}) = \mathbf{G}\mathbf{m}$ and the objective function becomes

$$\begin{aligned} \mathcal{O}(\mathbf{m}) &= \frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) + \\ &\quad \frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}), \end{aligned} \quad (2.32)$$

whose minimum can be computed analytically. Setting the gradient of $\mathcal{O}(\mathbf{m})$ to zero leads to

$$\begin{aligned} \mathbf{0} &= \nabla \mathcal{O}(\mathbf{m}) \\ &= -\mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) + \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}) \\ &= -\mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}} + \mathbf{G}\mathbf{m}_{\text{pr}} - \mathbf{G}\mathbf{m}_{\text{pr}}) + \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}) \\ &= -\mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}}) + (\mathbf{C}_m^{-1} + \mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G}) (\mathbf{m} - \mathbf{m}_{\text{pr}}). \end{aligned} \quad (2.33)$$

Hence,

$$(\mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G} + \mathbf{C}_m^{-1}) (\mathbf{m} - \mathbf{m}_{\text{pr}}) = \mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}}). \quad (2.34)$$

Let

$$\mathbf{B} \equiv \mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G} + \mathbf{C}_m^{-1}, \quad (2.35)$$

then Eq. (2.34) becomes

$$\mathbf{B} (\mathbf{m} - \mathbf{m}_{\text{pr}}) = \mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}}). \quad (2.36)$$

Note that $\mathbf{B}$ is positive definite because $\mathbf{C}_m^{-1}$ is positive definite and $\mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G}$ is at least positive semidefinite. Therefore, we can solve for $\mathbf{m}$. Calling this result as $\mathbf{m}_{\text{map}}$ results in

$$\mathbf{m}_{\text{map}} = \mathbf{m}_{\text{pr}} + \mathbf{B}^{-1} \mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}}). \quad (2.37)$$

The MAP estimate in Eq. (2.37) assumes a Gaussian prior PDF, $p(\mathbf{m})$, and a linear relationship between data and model, $\mathbf{g}(\mathbf{m}) = \mathbf{G}\mathbf{m}$. Under these assumptions, the posterior PDF, $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$, is also Gaussian [428]. To demonstrate this, it suffices to write the second-order Taylor series expansion of $\mathcal{O}(\mathbf{m})$ around $\mathbf{m}_{\text{map}}$:

$$\begin{aligned} \mathcal{O}(\mathbf{m}) &= \mathcal{O}(\mathbf{m}_{\text{map}}) + [\nabla \mathcal{O}(\mathbf{m})]_{\mathbf{m}_{\text{map}}}^\top (\mathbf{m} - \mathbf{m}_{\text{map}}) + \\ &\quad \frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{map}})^\top \nabla \left\{ [\nabla \mathcal{O}(\mathbf{m})]_{\mathbf{m}_{\text{map}}}^\top \right\} (\mathbf{m} - \mathbf{m}_{\text{map}}), \end{aligned} \quad (2.38)$$

and note that

$$\nabla [\mathcal{O}(\mathbf{m})]_{\mathbf{m}_{\text{map}}} = \mathbf{0}. \quad (2.39)$$

The Hessian $\nabla \left\{ [\nabla \mathcal{O}(\mathbf{m})]^\top \right\}$ is given by

$$\begin{aligned} \nabla \left\{ [\nabla \mathcal{O}(\mathbf{m})]^\top \right\} &= \nabla \left\{ [-\mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) + \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}})]^\top \right\} \\ &= \nabla \left\{ -(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top \mathbf{C}_e^{-1} \mathbf{G} + (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_m^{-1} \right\} \\ &= \nabla \left\{ -(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top \mathbf{C}_e^{-1} \mathbf{G} \right\} + \\ &\quad \nabla \left\{ (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_m^{-1} \right\} \\ &= \nabla \left\{ \mathbf{m}^\top \mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G} \right\} + \nabla \left\{ \mathbf{m}^\top \mathbf{C}_m^{-1} \right\} \\ &= \nabla \mathbf{m}^\top (\mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G}) + \nabla (\mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G})^\top \mathbf{m} + \\ &\quad \nabla \mathbf{m}^\top \mathbf{C}_m^{-1} + \nabla (\mathbf{C}_m^{-1})^\top \mathbf{m} \\ &= \mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G} + \mathbf{C}_m^{-1} \\ &= \mathbf{B}. \end{aligned} \quad (2.40)$$

We used the fact that $\mathbf{C}_m$ is symmetric positive definite, which means that $\mathbf{C}_m^{-1} = \mathbf{C}_m^{-\top}$ to obtain Eq. (2.40). This result shows that the Hessian matrix is equal to $\mathbf{B}$ defined in Eq. (2.35), which is constant (independent of $\mathbf{m}$). Hence, there are no higher-order terms in the Taylor series and the quadratic form

$$\mathcal{O}(\mathbf{m}) = \mathcal{O}(\mathbf{m}_{\text{map}}) + \frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{map}})^\top \mathbf{B} (\mathbf{m} - \mathbf{m}_{\text{map}}) \quad (2.41)$$

is exact. Recall that

$$p(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \text{const} \times \exp \{-\mathcal{O}(\mathbf{m})\}. \quad (2.42)$$

Hence

$$\begin{aligned} p(\mathbf{m}|\mathbf{d}_{\text{obs}}) &= \text{const} \times \exp \left\{ -\mathcal{O}(\mathbf{m}_{\text{map}}) - \frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{map}})^\top \mathbf{B} (\mathbf{m} - \mathbf{m}_{\text{map}}) \right\} \\ &= \text{const} \times \exp \{-\mathcal{O}(\mathbf{m}_{\text{map}})\} \times \\ &\quad \exp \left\{ -\frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{map}})^\top \mathbf{B} (\mathbf{m} - \mathbf{m}_{\text{map}}) \right\} \\ &= \text{const} \times \exp \left\{ -\frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{map}})^\top \mathbf{B} (\mathbf{m} - \mathbf{m}_{\text{map}}) \right\}. \end{aligned} \quad (2.43)$$

The last result shows that $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$ is Gaussian with mean equals to $\mathbf{m}_{\text{map}}$ and covariance given by

$$\mathbf{C}_{\mathbf{m}_c} = \mathbf{B}^{-1} = (\mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G} + \mathbf{C}_{\mathbf{m}}^{-1})^{-1}. \quad (2.44)$$

We use the subscript c to denote that $\mathbf{C}_{\mathbf{m}_c}$ is the covariance for the conditional (*a posteriori*) vector of model parameters.

The formulas derived for the MAP estimate and the posterior covariance involve the inversion of a $N_m \times N_m$ matrix. We can rewrite these expressions in a more convenient way using the Sherman-Morrison-Woodbury (A.9)⁵ formula as

$$\begin{aligned} \mathbf{m}_{\text{map}} &= \mathbf{m}_{\text{pr}} + (\mathbf{C}_{\mathbf{m}}^{-1} + \mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G})^{-1} \mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G} \mathbf{m}_{\text{pr}}) \\ &= \mathbf{m}_{\text{pr}} + \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top (\mathbf{C}_e + \mathbf{G} \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top)^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G} \mathbf{m}_{\text{pr}}) \end{aligned} \quad (2.45)$$

and

$$\begin{aligned} \mathbf{C}_{\mathbf{m}_c} &= (\mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G} + \mathbf{C}_{\mathbf{m}}^{-1})^{-1} \\ &= \mathbf{C}_{\mathbf{m}} - \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top (\mathbf{C}_e + \mathbf{G} \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top)^{-1} \mathbf{G} \mathbf{C}_{\mathbf{m}}. \end{aligned} \quad (2.46)$$

The second line in Eqs. (2.45) and (2.46) involve the inversion of $N_d \times N_d$ matrices. Therefore, these forms are preferable when $N_d < N_m$, which is the typical case in reservoir data assimilation applications.

2.3.1.1 Example: Porosity of a Rock Sample

Consider the problem of estimating the porosity of a rock sample combining a prior information that $\phi \sim \mathcal{N}(\phi_{\text{pr}}, \sigma_{\text{pr}}^2)$ and a measurement $d_{\text{obs}} \sim \mathcal{N}(\phi_{\text{obs}}, \sigma_e^2)$. In this case, the posterior distribution of the porosity corresponds to the Gaussian distribution

$$p(\phi | \phi_{\text{obs}}) = \mathcal{N}(\phi_{\text{map}}, \sigma_{\text{post}}^2). \quad (2.47)$$

ϕ_{map} and σ_{post}^2 can be obtained by the direct application of the expressions developed for $\mathbf{m}_{\text{map}}$ (2.45) and $\mathbf{C}_{\mathbf{m}_c}$ (2.46):

$$\phi_{\text{map}} = \phi_{\text{pr}} + \frac{\sigma_{\text{pr}}^2}{\sigma_e^2 + \sigma_{\text{pr}}^2} (\phi_{\text{obs}} - \phi_{\text{pr}}) \quad (2.48)$$

and

$$\sigma_{\text{post}}^2 = \sigma_{\text{pr}}^2 - \frac{\sigma_{\text{pr}}^4}{\sigma_e^2 + \sigma_{\text{pr}}^2}. \quad (2.49)$$

⁵ See, for example, Oliver et al. [334, Chap. 7] for the algebraic steps to obtain (2.45) and (2.46).

Sometimes, we use the ratio between the posterior and prior variances (normalized variance) as an approximate measure of the uncertainty reduction [334]. From (2.49), the normalized variance, NV, is

$$\text{NV} = \frac{\sigma_{\text{post}}^2}{\sigma_{\text{pr}}^2} = 1 - \frac{\sigma_{\text{pr}}^2}{\sigma_{\text{pr}}^2 + \sigma_e^2} < 1, \quad (2.50)$$

which means that we observe an uncertainty reduction due to the assimilation of ϕ_{obs} . Also note that if we have a perfect (noiseless) measurement, i.e., $\sigma_e = 0$, the normalized variance collapses to zero. Conversely, if we assume an infinitely noisy observation, i.e., $\sigma_e \rightarrow \infty$, we obtain $\phi_{\text{post}} \rightarrow \phi_{\text{pr}}$ and $\text{NV} \rightarrow 1$, which means that there is no uncertainty reduction assimilating a non-informative datum.

2.3.1.2 Example: Porosity of a Rock Sample - Effect of Non-Diagonal Data-Error Covariance

Consider the same problem but with two measurements of porosity. For the sake of this illustration, assume that both measurements gave the same observed value ϕ_{obs} . Moreover, assume that the data-error covariance matrix has the form

$$\mathbf{C}_e = \begin{bmatrix} \sigma_e^2 & \rho\sigma_e^2 \\ \rho\sigma_e^2 & \sigma_e^2 \end{bmatrix}, \quad (2.51)$$

where ρ is the correlation coefficient between the errors of the two observations. In this case, we obtain the following expressions for the posterior mean and NV

$$\phi_{\text{post}}(\rho) = \phi_{\text{pr}} + \frac{2\sigma_{\text{pr}}^2}{2\sigma_{\text{pr}}^2 + \sigma_e^2(1 + \rho)} (\phi_{\text{obs}} - \phi_{\text{pr}}), \quad (2.52)$$

$$\text{NV}(\rho) = 1 - \frac{2\sigma_{\text{pr}}^2}{2\sigma_{\text{pr}}^2 + \sigma_e^2(1 + \rho)}. \quad (2.53)$$

For $\rho = 1$, i.e., the errors in the two observations are perfectly correlated, we have

$$\phi_{\text{post}}(\rho = 1) = \phi_{\text{pr}} + \frac{\sigma_{\text{pr}}^2}{\sigma_{\text{pr}}^2 + \sigma_e^2} (\phi_{\text{obs}} - \phi_{\text{pr}}) \quad (2.54)$$

and

$$\text{NV}(\rho = 1) = 1 - \frac{\sigma_{\text{pr}}^2}{\sigma_{\text{pr}}^2 + \sigma_e^2}, \quad (2.55)$$

which are the same results for the assimilation of a single measurement. To provide a more concrete example, let us consider the following numerical values: $\phi_{\text{pr}} = 0.20$, $\sigma_{\text{pr}}^2 = 0.0025$, $\phi_{\text{obs}} = 0.30$, and $\sigma_e^2 = 0.0025$. Fig. 2.1

shows the plots of ϕ_{post} and NV as a function of ρ . The figure clearly shows that increasing ρ leads to an increase in NV, indicating a reduction in the weight of the measurements. When ρ approaches 1, the problem becomes equivalent to assimilating a single measurement.

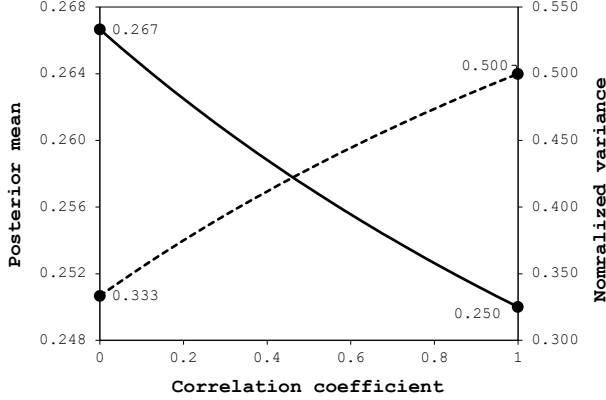


Fig. 2.1: Posterior mean and normalized variance as a function of the correlation coefficient between the observation error of two measurements. Example 2.3.1.2. Reproduced from Emerick and Neto [124] with permission from Elsevier.

2.3.1.3 Example: Porosity and Permeability of a Rock Sample

Consider the same rock sample problem, but with the goal of estimating the porosity and the logarithm of permeability simultaneously, $\mathbf{m} = [\phi \ \ln \kappa]^\top$, from a single measurement of porosity. Assume the prior mean and covariance have the forms

$$\mathbf{m}_{\text{pr}} = \begin{bmatrix} \phi_{\text{pr}} \\ (\ln \kappa)_{\text{pr}} \end{bmatrix} \quad (2.56)$$

and

$$\mathbf{C}_{\mathbf{m}} = \begin{bmatrix} \sigma_{\phi}^2 & \rho \sigma_{\phi} \sigma_{\ln \kappa} \\ \rho \sigma_{\phi} \sigma_{\ln \kappa} & \sigma_{\ln \kappa}^2 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}. \quad (2.57)$$

Because there is only a porosity measurement, the sensitivity matrix is

$$\mathbf{G} = \begin{bmatrix} 1 & 0 \end{bmatrix}. \quad (2.58)$$

The MAP estimate becomes

$$\begin{aligned}
 \mathbf{m}_{\text{map}} &= \mathbf{m}_{\text{pr}} + \mathbf{C}_{\mathbf{m}} \mathbf{G}^{\top} (\mathbf{C}_{\mathbf{e}} + \mathbf{G} \mathbf{C}_{\mathbf{m}} \mathbf{G}^{\top})^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G} \mathbf{m}_{\text{pr}}) \\
 &= \begin{bmatrix} \phi_{\text{pr}} \\ (\ln \kappa)_{\text{pr}} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left(\sigma_e^2 + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right)^{-1} \\
 &\quad \times \left(\phi_{\text{obs}} - \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_{\text{pr}} \\ (\ln \kappa)_{\text{pr}} \end{bmatrix} \right) \\
 &= \begin{bmatrix} \phi_{\text{pr}} + \frac{\sigma_{\phi}^2}{\sigma_e^2 + \sigma_{\phi}^2} (\phi_{\text{obs}} - \phi_{\text{pr}}) \\ (\ln \kappa)_{\text{pr}} + \frac{\rho \sigma_{\phi} \sigma_{\ln \kappa}}{\sigma_e^2 + \sigma_{\phi}^2} (\phi_{\text{obs}} - \phi_{\text{pr}}) \end{bmatrix}. \tag{2.59}
 \end{aligned}$$

The posterior covariance becomes

$$\begin{aligned}
 \mathbf{C}_{\mathbf{m}_c} &= \mathbf{C}_{\mathbf{m}} - \mathbf{C}_{\mathbf{m}} \mathbf{G}^{\top} (\mathbf{C}_{\mathbf{e}} + \mathbf{G} \mathbf{C}_{\mathbf{m}} \mathbf{G}^{\top})^{-1} \mathbf{G} \mathbf{C}_{\mathbf{m}} \\
 &= \begin{bmatrix} \sigma_{\phi}^2 \left(1 - \frac{\sigma_{\phi}^2}{\sigma_e^2 + \sigma_{\phi}^2} \right) & \rho \sigma_{\phi} \sigma_{\ln \kappa} \left(1 - \frac{\sigma_{\phi}^2}{\sigma_e^2 + \sigma_{\phi}^2} \right) \\ \rho \sigma_{\phi} \sigma_{\ln \kappa} \left(1 - \frac{\sigma_{\phi}^2}{\sigma_e^2 + \sigma_{\phi}^2} \right) & \sigma_{\ln \kappa}^2 \left(1 - \frac{\rho^2 \sigma_{\phi}^2}{\sigma_e^2 + \sigma_{\phi}^2} \right) \end{bmatrix}. \tag{2.60}
 \end{aligned}$$

It is interesting to analyze some particular cases:

Case 1: uncorrelated porosity and log-permeability

In this case, the prior correlation coefficient between ϕ and $\ln \kappa$ is $\rho = 0$. The MAP estimate and the posterior covariance become

$$\mathbf{m}_{\text{map}} = \begin{bmatrix} \phi_{\text{pr}} + \frac{\sigma_{\phi}^2}{\sigma_e^2 + \sigma_{\phi}^2} (\phi_{\text{obs}} - \phi_{\text{pr}}) \\ (\ln \kappa)_{\text{pr}} \end{bmatrix} \tag{2.61}$$

and

$$\mathbf{C}_{\mathbf{m}_c} = \begin{bmatrix} \sigma_{\phi}^2 \left(1 - \frac{\sigma_{\phi}^2}{\sigma_e^2 + \sigma_{\phi}^2} \right) & 0 \\ 0 & \sigma_{\ln \kappa}^2 \end{bmatrix}, \tag{2.62}$$

which means that the assimilation of ϕ_{obs} does not reduce the uncertainty in $\ln \kappa$.

Case 2: perfect correlation between porosity and log-permeability

If $\rho = 1$, the posterior covariance become

$$\mathbf{C}_{\mathbf{m}_c} = \left(1 - \frac{\sigma_{\phi}^2}{\sigma_e^2 + \sigma_{\phi}^2} \right) \begin{bmatrix} \sigma_{\phi}^2 & \sigma_{\phi} \sigma_{\ln \kappa} \\ \sigma_{\phi} \sigma_{\ln \kappa} & \sigma_{\ln \kappa}^2 \end{bmatrix} = a \mathbf{C}_{\mathbf{m}}. \tag{2.63}$$

Hence, for $\rho = 1$ all elements of $\mathbf{C}_{\mathbf{m}}$ are updated by the same amount, a . Note that the determinant of $\mathbf{C}_{\mathbf{m}_c}$ is

$$\det \mathbf{C}_{\mathbf{m}_c} = a \left[\sigma_{\phi}^2 \sigma_{\ln \kappa}^2 - (\sigma_{\phi} \sigma_{\ln \kappa}) (\sigma_{\phi} \sigma_{\ln \kappa}) \right] = 0, \tag{2.64}$$

which means that $\mathbf{C}_{\mathbf{m}_c}$ is singular. This result shows that there was only one degree of freedom as the observation of porosity resolves porosity and permeability simultaneously.

Case 3: infinitely noisy data

If $\sigma_e \rightarrow \infty$, we have $\mathbf{m}_{\text{map}} \rightarrow \mathbf{m}_{\text{pr}}$ and $\mathbf{C}_{\mathbf{m}_c} \rightarrow \mathbf{C}_{\mathbf{m}}$. Thus, $p(\mathbf{m}|\mathbf{d}_{\text{obs}}) \rightarrow p(\mathbf{m})$, which means that there is no reduction in uncertainty.

Case 4: perfect (noiseless) data

If $\sigma_e = 0$ the posterior covariance reduces to

$$\mathbf{C}_{\mathbf{m}_c} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{\ln \kappa}^2 (1 - \rho^2) \end{bmatrix}, \quad (2.65)$$

which is singular. Therefore, it is no longer possible to assimilate another observation of porosity because there is no uncertainty left and

$$\mathbf{m}_{\text{map}} = \left[(\ln \kappa)_{\text{pr}} - \frac{\phi_{\text{obs}}}{\sigma_{\phi}} \left(\phi_{\text{obs}} - \phi_{\text{pr}} \right) \right]. \quad (2.66)$$

2.3.1.4 Example: Porosity Distribution in a Core Sample

In the previous examples, we considered the estimation of a single value of porosity for a rock sample based on a measurement and prior information. Now, consider a similar situation, but for a core sample discretized in 50 (equally sized) gridblocks. Assume the prior mean is constant for all gridblocks with value $m_{\text{pr},i} = 0.25$. Moreover, assume that the prior covariance can be modeled with a spherical covariance function with constant prior variance $\sigma_{\text{pr}} = 0.0025$ and range of $L = 10$ gridblocks.

$$\mathcal{C}(h) = \sigma_{\text{pr}}^2 \begin{cases} 1 - \frac{3h}{2L} + \frac{h^3}{2L^3} & \text{for } 0 \leq h \leq L \\ 0 & \text{for } h > L \end{cases} \quad (2.67)$$

For this example, assume that we have ten equally spaced porosity measurements with $\sigma_e = 0.005$. Fig. 2.2a shows the prior mean, the mean $\pm$ three prior standard deviations, and the porosity measurements. The ground truth of porosity is also presented in this figure. Using Eqs. (2.45) and (2.46), we can compute the values of the MAP estimate and posterior variances (Fig. 2.2b). This figure shows that the MAP estimate, which represents the posterior mean, is very smooth compared to the ground truth. This is a consequence of the minimization of Eq. (2.28), which results in a compromise between data matches and changes in $\mathbf{m}$ from the prior, which is smooth. Fig. 2.2b also demonstrates that the posterior uncertainty in the porosity estimate is minimal at the measurement locations and maximal between them.

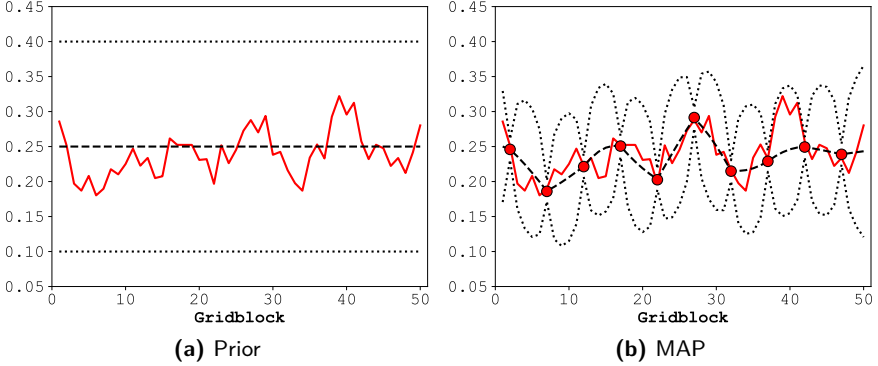


Fig. 2.2: Prior and MAP estimate of porosity in a core sample. The solid red line is the ground truth. The red circles are the observations. The dashed lines are the prior mean in (a) and MAP estimate in (b). The dotted lines correspond to the mean or MAP $\pm$ three standard deviations. Example 2.3.1.4.

2.3.2 Nonlinear Case

When the forward model, $\mathbf{g}(\mathbf{m})$, exhibits nonlinearity with respect to the parameter vector $\mathbf{m}$, the posterior PDF deviates from a Gaussian distribution. In this situation, the posterior may manifest multiple modes, with each mode corresponding to a distinct local minimum of the objective function $\mathcal{O}(\mathbf{m})$. Furthermore, $\mathcal{O}(\mathbf{m})$ can possess several global minima, thereby giving rise to multiple MAP estimates.

In the context of nonlinearity, computing the MAP estimate entails solving a nonlinear minimization problem. In practical applications, the effectiveness of this process is greatly influenced by the performance of the optimization algorithm. For instance, in cases where local minima exist, relying solely on a derivative-based algorithm may not guarantee convergence to an actual MAP estimate.

2.3.2.1 Example: Permeability in a Single Phase Flow

Consider the one-dimensional horizontal reservoir with single-phase flow introduced in [128]. The reservoir model is discretized into 300 gridblocks with a single production well situated at the 300th gridblock. Observations consist of pressure measurements taken over a sequence of draw-down and build-up periods. The model parameters under consideration are the log-permeabilities assigned to each gridblock. The prior value for log-permeability is 5 ln-mD for all gridblocks.

Fig. 2.3 shows the prior mean and the MAP estimate obtained by the minimization of Eq. (2.28) using a quasi-Newton [350] method and adjoint

gradients [259]⁶. For comparison, the ground truth of log-permeability is also included in Fig. 2.3. Remarkably, the MAP estimate exhibits a highly smooth log-permeability profile. Notably, the observed data effectively resolve the model parameters up to approximately 100 gridblocks from the production well. Furthermore, in the first half of the model (gridblocks 1–150), the estimated log-permeabilities revert to the prior mean.

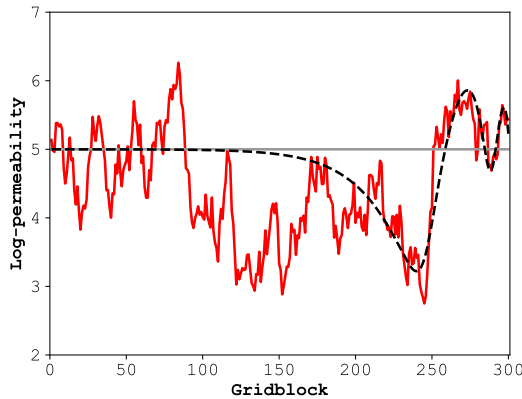


Fig. 2.3: True, prior, and MAP estimate of log-permeability in a single-phase flow. The red line represents the ground truth, the dashed line represents the MAP estimate, and the solid gray line represents the prior mean. Example 2.3.2.1. Reproduced from Emerick and Reynolds [128] with permission from Springer Nature.

2.3.2.2 Example: Permeability in a Two-Phase Flow

Consider a data assimilation problem on a single-layer reservoir model with a two-phase (oil-water) flow. Fig. 2.4a shows the ground truth for the values of log-permeability. This figure also indicates the position of four production and two water injection wells. The prior mean corresponds to a constant value of log-permeability of 5 ln-mD. The ML and MAP estimates were obtained for this problem minimizing Eqs. (2.10) and (2.28), respectively using a quasi-Newton [350] method and adjoint gradients [259].

Fig. 2.4b shows that the lack of regularization resulted in a completely unrealistic permeability field for a ML estimate, although the model is consistent with data as illustrated in Fig. 2.5. The regularization introduced by the term $\mathcal{O}_m(\mathbf{m})$ ensured a smooth MAP estimate (Fig. 2.4c).

⁶ Chapter 3 presents an overview of optimization methods used in the context of reservoir data assimilation.

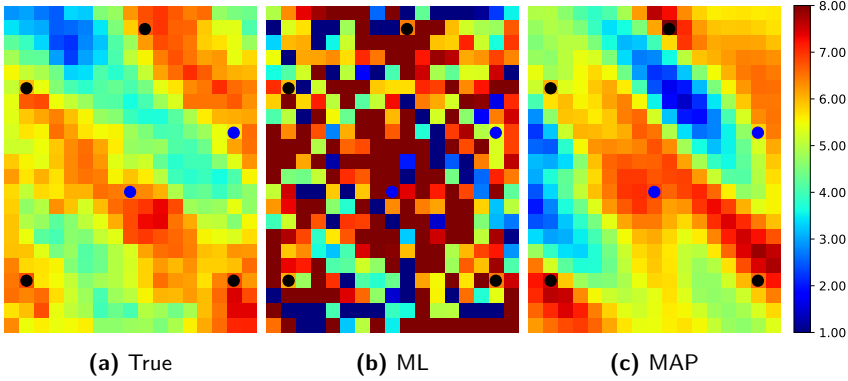


Fig. 2.4: True, ML, and MAP estimates of log-permeability in a two-phase flow. Black circles corresponds to the position of oil producing wells and blue circles the position of water injection wells. Example 2.3.2.2.

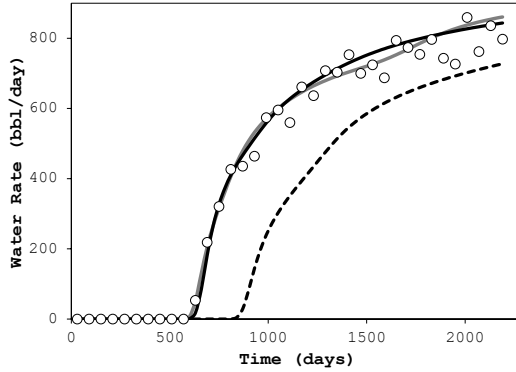


Fig. 2.5: Water production rate in one of the wells of the two-phase flow example. White circles are the observations, solid black curve corresponds to the MAP estimate, gray curve corresponds to ML estimate, and dashed line corresponds to the prior model. Example 2.3.2.2.

2.3.3 Data and Model Errors II

Section 2.2.2 examined the case where data and model errors are additive with zero mean. In this scenario, the equations derived for the ML and MAP estimate can be used with $\mathbf{C}_e$ representing the total data-error covariance, i.e.,

$$\mathbf{C}_e = \mathbf{C}_{e_g} + \mathbf{C}_{e_d}. \quad (2.68)$$

This section considers a less restrictive assumption following Evensen [141]. Rather than assuming an additive model error, we assume that the predicted data vector, $\mathbf{d}$, takes the form

$$\mathbf{d} = \mathbf{g}(\mathbf{m}, \mathbf{e}_g). \quad (2.69)$$

This condition is a more general representation of model errors because the predicted data are nonlinear functions of $\mathbf{m}$ and $\mathbf{e}_g$. Such scenarios are not uncommon in practical applications. For instance, Evensen and Eikrem [143] points out a typical situation in petroleum reservoir simulation where observed liquid rates are imposed to control the wells during the historical period. In this case, the measurement errors in liquid rates introduce nonlinear perturbations in the predicted data.

Assuming that the prior PDFs of $\mathbf{m}$ and $\mathbf{e}_g$ are independent,

$$p(\mathbf{m}, \mathbf{e}_g) = p(\mathbf{m})p(\mathbf{e}_g), \quad (2.70)$$

the joint PDF becomes

$$p(\mathbf{d}, \mathbf{m}, \mathbf{e}_g) = \delta(\mathbf{d} - \mathbf{g}(\mathbf{m}, \mathbf{e}_g)) p(\mathbf{m})p(\mathbf{e}_g), \quad (2.71)$$

where $\delta(\mathbf{d} - \mathbf{g}(\mathbf{m}, \mathbf{e}_g))$ is the Dirac delta function.

The likelihood function for the observations given the predictions is

$$\mathcal{L}(\mathbf{d}) = p(\mathbf{d}_{\text{obs}}|\mathbf{d}). \quad (2.72)$$

Thus, the joint posterior PDF becomes

$$p(\mathbf{d}, \mathbf{m}, \mathbf{e}_g|\mathbf{d}_{\text{obs}}) \propto p(\mathbf{d}_{\text{obs}}|\mathbf{d})\delta(\mathbf{d} - \mathbf{g}(\mathbf{m}, \mathbf{e}_g)) p(\mathbf{m})p(\mathbf{e}_g). \quad (2.73)$$

Because $\mathbf{d}$ is a function of $\mathbf{m}$ and $\mathbf{e}_g$, we can compute the marginal PDF

$$\begin{aligned} p(\mathbf{m}, \mathbf{e}_g|\mathbf{d}_{\text{obs}}) &\propto \int p(\mathbf{d}_{\text{obs}}|\mathbf{d})\delta(\mathbf{d} - \mathbf{g}(\mathbf{m}, \mathbf{e}_g)) p(\mathbf{m})p(\mathbf{e}_g)d\mathbf{d} \\ &= p(\mathbf{d}_{\text{obs}}|\mathbf{g}(\mathbf{m}, \mathbf{e}_g)) p(\mathbf{m})p(\mathbf{e}_g). \end{aligned} \quad (2.74)$$

Eq. (2.74) gives the posterior joint PDF of $\mathbf{m}$ and $\mathbf{e}_g$ given the observations. Assuming Gaussian priors

$$p(\mathbf{m}) = \mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_{\mathbf{m}}), \quad (2.75)$$

$$p(\mathbf{e}_g) = \mathcal{N}(\mathbf{e}_{g,\text{pr}}, \mathbf{C}_{\mathbf{e}_g}), \quad (2.76)$$

and

$$p(\mathbf{d}_{\text{obs}}|\mathbf{g}(\mathbf{m}, \mathbf{e}_g)) = \mathcal{N}(\mathbf{g}(\mathbf{m}, \mathbf{e}_g) - \mathbf{d}_{\text{obs}}, \mathbf{C}_{\mathbf{e}_d}) \quad (2.77)$$

leads to

$$\begin{aligned}
 p(\mathbf{m}, \mathbf{e}_g | \mathbf{d}_{\text{obs}}) \propto \exp \left\{ -\frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}) - \right. \\
 \frac{1}{2} (\mathbf{e}_g - \mathbf{e}_{g,\text{pr}})^\top \mathbf{C}_{\mathbf{e}_g}^{-1} (\mathbf{e}_g - \mathbf{e}_{g,\text{pr}}) - \\
 \left. \frac{1}{2} (\mathbf{g}(\mathbf{m}, \mathbf{e}_g) - \mathbf{d}_{\text{obs}})^\top \mathbf{C}_{\mathbf{e}_d}^{-1} (\mathbf{g}(\mathbf{m}, \mathbf{e}_g) - \mathbf{d}_{\text{obs}}) \right\}. \quad (2.78)
 \end{aligned}$$

Eq. (2.78) assumed that $\mathbf{m}$ and $\mathbf{e}_g$ are independent. This condition can be relaxed defining an augmented vector $\mathbf{y}^\top = [\mathbf{m}^\top \ \mathbf{e}_g^\top]$ and writing the covariance as

$$\mathbf{C}_{\mathbf{y}} = \begin{bmatrix} \mathbf{C}_{\mathbf{m}} & \mathbf{C}_{\mathbf{m}, \mathbf{e}_g} \\ \mathbf{C}_{\mathbf{e}_g, \mathbf{m}} & \mathbf{C}_{\mathbf{e}_g} \end{bmatrix}. \quad (2.79)$$

Hence

$$\begin{aligned}
 p(\mathbf{y} | \mathbf{d}_{\text{obs}}) \propto \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{y}_{\text{pr}})^\top \mathbf{C}_{\mathbf{y}}^{-1} (\mathbf{y} - \mathbf{y}_{\text{pr}}) - \right. \\
 \left. \frac{1}{2} (\mathbf{g}(\mathbf{y}) - \mathbf{d}_{\text{obs}})^\top \mathbf{C}_{\mathbf{e}_d}^{-1} (\mathbf{g}(\mathbf{y}) - \mathbf{d}_{\text{obs}}) \right\}. \quad (2.80)
 \end{aligned}$$

This expression represents the joint posterior, which can be used to compute the MAP estimate, $\mathbf{y}_{\text{map}}$, or to generate samples with the methods discussed in the sequence of this book. In this formulation, we estimate simultaneously the model parameters $\mathbf{m}$ and model errors $\mathbf{e}_g$. However, because $\mathbf{e}_g$ varies with time, the estimated values are valid only for the historical period. They cannot be directly applied to the forecast period.

2.4 Data Assimilation as a Sampling Problem

Data assimilation is frequently characterized as an ill-posed inverse problem due to the fact that the quantity of independent data accessible is substantially less than the number of uncertain parameters. Consequently, there exists an infinite number of combinations of these parameters capable of generating models that can match the observations within a tolerance range. Furthermore, the available data are always inaccurate and occasionally inconsistent. As a result, reservoir models are built under conditions of significant uncertainty, leading to inherently uncertain predictions.

Bayesian statistics offers a straightforward framework for handling uncertainty. Bayes' rule enables us to establish the posterior PDF for model parameters based on field measurements. Despite our ability to formulate these PDF expressions, they often have limited practical utility. Practitioners truly seek access to models conditioned on observations, allowing them to make performance predictions under various operational conditions and informed decisions.

The MAP estimate represents a mode of this distribution. However, a single model is insufficient for a useful characterization of uncertainty. Consequently, sampling the PDF emerges as the practical solution for quantifying uncertainty. If a collection of realizations for the model parameter vector serves as a set of samples drawn from the posterior PDF, then an accurate evaluation of uncertainty in specific outcomes of model predictions can be achieved by generating predictions with each model and constructing statistics for the resulting set of outcomes.

For example, consider the problem of computing the expectation of $\mathbf{g}(\mathbf{m})$. By definition

$$\mathbb{E}[\mathbf{g}(\mathbf{m})] = \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{m})p(\mathbf{m})d\mathbf{m}, \quad (2.81)$$

where $p(\mathbf{m})$ is the PDF of $\mathbf{m}$. Monte Carlo methods evaluate $\mathbb{E}[\mathbf{g}(\mathbf{m})]$ by drawing N_e samples from $p(\mathbf{m})$ and approximating the integral as

$$\mathbb{E}[\mathbf{g}(\mathbf{m})] \approx \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{g}(\mathbf{m}_j). \quad (2.82)$$

If the samples $\{\mathbf{m}_j\}$ are independent, the Law of Large Numbers⁷ ensures that the approximation converges as $N_e \rightarrow \infty$ (we control the accuracy of our estimate by increasing N_e). It is required, however, that $\{\mathbf{m}_j\}$ are generated from a process that draws samples from $p(\mathbf{m})$ in the correct proportions⁸.

2.4.1 Sampling in the Linear Case

Let's first analyze the problem of sampling from a multivariate prior Gaussian distribution and the posterior for the case where the predicted data are linearly related to the model⁹.

2.4.1.1 Sampling the Prior PDF

Let the prior PDF of $\mathbf{m}$ be a multivariate Gaussian, i.e., $p(\mathbf{m}) = \mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_{\mathbf{m}})$, in which case we can write

⁷ The law of large numbers is a fundamental theorem in probability and statistics that states that as the size of a sample increases, the sample mean converges to the expected value of the underlying distribution.

⁸ Note that the samples don't need to be "equiprobable." The requirement is that they are samples drawn from $p(\mathbf{m})$.

⁹ Practical data assimilation is typically nonlinear. However, the solutions for linear problems serve as the basis for developing methods for the nonlinear.

$$p(\mathbf{m}) = \text{const} \times \exp \left\{ -\frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}) \right\}. \quad (2.83)$$

There are various methods available to sample this PDF. The simplest procedure is based on the square root of $\mathbf{C}_{\mathbf{m}}$. In this case, a sample from $p(\mathbf{m})$ is given by

$$\mathbf{m}_j = \mathbf{m}_{\text{pr}} + \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_j, \quad (2.84)$$

where, $\mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, i.e., $\mathbf{z}_j$ is a vector of random normal deviates. $\mathbf{C}_{\mathbf{m}}^{1/2}$ is a square root of $\mathbf{C}_{\mathbf{m}}$ (see Appendix A, Section A.6).

In order to show that $\mathbf{m}_j$ generated with Eq. (2.84) is a sample from $\mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_{\mathbf{m}})$, suffices to show that $\mathbb{E}[\mathbf{m}_j] = \mathbf{m}_{\text{pr}}$ and $\mathbb{C}[\mathbf{m}_j] = \mathbf{C}_{\mathbf{m}}$. These results follow directly from the definitions of expectation and covariance

$$\begin{aligned} \mathbb{E}[\mathbf{m}_j] &= \mathbb{E} \left[\mathbf{m}_{\text{pr}} + \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_j \right] \\ &= \mathbb{E}[\mathbf{m}_{\text{pr}}] + \mathbf{C}_{\mathbf{m}}^{1/2} \mathbb{E}[\mathbf{z}_j] \\ &= \mathbf{m}_{\text{pr}} \end{aligned} \quad (2.85)$$

and

$$\begin{aligned} \mathbb{C}[\mathbf{m}_j] &= \mathbb{E} \left[(\mathbf{m}_j - \mathbb{E}[\mathbf{m}_j]) (\mathbf{m}_j - \mathbb{E}[\mathbf{m}_j])^\top \right] \\ &= \mathbb{E} \left[(\mathbf{m}_j - \mathbf{m}_{\text{pr}}) (\mathbf{m}_j - \mathbf{m}_{\text{pr}})^\top \right] \\ &= \mathbb{E} \left[\left(\mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_j \right) \left(\mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_j \right)^\top \right] \\ &= \mathbb{E} \left[\mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_j \mathbf{z}_j^\top \mathbf{C}_{\mathbf{m}}^{\top/2} \right] \\ &= \mathbf{C}_{\mathbf{m}}^{1/2} \mathbb{E}[\mathbf{z}_j \mathbf{z}_j^\top] \mathbf{C}_{\mathbf{m}}^{\top/2} \\ &= \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{C}_{\mathbf{m}}^{\top/2} \\ &= \mathbf{C}_{\mathbf{m}}. \end{aligned} \quad (2.86)$$

Any valid square root of $\mathbf{C}_{\mathbf{m}}$ can be used in Eq. (2.84). Often, the Cholesky decomposition is employed, in which case we write

$$\mathbf{C}_{\mathbf{m}} = \mathbf{L} \mathbf{L}^\top, \quad (2.87)$$

where $\mathbf{L}$ is a lower triangular matrix and $\mathbf{C}_{\mathbf{m}}^{1/2} = \mathbf{L}$.

Unfortunately, for large models, the factorization of $\mathbf{C}_{\mathbf{m}}$ becomes unfeasible. There are more efficient alternatives [454]. However, in reservoir modeling, we typically resort to geostatistical methods, in which case the most common is the sequential Gaussian simulation (SGS) method [171, 104].

2.4.1.2 Example: Cross-covariances

Often there are prior correlations between properties, e.g., porosity and log-permeability. In this case, we have non-zero cross-covariance terms in $\mathbf{C}_m$, i.e.,

$$\mathbf{C}_m = \begin{bmatrix} \mathbf{C}_\phi & \mathbf{C}_{\phi, \ln \kappa} \\ \mathbf{C}_{\ln \kappa, \phi} & \mathbf{C}_{\ln \kappa} \end{bmatrix}, \quad (2.88)$$

where $\mathbf{C}_{\phi, \ln \kappa} = \mathbf{C}_{\ln \kappa, \phi}^\top$.

One problem is ensuring that the cross-covariance matrix is positive-definite. A fairly simple way to do this is to use the screening hypothesis [471], in which case the entries of the cross-covariance matrices are computed as

$$\mathbb{C}[\phi_i, \ln \kappa_j] = \mathbb{C}[\ln \kappa_j, \phi_i] = \frac{\rho \sigma_{\phi_i}}{\sigma_{\ln \kappa_j}} \mathbb{C}[\ln \kappa_j] = \frac{\rho \sigma_{\ln \kappa_j}}{\sigma_{\phi_i}} \mathbb{C}[\phi_i]. \quad (2.89)$$

Consider the following prior information about the porosity and log-permeability of a rock sample:

- $p(\phi) = \mathcal{N}(0.2, 0.05^2)$.
- $p(\ln \kappa) = \mathcal{N}(1, 0.5^2)$.
- $\rho = 0.8$.

Let $\mathbf{m} = [\phi \quad \ln \kappa]^\top$. The prior covariance of $\mathbf{m}$ is

$$\begin{aligned} \mathbf{C}_m &= \begin{bmatrix} \sigma_\phi^2 & \rho \sigma_\phi \sigma_{\ln \kappa} \\ \rho \sigma_\phi \sigma_{\ln \kappa} & \sigma_{\ln \kappa}^2 \end{bmatrix} \\ &= \begin{bmatrix} 0.05^2 & 0.8 \times 0.05 \times 0.5 \\ 0.8 \times 0.05 \times 0.5 & 0.5^2 \end{bmatrix} \\ &= \begin{bmatrix} 0.0025 & 0.02 \\ 0.02 & 0.25 \end{bmatrix}. \end{aligned} \quad (2.90)$$

Because ϕ and $\ln \kappa$ are correlated, we need to compute $\mathbf{C}_m^{1/2}$ to draw samples from $p(\phi, \ln \kappa)$. This can be done using the Cholesky decomposition. The Cholesky of a 2×2 is trivial to compute. First, we write

$$\begin{aligned} \mathbf{C}_m = \mathbf{L}\mathbf{L}^\top &= \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \\ &= \begin{bmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{bmatrix} = \begin{bmatrix} 0.0025 & 0.02 \\ 0.02 & 0.25 \end{bmatrix}. \end{aligned} \quad (2.91)$$

Hence $a = 0.05$, $b = 0.4$ and $c = 0.3$.

Now we can draw samples using

$$\mathbf{m} = \mathbf{m}_{\text{pr}} + \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}, \quad (2.92)$$

where $\mathbf{m}_{\text{pr}} = [0.2 \ 1]^\top$ and $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Fig. 2.6 shows a cross-plot between ϕ and $\ln \kappa$ obtained using Eq. (2.92) illustrating that the resulting samples are correlated.

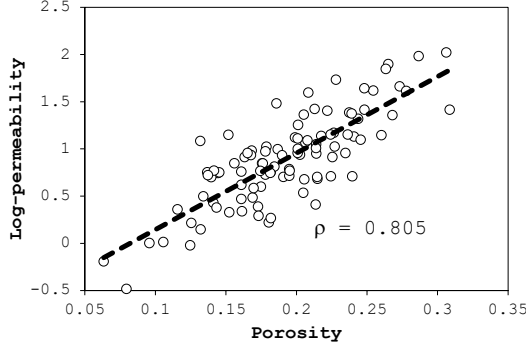


Fig. 2.6: Cross-plot between porosity and log-permeability obtained with 100 correlated samples. Example 2.4.1.2.

2.4.1.3 Example: Porosity Realizations

Fig. 2.7 shows two samples from a multivariate Gaussian distribution used to model the spatial distribution of porosity in a 2D uniform grid with 100×100 gridblocks. These samples were generated using the Cholesky decomposition. The prior mean is assumed to be constant and equal to 0.2, and the prior standard deviation is 0.05. The covariance was computed using a spherical model (see Appendix C, Section C.4.3) with and without anisotropy.

2.4.1.4 Sampling the Posterior PDF

As discussed before, if the prior model is Gaussian and the relationship between model and data is linear, then the posterior PDF is also Gaussian with the form

$$p(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \text{const} \times \exp \left\{ -\frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{map}})^\top \mathbf{C}_{\mathbf{m}_c}^{-1} (\mathbf{m} - \mathbf{m}_{\text{map}}) \right\}. \quad (2.93)$$

We can use the same procedure described to sample the prior and compute the square root of $\mathbf{C}_{\mathbf{m}_c}$ to generate samples from $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$. Nevertheless,

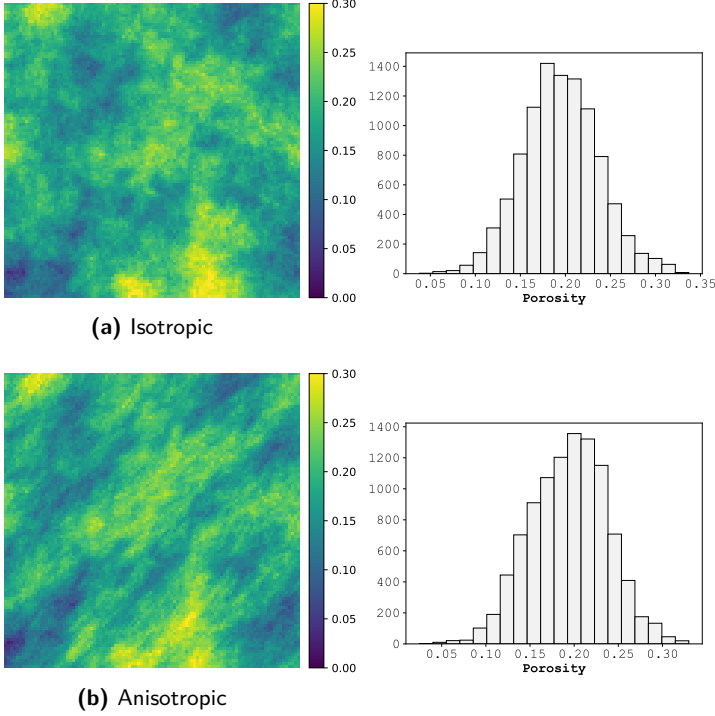


Fig. 2.7: Porosity models sampled from a multivariate Gaussian using Cholesky decomposition of the prior covariance matrix. The isotropic case uses a covariance matrix with range corresponding to 30 gridblocks. The anisotropic case uses a major range of 40 and a minor of 20 gridblocks, rotated by 45° . Example 2.4.1.3.

this approach is only practical when dealing with a low-dimensional model. Fortunately, there is an alternative method for generating conditional samples $\mathbf{m}_{c,j} \sim \mathcal{N}(\mathbf{m}_{\text{map}}, \mathbf{C}_{\mathbf{m}_c})$ by solving the following optimization problem:

$$\mathbf{m}_{c,j} = \arg \min_{\mathbf{m}} \mathcal{O}_r(\mathbf{m}), \quad (2.94)$$

where $\mathcal{O}_r(\mathbf{m})$ is a modified (randomized) objective function given by

$$\begin{aligned} \mathcal{O}_r(\mathbf{m}) = & \frac{1}{2} (\mathbf{d}_{\text{obs},j} - \mathbf{G}\mathbf{m})^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{G}\mathbf{m}) + \\ & \frac{1}{2} (\mathbf{m} - \mathbf{m}_j)^\top \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_j). \end{aligned} \quad (2.95)$$

Note the differences between the original $\mathcal{O}(\mathbf{m})$ and $\mathcal{O}_r(\mathbf{m})$. In Eq. (2.95), we replaced $\mathbf{d}_{\text{obs}}$ by a random sample $\mathbf{d}_{\text{obs},j} \sim \mathcal{N}(\mathbf{d}_{\text{obs}}, \mathbf{C}_{\mathbf{e}})$ and $\mathbf{m}_{\text{pr}}$ by a prior sample $\mathbf{m}_j \sim \mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_{\mathbf{m}})$.

The minimum of $\mathcal{O}_r(\mathbf{m})$ can be obtained with the same procedure used to derive $\mathbf{m}_{\text{map}}$. The resulting expressions are similar to the ones obtained

for the MAP estimate

$$\begin{aligned}
 \mathbf{m}_{c,j} &= (\mathbf{C}_{\mathbf{m}}^{-1} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G})^{-1} (\mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_j + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs},j}) \\
 &= \mathbf{m}_j + (\mathbf{C}_{\mathbf{m}}^{-1} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G})^{-1} \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{G} \mathbf{m}_j) \\
 &= \mathbf{m}_j + \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top (\mathbf{C}_{\mathbf{e}} + \mathbf{G} \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top)^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{G} \mathbf{m}_j), \quad (2.96)
 \end{aligned}$$

where $\mathbf{m}_{c,j}$ is a sample from the posterior. This procedure is known as randomized maximum likelihood (RML) [332].

In order to show that $\mathbf{m}_{c,j} \sim \mathcal{N}(\mathbf{m}_{\text{map}}, \mathbf{C}_{\mathbf{m}_c})$, suffices to show that $\mathbb{E}[\mathbf{m}_{c,j}] = \mathbf{m}_{\text{map}}$ and $\mathbb{C}[\mathbf{m}_{c,j}] = \mathbf{C}_{\mathbf{m}_c}$. For the expectation follows that

$$\begin{aligned}
 \mathbb{E}[\mathbf{m}_{c,j}] &= \mathbb{E} \left[(\mathbf{C}_{\mathbf{m}}^{-1} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G})^{-1} (\mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_j + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs},j}) \right] \\
 &= \mathbb{E} [\mathbf{C}_{\mathbf{m}_c} (\mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_j + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs},j})] \\
 &= \mathbf{C}_{\mathbf{m}_c} \{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbb{E}[\mathbf{m}_j] + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbb{E}[\mathbf{d}_{\text{obs},j}] \} \\
 &= \mathbf{C}_{\mathbf{m}_c} \{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_{\text{pr}} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs}} \} \\
 &= \mathbf{m}_{\text{map}}. \quad (2.97)
 \end{aligned}$$

The calculation of the covariance requires more steps, but it is relatively straightforward starting from the definition of covariances

$$\begin{aligned}
 \mathbb{C}[\mathbf{m}_{c,j}] &= \mathbb{E} \left[(\mathbf{m}_{c,j} - \mathbb{E}[\mathbf{m}_{c,j}]) (\mathbf{m}_{c,j} - \mathbb{E}[\mathbf{m}_{c,j}])^\top \right] \\
 &= \mathbb{E} \left[(\mathbf{m}_{c,j} - \mathbf{m}_{\text{map}}) (\mathbf{m}_{c,j} - \mathbf{m}_{\text{map}})^\top \right]. \quad (2.98)
 \end{aligned}$$

Let's first develop an expression for $\mathbf{m}_{c,j} - \mathbf{m}_{\text{map}}$:

$$\begin{aligned}
 \mathbf{m}_{c,j} - \mathbf{m}_{\text{map}} &= \mathbf{C}_{\mathbf{m}_c} \{ (\mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_j + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs},j}) - \\
 &\quad (\mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_{\text{pr}} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs}}) \} \\
 &= \mathbf{C}_{\mathbf{m}_c} \{ \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m}_j - \mathbf{m}_{\text{pr}}) + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{d}_{\text{obs}}) \} \\
 &= \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_{\mathbf{m},j} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{C}_{\mathbf{e}}^{1/2} \mathbf{z}_{\mathbf{d},j} \right\}. \quad (2.99)
 \end{aligned}$$

In (2.99), we used

$$\mathbf{m}_j = \mathbf{m}_{\text{pr}} + \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_{\mathbf{m},j} \quad (2.100)$$

$$\mathbf{d}_{\text{obs},j} = \mathbf{d}_{\text{obs}} + \mathbf{C}_{\mathbf{e}}^{1/2} \mathbf{z}_{\mathbf{d},j}, \quad (2.101)$$

where both $\mathbf{z}_{\mathbf{m},j}$ and $\mathbf{z}_{\mathbf{d},j}$ are random vectors sampled from $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

Now, we compute the cross product

$$\begin{aligned}
 & (\mathbf{m}_{c,j} - \mathbf{m}_{\text{map}})(\mathbf{m}_{c,j} - \mathbf{m}_{\text{map}})^\top = \\
 & = \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_{\mathbf{m},j} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{C}_{\mathbf{e}}^{1/2} \mathbf{z}_{\mathbf{d},j} \right\} \times \\
 & \quad \left\{ \mathbf{z}_{\mathbf{m},j}^\top \mathbf{C}_{\mathbf{m}}^{\top/2} \mathbf{C}_{\mathbf{m}}^{-1} + \mathbf{z}_{\mathbf{d},j}^\top \mathbf{C}_{\mathbf{e}}^{\top/2} \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G} \right\} \mathbf{C}_{\mathbf{m}_c} \\
 & = \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_{\mathbf{m},j} \mathbf{z}_{\mathbf{m},j}^\top \mathbf{C}_{\mathbf{m}}^{\top/2} \mathbf{C}_{\mathbf{m}}^{-1} + \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_{\mathbf{m},j} \mathbf{z}_{\mathbf{d},j}^\top \mathbf{C}_{\mathbf{e}}^{\top/2} \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G} + \right. \\
 & \quad \left. \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{C}_{\mathbf{e}}^{1/2} \mathbf{z}_{\mathbf{d},j} \mathbf{z}_{\mathbf{m},j}^\top \mathbf{C}_{\mathbf{m}}^{\top/2} \mathbf{C}_{\mathbf{m}}^{-1} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{C}_{\mathbf{e}}^{1/2} \mathbf{z}_{\mathbf{d},j} \mathbf{z}_{\mathbf{d},j}^\top \mathbf{C}_{\mathbf{e}}^{\top/2} \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G} \right\} \mathbf{C}_{\mathbf{m}_c}.
 \end{aligned} \tag{2.102}$$

Finally, we note that $\mathbb{E}[\mathbf{z}_{\mathbf{m},j} \mathbf{z}_{\mathbf{m},j}^\top] = \mathbf{I}$, $\mathbb{E}[\mathbf{z}_{\mathbf{d},j} \mathbf{z}_{\mathbf{d},j}^\top] = \mathbf{I}$ and $\mathbb{E}[\mathbf{z}_{\mathbf{m},j} \mathbf{z}_{\mathbf{d},j}^\top] = \mathbb{E}[\mathbf{z}_{\mathbf{d},j} \mathbf{z}_{\mathbf{m},j}^\top] = \mathbf{O}$ (null matrix) because $\mathbf{z}_{\mathbf{m},j}$ and $\mathbf{z}_{\mathbf{d},j}$ are independent. The covariance becomes

$$\begin{aligned}
 \mathbb{C}[\mathbf{m}_{c,j}] &= \mathbb{E}[(\mathbf{m}_{c,j} - \mathbf{m}_{\text{map}})(\mathbf{m}_{c,j} - \mathbf{m}_{\text{map}})^\top] \\
 &= \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G} \right\} \mathbf{C}_{\mathbf{m}_c} \\
 &= \mathbf{C}_{\mathbf{m}_c} \mathbf{C}_{\mathbf{m}_c}^{-1} \mathbf{C}_{\mathbf{m}_c} \\
 &= \mathbf{C}_{\mathbf{m}_c},
 \end{aligned} \tag{2.103}$$

which completes the demonstration that $\mathbf{m}_{c,j} \sim \mathcal{N}(\mathbf{m}_{\text{map}}, \mathbf{C}_{\mathbf{m}_c})$.

2.4.1.5 Example: Porosity Distribution in a Core Sample

Consider the same problem of Example 2.3.1.4. Instead of the MAP estimate, we can generate 100 samples of the prior using the square root method (Fig. 2.8a). For each prior realization, we applied Eq. (2.96) to obtain a posterior sample (Fig. 2.8b). Note that all posterior realizations are consistent with the observed data points. The average of the posterior realizations is an approximation of the MAP estimate (compare Figs 2.2 and 2.8b)¹⁰.

¹⁰ The average converges to the MAP as we increase the number of samples.

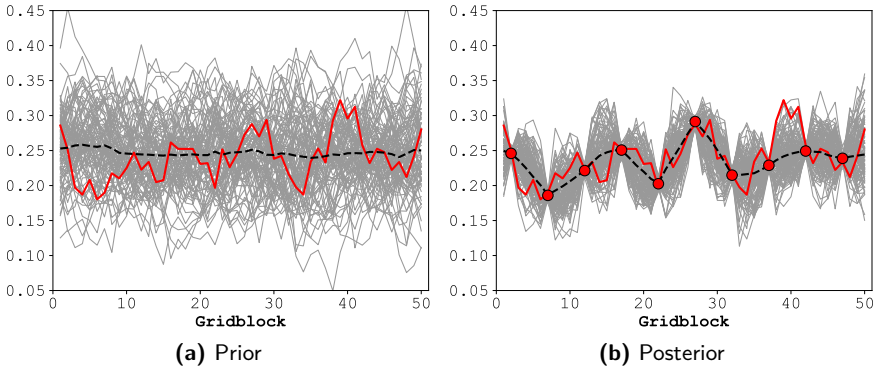


Fig. 2.8: Prior and posterior realizations of porosity in a core sample. The solid red line is the ground truth. The red circles are the observations. The gray lines are samples from the prior (a) and posterior (b). The dashed lines are the average computed over the 100 samples. Example 2.4.1.5.

Remark 2.4: Relation to Geostatistics

Geostatistics, as a statistical discipline, primarily focuses on analyzing spatial and temporal datasets. Its roots can be traced back to the early days when it was closely associated with interpolation techniques, particularly in the development of kriging methods [242, 295]. However, contemporary geostatistics has expanded well beyond the realms of basic interpolation. Modern geostatistics embraces a Bayesian framework, with its core objective being the generation of samples from complex, high-dimensional probability distributions. Notably, traditional geostatistical techniques like simple kriging and SGS find direct parallels in the linear-Gaussian context: simple kriging corresponds to the posterior mean^a estimate, while models generated using SGS are akin to samples drawn from the posterior distribution.

^a One important difference between kriging and the MAP estimate is the fact that in kriging, we typically assume that data are perfect. This is the so-called “hard data assumption.”

2.4.2 Sampling in the Nonlinear Case

In the nonlinear case, estimators such as mean and covariance are not enough to fully characterize the PDF, and there are no general analytical solutions. Moreover, apart from problems with low dimensionality, a “brutal force” exploratory sampling is unfeasible. This happens because high-dimensional

PDFs are extremely diffusive. This phenomenon is often referred to as the “curse of dimensionality”¹¹.

To illustrate this issue, consider the example of sampling from a multivariate Gaussian distribution $\mathbf{m} \sim p(\mathbf{m}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. In this standard Gaussian distribution, the mode is centered at the origin, suggesting that a significant portion of the probability mass should concentrate near this point. However, this expectation holds only in cases of very low dimensions. To quantify this behavior, we can compute the probability of a random vector $\mathbf{m}$ being within a distance of 1 from the origin, denoted as $P(\mathbf{m}^\top \mathbf{m} \leq 1)$. We note that:

$$\mathbf{m}^\top \mathbf{m} = \sum_{i=1}^{N_m} m_i^2 = x, \quad (2.104)$$

where x follows a chi-squared distribution with N_m degrees of freedom (see Appendix C, Section C.1.18.4), i.e., $x \sim \chi^2(N_m)$. Therefore, we can compute the probability $P(x = \mathbf{m}^\top \mathbf{m} \leq 1)$ using the cumulative density function of $\chi^2(N_m)$

$$P(x \leq 1) = \frac{1}{\Gamma(\frac{N_m}{2})} \gamma\left(\frac{N_m}{2}, \frac{x}{2}\right), \quad (2.105)$$

where $\Gamma(\cdot)$ represents the gamma function, and $\gamma(\cdot, \cdot)$ denotes the incomplete gamma function. As illustrated in Fig. 2.9, we observe that the probability mass around the origin in a multivariate Gaussian distribution diminishes rapidly as we increase the dimensionality N_m . For instance, with $N_m = 10$, less than 10^{-3} of the probability mass lies within a radius of 1 from the origin. For $N_m = 20$, this probability is on the order of 10^{-10} . For this reason, sampling high-dimensional PDFs can be very challenging.

In the context of reservoir data assimilation, we are interested in sampling the posterior PDF of model parameters given a set of measurements. However, the number of parameters can be extremely high; we may have problems with 10^7 parameters or even more. The number of observations is also high, often more than 10^3 data points. This means that the probability of getting a model consistent with the observation with a naive acceptance-rejection algorithm is extremely low.

2.4.3 Markov Chain Monte Carlo

There are general sampling methods that can be used to sample $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$ in the nonlinear case. Among these methods, Markov chain Monte Carlo (MCMC) is perhaps the best known. In MCMC, a sequence of “states” is generated by simulating a Markov chain, where the probability of transition-

¹¹ The term “curse of dimensionality” was originally coined by the mathematician Richard E. Bellman within the context of dynamic programming [33].

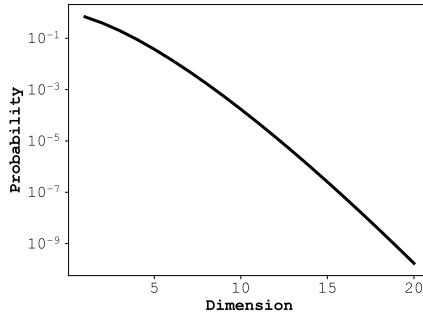


Fig. 2.9: Probability of a random vector be at a distance less than one in a standard multivariate Gaussian distribution.

ing to a new state depends solely on the previous state. As the chain becomes sufficiently long, the sequence of states effectively constitutes a sampling of the desired target distribution. Under reasonable assumptions, it is possible to demonstrate that MCMC accurately samples the target distribution as the number of states in the chain approaches infinity¹².

2.4.3.1 Metropolis-Hastings Algorithm

The foundation of many MCMC implementations rests upon the seminal work from Metropolis et al. [302]¹³, which introduced the concept of defining transition probabilities. In the Metropolis algorithm, the probability of transitioning from a state $\mathbf{m}$ to $\hat{\mathbf{m}}$ is computed as the product of the probability of proposing this transition, denoted as $q(\mathbf{m}, \hat{\mathbf{m}})$, and the probability of accepting the proposal. Hastings [187] expanded the Metropolis algorithm, giving rise to what is now one of the most widely adopted MCMC techniques – the Metropolis-Hastings algorithm.

Consider $p^*(\mathbf{m})$ as the target PDF we aim to sample, and let $q(\cdot, \cdot)$ represent a proposal PDF easy to sample from. As long as the proposal mechanism allows for reaching any state with $p^*(\mathbf{m}) \neq 0$ within a finite number of transitions, the states in the chain generated by the Metropolis-Hastings algorithm will eventually represent samples from $p^*(\mathbf{m})$, regardless of the initial state $\mathbf{m}^0$. Computing the acceptance probability, denoted as $P(\mathbf{m}^\ell, \hat{\mathbf{m}}^{\ell+1})$, only requires the evaluation of the target PDF up to its normalization constant.

¹² The body of literature dedicated to MCMC methods is extensive and wide-ranging. A good introduction discussion on MCMC methods is presented in [166].

¹³ The Metropolis algorithm was selected as one of the ten algorithms that have had the greatest influence on the development and practice of science and engineering in the 20th Century – Cipra, B.A., *The Best of the 20th Century: Editors Name Top 10 Algorithms*, SIAM News (2000) [88].

The main challenge in MCMC, especially for high-dimensional problems, is the slow convergence. The selection of the proposal distribution is crucial, as a well-chosen proposal can greatly influence the acceptance rate of new states within the chain, thereby reducing the number of iterations needed to converge to the target distribution.

Pseudo-code 2.1 provides a concise summary of the Metropolis-Hastings MCMC algorithm.

Pseudo-code 2.1: MCMC - Metropolis-Hastings

1. Set $\ell = 0$ (iteration index) and $\mathbf{m}^\ell = \mathbf{m}^0$ (initial state).
2. Propose a new state $\hat{\mathbf{m}}^{\ell+1}$ from $q(\mathbf{m}^\ell, \hat{\mathbf{m}}^{\ell+1})$.
3. Compute the probability of accepting the proposed state

$$P(\mathbf{m}^\ell, \hat{\mathbf{m}}^{\ell+1}) = \min \left\{ 1, \frac{q(\hat{\mathbf{m}}^{\ell+1}, \mathbf{m}^\ell) p^*(\hat{\mathbf{m}}^{\ell+1})}{q(\mathbf{m}^\ell, \hat{\mathbf{m}}^{\ell+1}) p^*(\mathbf{m}^\ell)} \right\}. \quad (2.106)$$

4. Sample $u \sim \mathcal{U}[0, 1]$. If $u \leq P(\mathbf{m}^\ell, \hat{\mathbf{m}}^{\ell+1})$, accept the new state, $\mathbf{m}^{\ell+1} = \hat{\mathbf{m}}^{\ell+1}$. Otherwise, repeat the old state, $\mathbf{m}^{\ell+1} = \mathbf{m}^\ell$. Then add 1 to ℓ and return to step 2.

In data assimilation, the target is the posterior PDF, $p^*(\mathbf{m}) \equiv p(\mathbf{m}|\mathbf{d}_{\text{obs}})$. Therefore, evaluation of $p^*(\mathbf{m})$ requires the evaluation of the forward model, which can render the sampling process computationally prohibitive unless (1) the forward model is extremely fast, or (2) there is very few model parameters and/or few data points.

In practical reservoir data assimilation scenarios, MCMC techniques are frequently employed in conjunction with the substitution of the original forward model with a surrogate (proxy) model. In such cases, the quality of the sampling hinges on the surrogate's capacity to replicate the nonlinearity inherent in the original forward model. However, the generation of reliable proxies can be very challenging for high-dimensional problems¹⁴.

2.4.4 Randomized Maximum Likelihood

RML was independently introduced by Kitanidis [239] and Oliver et al. [332]¹⁵ as an approximate sampling method for nonlinear problems. Orig-

¹⁴ Chapter 3, Section 3.5 discusses the use of proxies in the context of reservoir data assimilation.

¹⁵ In a more recent publication, Bardsley et al. [28] proposed the “Randomize-Then-Optimize” method for sampling posterior distributions. This method is essentially identical to RML.

inally, Oliver et al. [332] introduced RML as a proposal mechanism for MCMC, where each proposed state is derived from an optimization problem. However, given the complexities associated with evaluating $q(\cdot, \cdot)$ and the consistently high acceptance rates observed, the authors subsequently suggested accepting all states that demonstrate proper convergence during optimization. This approach has since evolved into the current RML method, which serves as an approximate strategy for sampling from the posterior distribution¹⁶.

Pseudo-code 2.2 summarizes the RML process. Each sample obtained through RML requires the solution of an optimization problem. In theory, any optimization technique can be employed, provided it converges to a minimum of $\mathcal{O}_r(\mathbf{m})$. However, practical experience suggests that an efficient implementation of RML requires highly efficient optimization methods, such as the ones based on derivatives (Chapter 3, Section 3.2).

Upon inspecting $\mathcal{O}_r(\mathbf{m})$, it becomes evident that each RML sample tends to remain close to the prior realization unless the likelihood demands significant adjustments to the parameter values. This characteristic stands out as a distinctive advantage of the method, particularly when the prior realizations are derived from a rigorous ge modeling process. In this sense, the objective is to make the smallest changes in the model to honor the observations [328].

Pseudo-code 2.2: RML

1. Sample $\mathbf{m}_j \sim \mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_{\mathbf{m}})$.
2. Sample $\mathbf{d}_{\text{obs},j} \sim \mathcal{N}(\mathbf{d}_{\text{obs}}, \mathbf{C}_{\mathbf{e}})$.
3. Compute

$$\mathbf{m}_{c,j} = \arg \min_{\mathbf{m}} \mathcal{O}_r(\mathbf{m}), \quad (2.107)$$

where

$$\begin{aligned} \mathcal{O}_r(\mathbf{m}) = & \frac{1}{2} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m})) + \\ & \frac{1}{2} (\mathbf{m} - \mathbf{m}_j)^\top \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_j). \end{aligned} \quad (2.108)$$

2.4.5 Examples

2.4.5.1 Example: Linear Model

Consider a linear-Gaussian problem with three measurements and $N_m = 20$ model parameters. Fig. 2.10a shows five prior samples along with the three

¹⁶ For linear-Gaussian problems, RML results in a correct sampling (as discussed in Section 2.4.1.4).

data points. A sampling from the posterior was performed using Metropolis-Hasting MCMC with the prior PDF as proposal distribution. Out of a total of 50,000 proposals and a burn-in period of 1,000 proposals, only 66 samples were accepted, resulting in an acceptance rate of merely 0.13%. Fig. 2.10b shows five of these posterior samples. While there exist more efficient MCMC implementations and proposal distributions, this example illustrates the curse of dimensionality problem faced in sampling high-dimensional PDFs.

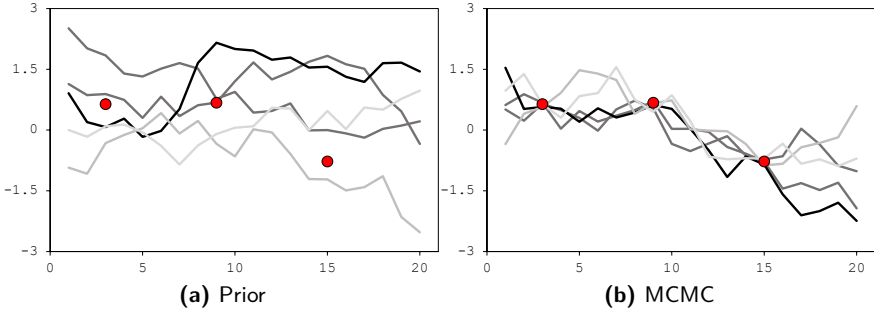


Fig. 2.10: Five prior and MCMC samples for a simple linear problem. The red circles are the observations and the lines are samples. Example. 2.4.5.1.

Using MCMC to sample the posterior can be computationally very expensive. However, sampling the prior tends to be relatively straightforward. This observation motivates the adoption of a procedure that goes beyond mere acceptance or rejection of proposals; instead, it corrects the proposals. This rationale underlies the concept of RML. Fig. 2.11a shows one realization from the prior and a hypothetical naive correction to this sample in order to match the observations. While this naive correction achieves the goal of matching the data, it compromises the notion of model plausibility. Furthermore, it offers no enhancements in predicting values at other locations. In contrast, RML produces smooth corrections to the prior realization that are consistent with the expected distribution of the property (Fig. 2.11b)

2.4.5.2 Example: Nonlinear Univariate Problem

Consider a single-parameter model with time evolution defined as

$$g(m, t) = 1 - \frac{9}{2} \left(m - \frac{2\pi}{3} \right)^2 + (t - 1) \sin(m). \quad (2.109)$$

This problem is an extension of the synthetic problem tested in [332]. Assume a Gaussian prior PDF $\mathcal{N}(2.4, 0.1)$ and five measurements corrupted with random error $e \sim \mathcal{N}(0, 0.01)$. The posterior PDF has the form

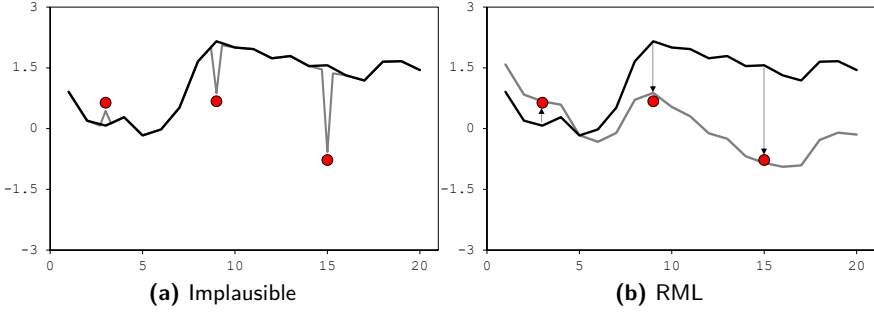


Fig. 2.11: Implausible and RML corrections to a prior sample for a simple linear problem. The red circles are the observations. The black line is the prior and the gray line is the corresponding realization after corrections. Example. 2.4.5.1.

$$p(m|\mathbf{d}_{\text{obs}}) = b \exp \left(-\frac{1}{2} \frac{(m^\ell - 2.4)^2}{0.1} - \frac{1}{2} \sum_{t=1}^{N_t} \frac{(d_{\text{obs},t} - g(m^\ell, t))^2}{0.01} \right), \quad (2.110)$$

where N_t is the number of time steps.

Fig. 2.12 presents histograms of the model parameters obtained through the sampling of both the prior PDF and the posterior using MCMC and RML. Each plot also includes the actual posterior distribution for reference. It is worth noting that the posterior exhibits a bi-modal behavior, particularly pronounced after the assimilation of the first datum. As more data points are incorporated, one of the modes becomes more prominent. Just as a reference, the true value of m is 1.884, which is closer to the first mode of the posterior. The prior distribution was intentionally chosen with a mean closer to the second mode of the posterior.

MCMC utilized 50,000 proposals from the prior, resulting in a near-perfect sampling of the PDFs. The acceptance rates for the chains were 0.274 and 0.06 for the posteriors with one and five data points, respectively. In contrast, for RML, we obtained the samples by generating 3,000 realizations while minimizing the objective function (2.108). In this example, the minimization problem was solved using Brent's method [350, Chap. 8]. For the first datum, RML was successful in capturing both modes of the posterior distribution, although not in the exact proportions. Notably, the region between the modes was under-sampled. After the fifth datum, RML over-sampled the mode of the distribution closer to the prior mean.

2.4.5.3 Example: Permeability in a Two-Phase Linear Flow

Emerick and Reynolds [131] proposed a benchmark problem to compare the sampling performance of ensemble-based methods. This problem entails es-

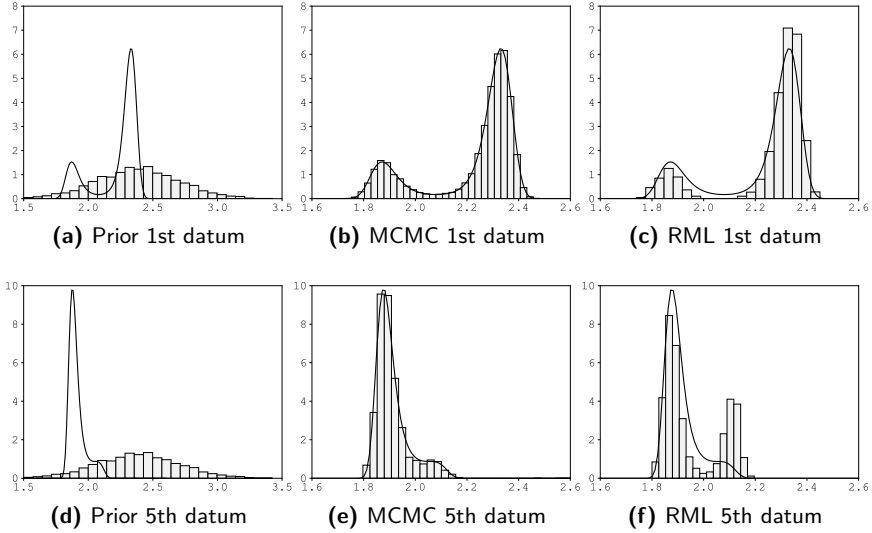


Fig. 2.12: Prior, MCMC and RML sampling for a nonlinear univariate problem. The black curve in each plot corresponds to the correct posterior distribution. Example 2.4.5.2.

timating the marginal log-permeability distribution within a small reservoir model subject to waterflooding, with pressure measurements as observed data. For reference, a Metropolis-Hastings MCMC algorithm was employed with 20 million proposals to sample the posterior PDF. In this case, the proposal distribution was centered around a small Gaussian perturbation applied to the chain's current state. Each proposal was computed as follows:

$$\hat{\mathbf{m}}^{\ell+1} = \mathbf{m}^{\ell} + \mathbf{C}_{\mathbf{m}}^{1/2} \delta \mathbf{z}, \quad (2.111)$$

where $\delta \mathbf{z} \sim N(\mathbf{0}, \gamma^2 \mathbf{I})$, with $\gamma = 0.05$ representing the scaling factor. This proposal scheme resulted in an acceptance rate of 23%.

Fig. 2.13 presents the percentiles (p2, p25, p50, p75, and p98) of the resulting marginal posterior distributions of log-permeability obtained with MCMC and RML. Fig. 2.14 shows the corresponding distributions obtained for a forecasted water production rate. For RML, the optimizations were solved using a quasi-Newton method with adjoint gradients to generate 100 conditional realizations; more details can be found in the original publication. These figures show that MCMC and RML yield remarkably similar samplings of the posterior PDF.

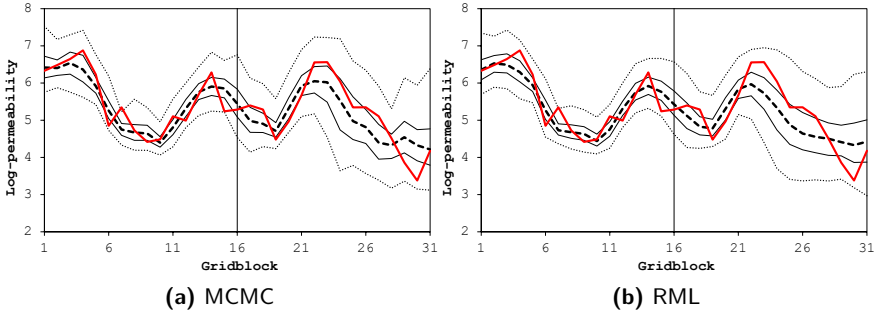


Fig. 2.13: Marginal log-permeability distributions for a two-phase linear flow. The red line represents the ground truth, the dashed line denotes the median, and the solid black lines correspond to the 25th and 75th percentiles. The outer bounding lines encompass the 2nd and 98th percentiles. Example 2.4.5.3. Reproduced from Emerick and Reynolds [131] with permission from Springer Nature.

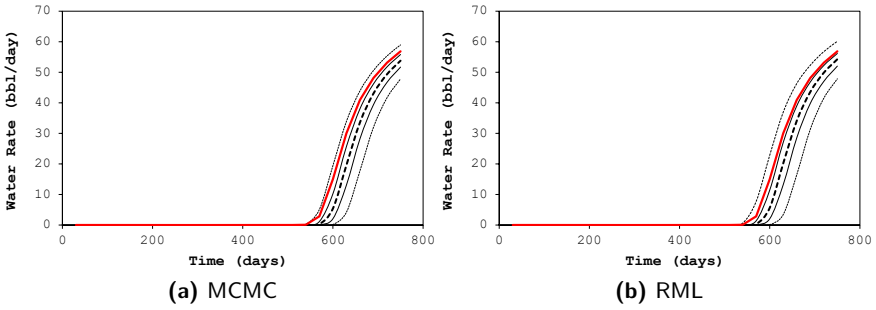


Fig. 2.14: Predicted water production rate. The red line represents the ground truth, the dashed line denotes the median, and the solid black lines correspond to the 25th and 75th percentiles. The outer bounding lines encompass the 2nd and 98th percentiles. Example 2.4.5.3. Reproduced from Emerick and Reynolds [131] with permission from Springer Nature.

2.4.5.4 Example: Two-Phase Flow in 2D Model

Consider a two-phase flow (oil and water) in a small 2D model discretized into 15×30 gridblocks. In this model, there are six producers and two water injection wells arranged in five-spot patterns, as illustrated in Fig. 2.15. The ground truth of log-permeability and the prior realizations were generated by sampling a multivariate Gaussian distribution with a constant mean of $5 \ln\text{-mD}$ and a constant variance of $1 (\ln\text{-mD})^2$ using the Cholesky decomposition method. The prior covariance was constructed using an exponential covariance function with ranges corresponding to 60 gridblocks in the major direction and five gridblocks in the minor direction, rotated by

an angle of 45° . The objective of the long correlation is to mimic a system of high-permeability channels.

The experiment involves the assimilation of oil and water production data for a period of 3900 days, with measurements every 150 days. The synthetic observations were obtained perturbing the data predicted by the model with the true log-permeability distribution (Fig. 2.15a) with random uncorrelated noise with standard deviation corresponding to 5% of the data values.

Starting from a constant prior mean, we obtained the MAP estimate by minimizing Eq. (2.28) using a quasi-Newton method and adjoint gradients. This process successfully recovered the high-permeability streak connecting one injector to two producers in the upper part of the model, as indicated in Fig. 2.15b. Figs. 2.15c–2.15e show three prior realizations while Figs. 2.15f–2.15h show the corresponding posteriors obtained with RML. The posterior realizations exhibit the expected structure with long-range permeability corridors.

Fig. 2.16 shows the predicted water production rate for three wells in the model, including a forecast period with an additional 3900 days of production. Each plot displays the observed data points, the predicted data obtained with the true log-permeability distribution, the MAP estimate, 100 prior realizations, and the corresponding 100 posterior realizations obtained with RML. Notably, there is significant variability in the predicted water breakthrough with the prior realizations. This occurs because these realizations were generated without any conditioning data at well locations, resulting in high-permeability channels appearing in different locations. This feature makes the data assimilation problem highly nonlinear despite the model's simplicity. It is also noteworthy that the predicted water production obtained with the MAP estimate deviates significantly from the ground truth during the forecast period, especially for the first well. However, the range of predictions from the posterior realizations obtained with RML encompasses the true prediction.

2.5 Normalized Objective Function

A recurring question posed by data assimilation practitioners pertains to the criteria for determining the success of a data assimilation procedure. Evidently, the response to this question encompasses various facets of the process, including considerations related to model plausibility, proper quantification of uncertainty, and the quality of data matches. In fact, defining strict criteria for acceptance of a data assimilation study can be challenging (and often controversial). Experienced practitioners consistently emphasize the importance of aligning the study's objectives with the criteria for evaluating its acceptability.

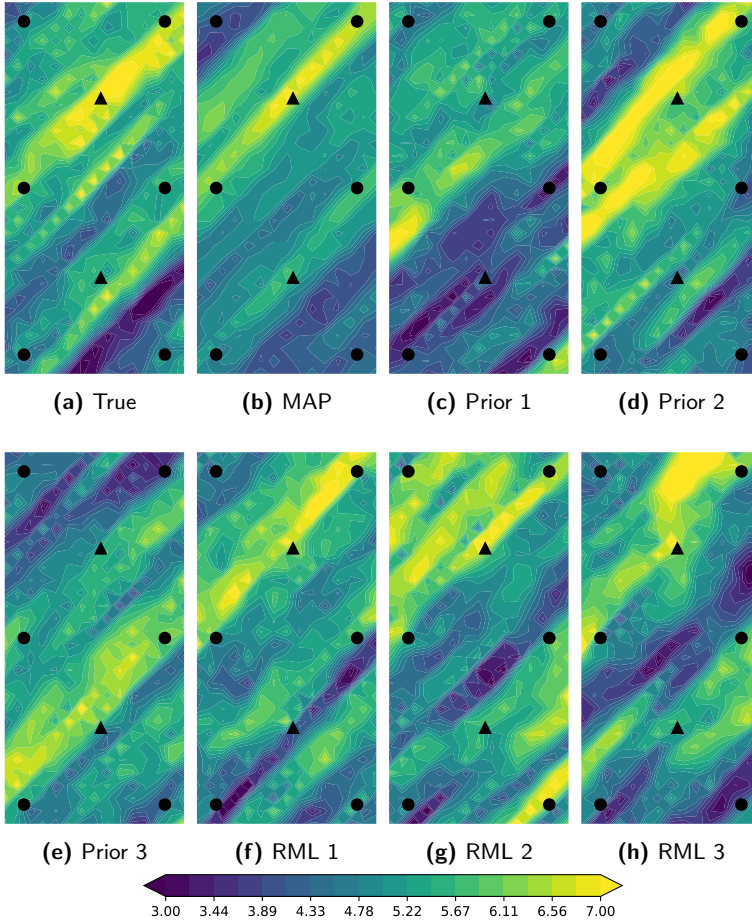


Fig. 2.15: True, MAP, and three prior and posterior realizations of log-permeability for a 2D model under two-phase flow. The posterior realizations were obtained with RML. Black circles indicate the position of producers and black triangles indicate the position of water injection wells. Example 2.4.5.4.

Among these aspects, the quality of data matches is perhaps the easiest to quantify, as it directly relates to the definition of the objective function minimized during data assimilation. However, a crucial question arises: What should be the acceptable value of the objective function following a successful data assimilation? Oliver et al. [334] observe that in the linear-Gaussian case, the value of $2\mathcal{O}(\mathbf{m})$ when evaluated at the MAP estimate follows a χ^2 distribution with N_d degrees of freedom [428]. For large N_d , the χ^2 distribution can be approximated by a Gaussian distribution with a mean of N_d and a variance of $2N_d$. Consequently, it is convenient to define a normalized objective function as

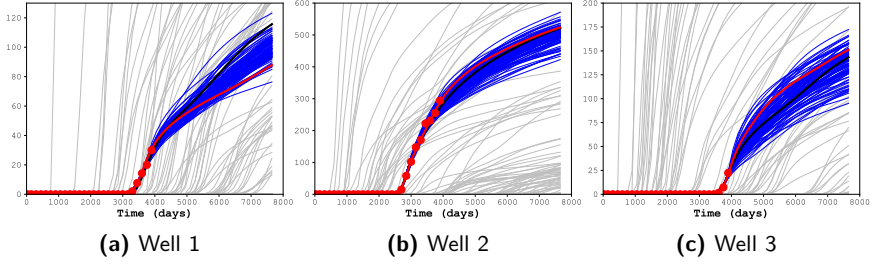


Fig. 2.16: Water production rate in bbl/day for three wells. Red circles are the observations, red line the prediction with the true log-permeability distribution, black line the prediction from the MAP estimate. The gray and blue lines show the predictions from 100 prior and posterior realizations, respectively. The posterior realizations were obtained with RML. Example 2.4.5.4.

$$\mathcal{O}_N(\mathbf{m}) = \frac{\mathcal{O}(\mathbf{m}_c)}{N_d}. \quad (2.112)$$

Based on the arguments presented in [334], the expected value of $\mathcal{O}_N(\mathbf{m}_{\text{map}})$ should be $1/2$. For samples of the posterior, Oliver et al. [334] drop the multiplication by two and defined the following acceptance criteria

$$\mathcal{O}_N(\mathbf{m}_{c,j}) = \frac{\mathcal{O}(\mathbf{m}_{c,j})}{N_d} \leq 1 + 5\sqrt{\frac{2}{N_d}}. \quad (2.113)$$

The reason for dropping the multiplication by two comes from the fact that when sampling the posterior with RML, we replace $\mathbf{m}_{\text{pr}}$ and $\mathbf{d}_{\text{obs}}$ by $\mathbf{m}_j$ and $\mathbf{d}_{\text{obs},j}$ [334].

The criterion of Eq. (2.113) is strictly valid only for linear problems, but Oliver et al. [334] argues that it should approximately hold for nonlinear problems. However, for practical data assimilation applications, the computation of $\mathcal{O}(\mathbf{m})$ is expensive because it includes the inverse of $\mathbf{C}_m$. In these cases, it is common to consider only the data mismatch part of the objective function with the justification that this part typically dominates the magnitude of the total objective function. Hence, we define the normalized data-mismatch objective function

$$\mathcal{O}_{N,d}(\mathbf{m}) \equiv \frac{\mathcal{O}_d(\mathbf{m})}{N_d}. \quad (2.114)$$

However, even considering only the data mismatch, the experience with real-life applications indicates that the criterion of Eq. (2.113) is rarely satisfied. This level of agreement is typically only observed in controlled synthetic problems, where model and data uncertainties are fully characterized and

there are no relevant model errors. Nevertheless, this criterion serves as a reference.

Perhaps one useful interpretation for the values of $\mathcal{O}_{N,d}(\mathbf{m})$ is how large is the average mismatch compared to the data error standard deviation. If we neglect the off-diagonal terms of $\mathbf{C}_e$, Eq. (2.114) reduces to

$$\mathcal{O}_{N,d}(\mathbf{m}) = \frac{1}{2N_d} \sum_{i=1}^{N_d} \left(\frac{d_{\text{obs},i} - g_i(\mathbf{m})}{\sigma_{e,i}} \right)^2. \quad (2.115)$$

Therefore, if the absolute difference $|d_{\text{obs},i} - g_i(\mathbf{m})|$ in Eq. (2.115) is exactly equals to $\sigma_{e,i}$ for all data points, we obtain $\mathcal{O}_{N,d}(\mathbf{m}) = 0.5$. Similarly, allowing mismatches of $2\sigma_{e,i}$ or $3\sigma_{e,i}$ correspond to $\mathcal{O}_{N,d}(\mathbf{m}) = 2$ and $\mathcal{O}_{N,d}(\mathbf{m}) = 4.5$, respectively, as illustrated in Fig. 2.17. These numbers serve as a reference for analyzing the results of the data assimilation. For example, $\mathcal{O}_{N,d}(\mathbf{m}) > 4.5$ may indicate that the validity of the model should be reconsidered. It also may indicate that the data-error covariance has been underestimated. Note that in the definition of $\mathbf{C}_e$ we should consider both measurement and model errors.

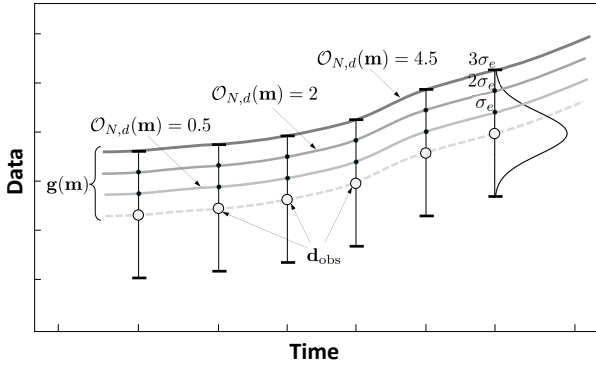


Fig. 2.17: Hypothetical model predictions with corresponding values of normalized data mismatch objective function.

This page has no content, but it has plenty of potential.

Optimization Methods for Data Assimilation

Abstract: *This chapter provides a concise overview of key optimization methods employed in reservoir data assimilation. The goal is to introduce these techniques and guide readers toward more detailed references for further study. The chapter concludes with a discussion of parametrization strategies commonly used in reservoir data assimilation.*

3.1 Introduction

A substantial portion of the literature on reservoir data assimilation is devoted to proposing efficient optimization strategies to minimize the distance between observed and predicted data vectors. The range of methods explored in this context is vast and diverse.

This book adopts the fundamental principle that reservoir data assimilation is better described as a sampling problem rather than an optimization problem. According to this principle, the goal is to explore the uncertainty space while searching for solutions (samples) that are consistent with geological information and capable of reproducing the observations within their confidence levels. Even in this context, however, optimization methods can be useful as components for building efficient sampling strategies, with the most prominent example being the RML method. In fact, as we will discuss in the following chapters, even ensemble methods, particularly the iterative forms of the ensemble smoother, have roots grounded in optimization.

This chapter presents a concise (and superficial) review of some optimization methods used in reservoir data assimilation. The objective is to provide an overview of these methods and direct readers to relevant references. The final section of this chapter discusses some parametrization strategies used in reservoir data assimilation.

3.2 Derivative-Based Methods

Derivative-based optimization methods were among the earliest techniques to emerge in the data assimilation literature in the context of petroleum reservoirs [331] and groundwater hydrology [473]. Most current developments in this area are grounded in the foundational works of Gavalas et al. [164] and Shah et al. [390]. These methods typically exhibit high convergence rates. In practical applications, however, the crux of the problem lies in computing the derivatives, as this factor heavily influences the overall performance of the process.

Pseudo-code 3.1 presents the general iterative process for minimizing $\mathcal{O}(\mathbf{m})$. Each iteration of this process requires the selection of a downhill direction followed by the solution of a line-search problem¹. Fig. 3.1 illustrates the process described in Pseudo-code 3.1.

Pseudo-code 3.1: Derivative-Based Methods

1. At the ℓ th iteration, find a downhill direction, $\mathbf{s}^\ell$.
2. Find an approximate solution for the line-search problem, i.e., find β_ℓ such that

$$\beta_\ell \approx \arg \min_{\beta} \mathcal{O}(\mathbf{m}^\ell + \beta \mathbf{s}^\ell). \quad (3.1)$$

3. Update the model estimate using

$$\mathbf{m}^{\ell+1} = \mathbf{m}^\ell + \beta_\ell \mathbf{s}^\ell = \mathbf{m}^\ell + \delta \mathbf{m}^{\ell+1}. \quad (3.2)$$

3.2.1 Steepest Descent Method

The direction with the fastest decrease in $\mathcal{O}(\mathbf{m}^\ell)$ is the opposite of the gradient of $\mathcal{O}(\mathbf{m}^\ell)$. Therefore, the steepest descent method corresponds to set the search direction $\mathbf{s}^\ell$ in Pseudo-code 3.1 as

$$\mathbf{s}^\ell = -\nabla \mathcal{O}(\mathbf{m}^\ell). \quad (3.3)$$

Note, however, that the name “steepest descent direction” does not mean that $\mathbf{s}^\ell = -\nabla \mathcal{O}(\mathbf{m}^\ell)$ is the direction which will necessarily lead to the minimum of $\mathcal{O}(\mathbf{m})$ with the smallest number of iterations. Actually, any direction

¹ Derivative-based optimization strategies are typically divided into line-search and trust-region methods. In this book, we limit our discussion to optimization schemes based on line search. There are several excellent references on derivative-based optimization methods. For example, Nocedal and Wright [325] presents a didactic discussion of both line-search and trust-region methods.

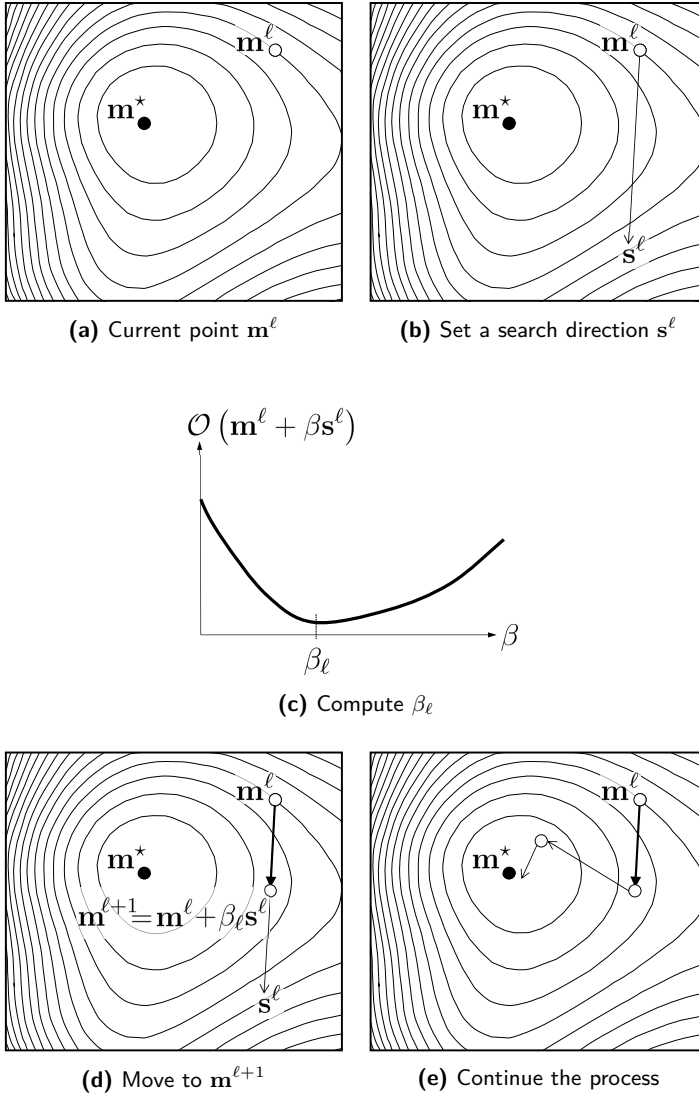


Fig. 3.1: Illustration of the general derivative-based optimization process based on line-search.

$\mathbf{s}^\ell$ such that $[-\nabla\mathcal{O}(\mathbf{m}^\ell)]^\top \mathbf{s}^\ell > 0$ is downhill. Note that

$$[-\nabla\mathcal{O}(\mathbf{m}^\ell)]^\top \mathbf{s}^\ell = \|-\nabla\mathcal{O}(\mathbf{m}^\ell)\| \|\mathbf{s}^\ell\| \cos \theta, \quad (3.4)$$

where θ is the angle between $-\nabla\mathcal{O}(\mathbf{m}^\ell)$ and $\mathbf{s}^\ell$. Therefore, any $\mathbf{s}^\ell$ such that $-\pi/2 < \theta < \pi/2$ is a downhill direction. Fig. 3.2a illustrates the steepest descent direction pointing in the opposite direction to the gradient of $\mathcal{O}(\mathbf{m}^\ell)$. Note that the alternate direction $\mathbf{s}^\ell$ in Fig. 3.2b is also downhill and points directly towards the optimum $\mathbf{m}^*$. In this case, $\mathbf{s}^\ell$ would require a single iteration to converge. In fact, for poorly scaled problems, the steepest descent method may present a poor performance, requiring several iterations to converge to an optimum.

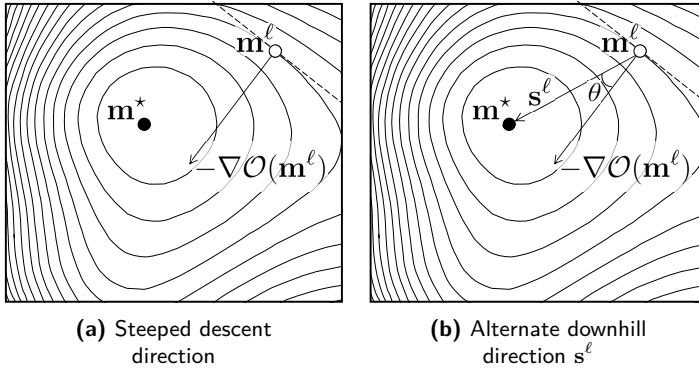


Fig. 3.2: Illustration of the steepest descent search direction.

3.2.2 Nonlinear Conjugate Gradient Method

The conjugate gradient method is an iterative technique for solving symmetric systems of linear equations or, equivalently, for minimizing convex quadratic functions. This method was extended to minimize general nonlinear functions by Fletcher and Reeves [149]. Pseudo-code 3.2 presents a standard version of the nonlinear conjugate gradient (NCG). The minimization process is similar to the steepest descent method; in fact, the first iteration of both methods is identical. However, unlike the steepest descent, NCG uses information from the previous iteration's search direction to compute the current search direction. Pseudo-code 3.2 utilizes the formula proposed in [149] to update the search direction, but other formulas also exist; see Nocedal and Wright [325] for further details. Similar to the steepest descent

method, NCG requires only the evaluation of the objective function and its gradient without involving any matrix operations. This makes the NCG method particularly attractive for large-scale problems.

Pseudo-code 3.2: Nonlinear Conjugate Gradient Method

1. Set $\ell = 0$ and initialize with the steepest descent search direction, $\mathbf{s}^0 = -\nabla \mathcal{O}(\mathbf{m}^0)$.
2. Find an approximate solution for the line-search problem

$$\beta_\ell \approx \arg \min_{\beta} \mathcal{O}(\mathbf{m}^\ell + \beta \mathbf{s}^\ell). \quad (3.5)$$

3. Update the model estimate

$$\mathbf{m}^{\ell+1} = \mathbf{m}^\ell + \beta_\ell \mathbf{s}^\ell. \quad (3.6)$$

4. Check the termination criteria.
5. If not terminated then:

- Compute

$$\gamma_{\ell+1} = \frac{[\nabla \mathcal{O}(\mathbf{m}^{\ell+1})]^\top \nabla \mathcal{O}(\mathbf{m}^{\ell+1})}{[\nabla \mathcal{O}(\mathbf{m}^\ell)]^\top \nabla \mathcal{O}(\mathbf{m}^\ell)}. \quad (3.7)$$

- If $\gamma_{\ell+1} < \epsilon$, set $\gamma_{\ell+1} = 0$ (typically $\epsilon \leq 0.1$).
- Update the search direction

$$\mathbf{s}^{\ell+1} = -\nabla \mathcal{O}(\mathbf{m}^{\ell+1}) + \gamma_{\ell+1} \mathbf{s}^\ell. \quad (3.8)$$

- Set $\ell = \ell + 1$ and return to step 2.

3.2.3 Newton's Method

Newton's method define the search direction by computing a quadratic function around the current estimate $\mathbf{m}^\ell$ as

$$\begin{aligned} \mathcal{Q}(\mathbf{m}) &= \mathcal{O}(\mathbf{m}^\ell) + (\nabla \mathcal{O}(\mathbf{m}^\ell))^\top (\mathbf{m} - \mathbf{m}^\ell) + \\ &\quad \frac{1}{2} (\mathbf{m} - \mathbf{m}^\ell)^\top \mathbf{B}_\ell (\mathbf{m} - \mathbf{m}^\ell). \end{aligned} \quad (3.9)$$

$\mathcal{Q}(\mathbf{m})$ has a unique minimum if the Hessian matrix $\mathbf{B}_\ell = \nabla [(\nabla \mathcal{O}(\mathbf{m}^\ell))^\top]$ is positive definite. In this case, to find the minimum of the quadratic, it suffices to compute the gradient $\nabla \mathcal{Q}(\mathbf{m})$ and set the resulting expression to zero:

$$\nabla \mathcal{Q}(\mathbf{m}) = \nabla \mathcal{O}(\mathbf{m}^\ell) + \mathbf{B}_\ell (\mathbf{m} - \mathbf{m}^\ell) = \mathbf{0}. \quad (3.10)$$

Solving for $(\mathbf{m} - \mathbf{m}^\ell)$ and denoting the result as $\delta \mathbf{m}^{\ell+1}$ leads to the Newton's search direction:

$$\mathbf{s}^\ell = \delta \mathbf{m}^{\ell+1} = -\mathbf{B}_\ell^{-1} \nabla \mathcal{O}(\mathbf{m}^\ell). \quad (3.11)$$

The original Newton's method uses a full step ($\beta_\ell = 1$), in which case there is no need for a line search. This condition is required for proving the quadratic convergence of the method [325]. In practice, however, it may be beneficial to restrict the step size, especially in the early iterations [334].

3.2.4 Gauss-Newton

Newton's method requires second-order derivatives to compute the Hessian. The Gauss-Newton (GN) method relaxes this requirement for least-squares problems by approximating the Hessian using only first-order derivatives.

Recall the objective function for finding the MAP²

$$\begin{aligned} \mathcal{O}(\mathbf{m}) = & \frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})) + \\ & \frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}). \end{aligned} \quad (3.12)$$

The gradient of $\mathcal{O}(\mathbf{m})$ has the form

$$\nabla \mathcal{O}(\mathbf{m}) = \mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}) + \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}), \quad (3.13)$$

and the Hessian of $\mathcal{O}(\mathbf{m})$ is

$$\begin{aligned} \mathbf{B} = & \nabla \left[(\nabla \mathcal{O}(\mathbf{m}))^\top \right] \\ = & \mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G} + \nabla \mathbf{G}^\top \mathbf{C}_e^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}) + \mathbf{C}_m^{-1}. \end{aligned} \quad (3.14)$$

$\mathbf{G}$ is the sensitivity (Jacobian) matrix, which contains the partial derivatives of data with respect to parameters

$$G_{ij} = \frac{\partial g_i(\mathbf{m})}{\partial m_j}, \quad \text{for } i = 1, \dots, N_d \quad \text{and} \quad j = 1, \dots, N_m. \quad (3.15)$$

² Here, we present the equations for finding the MAP. However, the equations are essentially the same for finding posterior realizations with RML, in which case it suffices to replace $\mathbf{m}_{\text{pr}}$ by $\mathbf{m}_j$ and $\mathbf{d}_{\text{obs}}$ by $\mathbf{d}_{\text{obs},j}$.

The term involving $\nabla \mathbf{G}^\top$ in Eq. (3.14) requires to compute the second partial derivatives. Neglecting these terms gives the GN Hessian:

$$\mathbf{B}_{\text{GN}} = \mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G} + \mathbf{C}_m^{-1}. \quad (3.16)$$

The GN Hessian is positive definite. This means that the GN search direction is always downhill, whereas this may not be true for Newton's method. The GN search direction for computation of the MAP estimate at the ℓ th iteration becomes

$$\begin{aligned} \mathbf{s}^\ell &= -\mathbf{B}_{\text{GN},\ell}^{-1} \nabla \mathcal{O}(\mathbf{m}^\ell) \\ &= -(\mathbf{G}_\ell^\top \mathbf{C}_e^{-1} \mathbf{G}_\ell + \mathbf{C}_m^{-1})^{-1} \times \\ &\quad [\mathbf{G}_\ell^\top \mathbf{C}_e^{-1} (\mathbf{g}(\mathbf{m}^\ell) - \mathbf{d}_{\text{obs}}) + \mathbf{C}_m^{-1} (\mathbf{m}^\ell - \mathbf{m}_{\text{pr}})]. \end{aligned} \quad (3.17)$$

Eq. (3.17) involves the inverse of $N_m \times N_m$ matrices. For $N_d \ll N_m$, it is convenient to use the Sherman-Morrison-Woodbury formula (A.9) to rewrite the search direction as

$$\begin{aligned} \mathbf{s}^\ell &= \mathbf{m}_{\text{pr}} - \mathbf{m}^\ell - \mathbf{C}_m \mathbf{G}_\ell^\top (\mathbf{C}_e + \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top)^{-1} \times \\ &\quad [\mathbf{g}(\mathbf{m}^\ell) - \mathbf{d}_{\text{obs}} - \mathbf{G}_\ell (\mathbf{m}^\ell - \mathbf{m}_{\text{pr}})]. \end{aligned} \quad (3.18)$$

If both N_m and N_d are large, the GN method may not be computationally feasible. In such case, we may use the steepest descent, NCG, or quasi-Newton methods.

3.2.5 Levenberg-Marquardt

The GN method tends to be very efficient when the objective function is well approximated by a quadratic. However, especially at the beginning of the minimization process, when the estimate is far from the optimum, a quadratic can result in a poor estimate of the objective function. In these cases, the steepest descent method can move faster toward the solution than GN. Fortunately, the Levenberg-Marquardt (LM) algorithm incorporates characteristics of both methods: it behaves like the steepest descent when the objective function is more complex than a quadratic and inherits the good convergence rate of GN when the quadratic becomes a good approximation.

The most basic form of the LM algorithm is obtained by modifying the Hessian as follows:

$$\mathbf{B}_{\text{LM}} = \lambda \mathbf{I} + \mathbf{B}, \quad (3.19)$$

where $\lambda > 0$ is known as the LM parameter. Note that $\mathbf{B}_{\text{LM}}$ is always positive definite, even if the Hessian is not. Typically, we use the GN Hessian, in which case the LM parameter has a favorable effect on improving the condition number of the matrix. LM does not require a line search to define

the step size. Instead, λ controls both the search direction and the size of the update. The basic LM process is summarized in Pseudo-code 3.3.

Pseudo-code 3.3: Levenberg-Marquardt

1. Set $\ell = 0$ and choose λ_0 (typically $\lambda_0 = \mathcal{O}(\mathbf{m}^0)/N_d$).
2. Solve

$$(\lambda_\ell \mathbf{I} + \mathbf{B}_\ell) \delta \mathbf{m}^{\ell+1} = -\nabla \mathcal{O}(\mathbf{m}^\ell), \quad (3.20)$$

and set

$$\mathbf{m}^{\ell+1} = \mathbf{m}^\ell + \delta \mathbf{m}^{\ell+1}. \quad (3.21)$$

3. If $\mathcal{O}(\mathbf{m}^{\ell+1}) \geq \mathcal{O}(\mathbf{m}^\ell)$ then

$$\lambda_\ell = \gamma \lambda_\ell, \quad \text{where } \gamma > 1 \quad (\text{typically } 4 \leq \gamma \leq 10) \quad (3.22)$$

and redo step 2.

Else check the termination criteria.

If not terminated then

$$\lambda_{\ell+1} = \lambda_\ell / \gamma, \quad \text{where } \gamma > 1. \quad (3.23)$$

Set $\ell = \ell + 1$ and return to step 2.

3.2.6 Quasi-Newton

For large-scale problems, the construction of the Hessian may not be feasible. Even the computation of individual components of $\mathbf{G}$ may become problematic. In these cases, one alternative is the use of quasi-Newton methods. These methods are based on recursive formulas to construct an estimate of $\mathbf{B}^{-1}$ during the iterations of the optimization.

The basic condition for quasi-Newton methods is that

$$\tilde{\mathbf{B}}_{\ell+1}^{-1} [\nabla \mathcal{O}(\mathbf{m}^{\ell+1}) - \nabla \mathcal{O}(\mathbf{m}^\ell)] = \mathbf{m}^{\ell+1} - \mathbf{m}^\ell. \quad (3.24)$$

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [325, 350] stands out as one of the most efficient quasi-Newton methods. BFGS uses information on the descent directions and gradients at previous iterations to construct an approximate inverse Hessian, $\tilde{\mathbf{B}}^{-1}$. In the pure quadratic case, the BFGS algorithm can eventually converge to the true inverse Hessian [325].

Calling $\mathbf{p}_\ell = \nabla \mathcal{O}(\mathbf{m}^{\ell+1}) - \nabla \mathcal{O}(\mathbf{m}^\ell)$ and $\mathbf{q}_\ell = \mathbf{m}^{\ell+1} - \mathbf{m}^\ell$ in Eq. (3.24) leads to

$$\tilde{\mathbf{B}}_{\ell+1}^{-1} \mathbf{p}_\ell = \mathbf{q}_\ell. \quad (3.25)$$

The BFGS updates the approximation of the inverse Hessian using

$$\tilde{\mathbf{B}}_{\ell+1}^{-1} = \mathbf{Q}_\ell^\top \tilde{\mathbf{B}}_\ell^{-1} \mathbf{Q}_\ell + v_\ell \mathbf{q}_\ell \mathbf{q}_\ell^\top, \quad (3.26)$$

where $v_\ell = 1/(\mathbf{p}_\ell^\top \mathbf{q}_\ell)$ and $\mathbf{Q}_\ell = \mathbf{I} - v_\ell \mathbf{p}_\ell \mathbf{q}_\ell^\top$. The theory behind BFGS is beyond the scope of this book³. Pseudo-code 3.4 summarizes the BFGS method.

Pseudo-code 3.4: BFGS

1. Set $\ell = 0$. Choose an initial symmetric positive definite inverse Hessian, $\tilde{\mathbf{B}}_0^{-1}$.
2. Compute the search direction $\mathbf{s}^\ell = -\tilde{\mathbf{B}}_\ell^{-1} \nabla \mathcal{O}(\mathbf{m}^\ell)$.
3. Find an approximate solution for the line-search problem

$$\beta_\ell \approx \arg \min_{\beta} \mathcal{O}(\mathbf{m}^\ell + \beta \mathbf{s}^\ell). \quad (3.27)$$

4. Update the estimate

$$\mathbf{m}^{\ell+1} = \mathbf{m}^\ell + \beta_\ell \mathbf{s}^\ell.$$

5. Check the termination criteria.
If not terminated then compute:

$$\mathbf{p}_\ell = \nabla \mathcal{O}(\mathbf{m}^{\ell+1}) - \nabla \mathcal{O}(\mathbf{m}^\ell), \quad (3.28)$$

$$\mathbf{q}_\ell = \mathbf{m}^{\ell+1} - \mathbf{m}^\ell, \quad (3.29)$$

$$\tilde{\mathbf{B}}_{\ell+1}^{-1} = \mathbf{Q}_\ell^\top \tilde{\mathbf{B}}_\ell^{-1} \mathbf{Q}_\ell + v_\ell \mathbf{q}_\ell \mathbf{q}_\ell^\top. \quad (3.30)$$

Set $\ell = \ell + 1$ and go to step 2.

The BFGS algorithm requires only the gradient of the objective function. A good initial guess for the inverse Hessian is $\tilde{\mathbf{B}}_0^{-1} = \mathbf{C}_\mathbf{m}$ whenever computing and storing $\mathbf{C}_\mathbf{m}$ is feasible. Alternatively, a reasonable choice is to initialize $\tilde{\mathbf{B}}_0^{-1} = \text{diag}(\mathbf{C}_\mathbf{m})$. However, the BFGS approximation of $\tilde{\mathbf{B}}_\ell^{-1}$ is usually dense [325], which can make manipulating this matrix prohibitive for large-scale problems. In this case, the alternative is to use the limited-memory BFGS (L-BFGS) algorithm [324], which stores only a subset of pairs of the vectors $\mathbf{p}_\ell$ and $\mathbf{q}_\ell$. Gao and Reynolds [158] presents a series of procedures to enhance the performance of an L-BFGS implementation in the context of reservoir data assimilation.

³ A good theoretical description of BFGS can be found in [325].

Remark 3.1: Termination Criteria

Oliver et al. [334] present three termination criteria for the iterative optimization process in the context of reservoir data assimilation:

1. The algorithm is terminated if

$$\frac{\|\nabla \mathcal{O}(\mathbf{m}^{\ell+1})\|}{\max\{\|\mathbf{m}^{\ell+1}\|_2, 1\}} < 10^{-8}. \quad (3.31)$$

2. If both of the following two conditions are satisfied, we also terminate the algorithm:

$$\frac{|\mathcal{O}(\mathbf{m}^\ell) - \mathcal{O}(\mathbf{m}^{\ell+1})|}{\max\{\mathcal{O}(\mathbf{m}^{\ell+1}), 1\}} < 10^{-4}, \quad (3.32)$$

and

$$\max_{1 \leq i \leq N_m} \left\{ \frac{|m_i^{\ell+1} - m_i^\ell|}{\max\{|m_i^{\ell+1}|, |\text{typ}(m_i)|\}} \right\} < 10^{-3}. \quad (3.33)$$

where $\text{typ}(m_i)$ is a typical value for m_i but nonzero.

3. We also terminate the process if the number of iterations exceeds a predefined value.

3.2.7 Scaling

The vector of model parameters may include properties with very different magnitudes. Therefore, it is good practice to re-scale $\mathbf{m}$ before applying an optimization procedure. One way to re-scale $\mathbf{m}$ is to write

$$\hat{\mathbf{m}} = \mathbf{C}_{\mathbf{m}}^{-1/2} (\mathbf{m} - \mathbf{m}_{\text{pr}}), \quad (3.34)$$

and the predicted data vector as

$$\hat{\mathbf{d}} = \mathbf{C}_{\mathbf{e}}^{-1/2} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}). \quad (3.35)$$

In this case, the objective function reduces to

$$\mathcal{O}(\hat{\mathbf{m}}) = \frac{1}{2} \hat{\mathbf{m}}^\top \hat{\mathbf{m}} + \frac{1}{2} \hat{\mathbf{d}}^\top \hat{\mathbf{d}}. \quad (3.36)$$

The gradient with respect to $\hat{\mathbf{m}}$ becomes

$$\nabla_{\hat{\mathbf{m}}} \mathcal{O}(\hat{\mathbf{m}}) = \hat{\mathbf{m}} + \mathbf{G}_d^\top \hat{\mathbf{d}}, \quad (3.37)$$

and the Hessian becomes

$$\hat{\mathbf{B}} = \nabla_{\hat{\mathbf{m}}} \left[(\nabla_{\hat{\mathbf{m}}} \mathcal{O}(\hat{\mathbf{m}}))^{\top} \right] = \mathbf{I} + \mathbf{G}_d^{\top} \mathbf{G}_d. \quad (3.38)$$

In these expressions, $\mathbf{G}_d$ is referred to as dimensionless sensitivity matrix [478], which is defined as

$$\begin{aligned} \mathbf{G}_d &= \left(\nabla_{\hat{\mathbf{m}}} \hat{\mathbf{d}}^{\top} \right)^{\top} \\ &= \left(\nabla_{\hat{\mathbf{m}}} \hat{\mathbf{d}}^{\top} \right)^{\top} (\nabla_{\hat{\mathbf{m}}} \mathbf{m}^{\top})^{\top} \\ &= \mathbf{C}_e^{-1/2} \mathbf{G} \mathbf{C}_m^{1/2}. \end{aligned} \quad (3.39)$$

3.2.8 Comments on Line Search

The line-search problem is formulated as the finding β_{ℓ} such that

$$\beta_{\ell} \approx \arg \min_{\beta} \mathcal{O}(\mathbf{m}^{\ell} + \beta \mathbf{s}^{\ell}). \quad (3.40)$$

Eq. (3.40) corresponds to a single-variable optimization problem. Numerous methods are readily available in the literature; see, e.g., [350, Chap. 10] for some classical algorithms. However, it is worth noting that each iteration of the line-search problem entails the evaluation of the forward model, and considering that the overall process may require multiple line searches, it is advisable to limit the number of steps for each line search. Remarkably, it is not necessary to employ an exact line search. In fact, there are theoretical results showing that an approximate line search suffices to establish convergence toward a local minimum (assuming it exists) [325]. Oliver et al. [334] presents a detailed discussion on quadratic and cubic line-search procedures for reservoir data assimilation.

3.2.9 Comments on Gradient and Sensitivity Calculation

Different optimization methods require different levels of derivative information. For instance, a complete implementation of Newton's method may yield a quadratic convergence rate, but it demands the computation of the Hessian matrix. In contrast, GN and LM methods alleviate this demand by only requiring the sensitivity matrix involving first-order derivatives. Steepest descent, NCG, and quasi-Newton methods, on the other hand, require only the gradient of the objective function.

The literature on reservoir data assimilation shows that various strategies for computing derivative information have been used, including finite-

difference approximation [368], the forward method [16], the adjoint method [73, 259, 367], simultaneous perturbation stochastic approximation (SPSA) [409, 159, 257], and streamline sensitivities [449, 230].

Finite difference, being the simplest approach, requires an additional execution of the forward model for each model parameter to estimate its corresponding derivative. Therefore, a problem with N_m parameters requires $N_m + 1$ evaluations of the forward model to compute the gradient vector. As a result, this method is viable only for problems featuring a limited number of parameters. The forward method computes the entire sensitivity matrix by solving an extra system of linear equations for each model parameter at each time step. The full sensitivity matrix allows the use of methods such as GN and LM, which exhibit high convergence rates, often rivaling Newton's method [334]. The primary drawback here is the added computational expense required to obtain the sensitivity matrix if the number of parameters is large.

The adjoint method provides the sensitivity matrix at the cost of solving a linear system for each data point in reverse chronological order. Typically, the number of data points is significantly less than the number of model parameters, rendering the adjoint method a preferred choice. Furthermore, the adjoint method can compute the gradient vector with the cost of solving a single linear system for each time step, regardless of the number of model parameters or data points [367]. This information suffices for using the steepest descent, NCG, or quasi-Newton methods. In fact, combining adjoint gradients with quasi-Newton methods results in a highly efficient data assimilation strategy [158]. However, unlike finite differences, the adjoint method demands substantial effort for implementation and maintenance, in addition to requiring integration within the source code of the forward model. In this context, the use of automatic differentiation (AD) strategies can be an alternative to simplify coding by eliminating the implementation of derivative calculations to construct the Jacobian matrix required in the forward simulations [260, 243]. AD can be used to calculate the partial derivatives required to construct the adjoint problem for sensitivity or gradient calculation [437].

Another gradient computation strategy explored in reservoir data assimilation literature is the SPSA method. In SPSA, the gradient vector is determined through simultaneous random perturbations of all model parameters. This approach yields a stochastic approximation of the gradient, which, under mild conditions, can be demonstrated to be downhill and have an expectation equal to the true gradient [159]. However, it is important to note that this method typically requires a higher number of iterations to converge. Nevertheless, it has the advantage of being relatively easy to implement and seamlessly integrate with complex forward models.

Streamline simulation presents another alternative for approximating sensitivities. In this approach, analytical derivatives of water-cut, gas-oil ratio, and even well-flowing pressures are derived. This method can be applied to

full-physics reservoir simulators by reconstructing the streamlines based on the flux fields stored during the simulation.

Remark 3.2: Derivative-Based Methods and Sampling

Derivative-based optimization methods are well-known for their tendency to converge toward local minima. This fact has always been a cause of concern in data assimilation applications. Typically, a premature convergence to a local minimum is quickly spotted because it corresponds to a model with poor data match. One potential remedy is to initiate the optimization from a different starting point. However, it is important to recall that the goal is to sample the posterior distribution. If we use RML, the sampling requires solving several independent minimization problems, some of which may not converge to acceptable estimates. Therefore, the use of an efficient optimization method is mandatory, in which case, the derivative-based methods become a natural choice.

3.3 Stochastic Optimization Methods

Stochastic optimization methods have been frequently used as data assimilation techniques, especially in the context of petroleum reservoirs. Among these techniques, prominent options include evolutionary algorithms [195], particle swarm optimization [234], and simulated annealing [238], among others. These methods offer some attractive features, including their ability to generate multiple solutions, robustness, simplicity of implementation, and seamless integration with the forward model, which is used as a “black box.” However, a primary drawback associated with these methods is their relatively slow convergence, which may constrain their applicability to problems featuring only a small number of uncertain parameters.

Stochastic optimization methods are well-known to be less prone to converge to local minima as they explore different parts of the search space. For this reason, these methods are sometimes used as sampling strategies, providing multiple solutions for data assimilation. Nevertheless, despite the inherently exploratory nature of stochastic methods in contrast to their derivative-based counterparts, there exists no formal assurance that they yield a judicious and unbiased sampling of the parameter space’s uncertainty. In his seminal work on Inverse Problems, Tarantola [428] aptly notes that “*genetic algorithms lack the fundamental theorem of the Metropolis algorithm: if you follow certain steps, you will produce samples from the distribution, precisely and in a technical sense.*” While Tarantola explicitly mentions genetic algorithms, the same assertion holds true for various other

stochastic optimization methods. It is worth noting, however, that none of the methods currently operational in reservoir data assimilation applications can guarantee an exact sampling, and a degree of approximation invariably persists.

3.3.1 Evolutionary Algorithms

Evolutionary algorithms (EA) are stochastic optimization methods inspired by Darwin's Evolution Theory and Natural Selection⁴ There are several variants, but the methods follow the base idea described in Pseudo-code 3.5.

Pseudo-code 3.5: Evolutionary Algorithms

1. Initialize the population.
2. Evaluate individuals.
3. While termination criteria are not met do:
 - Select individuals.
 - Recombine individuals.
 - Mutate individuals.
 - Evaluate individuals.

Genetic algorithms (GA) are among the most used variants of EA. In GA, each member of the population serves as a potential solution for the data assimilation problem. Various parametrization methods exist for individuals within GA. Traditional GA implementations employ binary arrays to represent individuals. However, floating-point representations of individuals can be used, particularly for describing real-valued parameters. Pseudo-code 3.6 provides a summarized outline of the main steps involved in employing GA to update models within a data assimilation process.

⁴ There are several good books about EA; for example, [90] is good reference.

Pseudo-code 3.6: Main Step of GA for Data Assimilation

For a predefined number of generations, repeat the following steps:

1. Selection of individuals:

- The probability of selecting an individual of the population is proportional to its fitness value, i.e.,

$$P_i = \frac{f_i}{\sum_{j=1}^{N_s} f_j}, \quad (3.41)$$

where the fitness is typically defined as $f_i = 1/\mathcal{O}_d(\mathbf{m}_i)$.

- Therefore, models with better data matches have a higher probability of being selected for the recombination.

2. Recombination:

- Individuals of the current generation (parents) are combined to generate new individuals (offsprings).
- There are several forms of recombination (crossover). For example, we can use a random linear combination of two models:

$$\hat{\mathbf{m}} = u\mathbf{m}_1 + (1 - u)\mathbf{m}_2, \quad (3.42)$$

where $u \sim \mathcal{U}[0, 1]$.

3. Mutation:

- Random perturbation in some individuals. For example, we can use a Gaussian perturbation

$$\hat{\mathbf{m}} = \mathbf{m} + \gamma \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}, \quad (3.43)$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $\gamma \in (0, 1]$ is a mutation factor.

4. Elitism:

- Preserve the best individuals of a generation to the next one (to ensure that the best-so-far solutions are not lost).

Several other algorithms belong to the EA class, among which the covariance matrix adaptation evolution strategy (CMA-ES) [186] is considered state-of-the-art [402]. However, its application to reservoir data assimilation remains limited. CMA-ES employs a mutation strategy based on Gaussian perturbations of the current solution, with a covariance matrix that is iteratively updated to adapt to the local structure of the objective function, thereby enhancing optimization efficiency.

3.3.2 Particle Swarm Optimization

Similar to GA, particle swarm optimization (PSO) [234] is a population-based method inspired by the collective behavior of animal groups, such as birds and fish. PSO operates by generating random solutions, with each solution represented as a particle forming a swarm. Within this swarm, each particle navigates the parameter space. At each step, the position and velocity of these particles are updated based on the best individual and global solutions from the previous step. This iterative process continues until the particles converge to the global optimum.

Pseudo-code 3.7 presents the basic PSO procedure. In this procedure, $\mathbf{m}_j^*$ is the previous best position for the j th particle, while $\mathbf{m}^*$ is the overall best position for all particles (sometimes computed as the best position in the neighborhood of the j th particle). The inertia parameter w controls the exploratory ability of the search: large values promote a more global search, while small values encourage local search. Typically, w is selected between 0.4 and 1. However, it may be beneficial to start with a higher value, such as 1.4, and reduce it during the iterations. The parameter c_1 defines the confidence that the particle has in its previous trajectory (the “cognitive” component) and is usually chosen between 1 and 2. The parameter c_2 defines the confidence in the swarm (the “social” component) and is also typically selected between 1 and 2. $\mathbf{R}_1$ and $\mathbf{R}_2$ are diagonal matrices with elements drawn from $\mathcal{U}[0, 1]$.

Pseudo-code 3.7: Basic PSO for Data Assimilation

1. Initialize the particles, $\{\mathbf{m}_j^0\}$, and the corresponding velocities, $\{\mathbf{v}_j^0\}$, for $j = 1, \dots, N_s$. Set $\ell = 0$ (iteration index).
2. Update the position of each particle using

$$\mathbf{m}_j^{\ell+1} = \mathbf{m}_j^\ell + \mathbf{v}_j^\ell \Delta t, \quad \text{for } j = 1, \dots, N_s \quad (3.44)$$

where $\Delta t = 1$.

3. Evaluate the objective function for each particle.
4. Update the velocity of each particle using

$$\mathbf{v}_j^{\ell+1} = w\mathbf{v}_j^\ell + \frac{c_1}{\Delta t} \mathbf{R}_1^{\ell+1} (\mathbf{m}_j^* - \mathbf{m}_j^\ell) + \frac{c_2}{\Delta t} \mathbf{R}_2^{\ell+1} (\mathbf{m}^* - \mathbf{m}_j^\ell). \quad (3.45)$$

5. Set $\ell = \ell + 1$.
6. If maximum number of iterations, then stop. Otherwise, go to step 2.

Before optimization, it is recommended to scale the parameters in the range $[-1, 1]$. Babin et al. [22] presents the following set of rules to ensure the parameter values to be within the range $[-1, 1]$ during the optimization:

- Set $m_{ij}^\ell = -2 - m_{ij}^\ell$ if $m_{ij}^\ell \leq -1$.
- Set $m_{ij}^\ell = 2 - m_{ij}^\ell$ if $m_{ij}^\ell \geq 1$.
- Set $v_{ij}^\ell = -v_{ij}^\ell$ if $|m_{ij}^\ell| \geq 1$.
- Set $v_{ij}^\ell = v_{\max}$ if $|v_{ij}^\ell| > v_{\max}$.

where m_{ij}^ℓ and v_{ij}^ℓ denote the i th component of $\mathbf{m}_j^\ell$ and $\mathbf{v}_j^\ell$, respectively. $v_{\max} \in (0, 2]$ is the maximum allowed velocity.

3.3.3 Simulated Annealing

Simulated Annealing [238] is a stochastic optimization algorithm inspired by the annealing process in metallurgy, where a material is heated and gradually cooled to reduce defects and attain a stable state. The algorithm iteratively explores the solution space by applying small random perturbations to the current solution. To avoid premature convergence, it accepts worse solutions with a probability that decreases over time. This mechanism helps the algorithm to escape local minima and enhance exploration.

The acceptance probability of a worse solution is governed by the Metropolis criterion:

$$P(\mathbf{m}, \hat{\mathbf{m}}) = \exp\left(-\frac{\Delta\mathcal{O}(\hat{\mathbf{m}})}{T}\right), \quad (3.46)$$

where $\Delta\mathcal{O}(\hat{\mathbf{m}}) = \mathcal{O}(\hat{\mathbf{m}}) - \mathcal{O}(\mathbf{m})$ denotes the change in the objective function from the proposed solution $\hat{\mathbf{m}}$ to the current solution $\mathbf{m}$. In simulated annealing, $\mathcal{O}(\hat{\mathbf{m}})$ is interpreted as the system's energy function, and T is the annealing parameter (temperature), which decreases throughout the optimization process, progressively reducing the probability of accepting worse solutions.

Pseudo-code 3.3.3 outlines the basic simulated annealing procedure for data assimilation.

Pseudo-code 3.3.3: Basic Simulated Annealing for Data Assimilation

1. Initialize the process with an initial solution $\mathbf{m}$, initial temperature T , cooling rate $\beta \in (0, 1)$, e.g., $\beta = 0.95$, and stopping temperature $T_{\min}$.
2. Keep the overall best solution $\mathbf{m}^* = \mathbf{m}$.
3. While $T > T_{\min}$:

- Propose a new solution $\hat{\mathbf{m}}$ with, for example, a Gaussian perturbation around the current solution $\mathbf{m}$:

$$\hat{\mathbf{m}} = \mathbf{m} + \gamma \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}, \quad (3.47)$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $\gamma \in (0, 1]$ is a perturbation factor.

- Compute the change in the objective function:

$$\Delta \mathcal{O}(\hat{\mathbf{m}}) = \mathcal{O}(\hat{\mathbf{m}}) - \mathcal{O}(\mathbf{m}). \quad (3.48)$$

- If $\Delta \mathcal{O}(\hat{\mathbf{m}}) < 0$ accept the proposal:
 - Set $\mathbf{m} = \hat{\mathbf{m}}$.
 - If $\mathcal{O}(\mathbf{m}) - \mathcal{O}(\mathbf{m}^*) < 0$, set $\mathbf{m}^* = \mathbf{m}$.

Else:

- Sample $u \sim \mathcal{U}[0, 1]$.
- Compute:

$$P(\mathbf{m}, \hat{\mathbf{m}}) = \exp \left(-\frac{\Delta \mathcal{O}(\hat{\mathbf{m}})}{T} \right). \quad (3.49)$$

- If $u < P(\mathbf{m}, \hat{\mathbf{m}})$, accept the worse solution and set $\mathbf{m} = \hat{\mathbf{m}}$.
- Reduce the temperature $T = \beta T$.

3.4 Other Derivative-Free Optimization Methods

There are numerous methods available for solving complex nonlinear optimization problems that do not rely on derivative information. The variety of derivative-free optimization (DFO) methods proposed in the literature is extensive⁵. Notable examples of DFO methods include the Nelder-Mead method [317], pattern search methods [442], Powell's conjugate direction method [348], and NEWUOA [349].

DFO methods provide a flexible alternative by relying exclusively on function evaluations to navigate the search for optimal solutions. This charac-

⁵ It is worth noting that stochastic methods, discussed in the previous section, could also be included in this category. However, for clarity and to emphasize their significance as data assimilation methods, we have chosen to address them separately.

teristic makes them particularly appealing for reservoir data assimilation, where black-box simulators are often employed, and derivative information is typically unavailable. However, as with stochastic methods, DFO approaches encounter significant challenges, especially in high-dimensional problems. These challenges include the potentially large number of function evaluations required for convergence and the ongoing difficulty of balancing exploration and exploitation. A comprehensive discussion of DFO methods can be found in Conn et al. [92].

3.5 Methods Based on Proxy Modeling

Another important group of methods used in the context of reservoir data assimilation is based on the combination of optimization and sampling methods with proxies of the forward model. The basic idea is to replace executions of a computationally intensive forward model with a faster substitute. Proxy modeling often appears with different names in the literature, including meta-models, surrogate, and response surface modeling.

The successful application of proxy modeling for reservoir data assimilation depends on two key factors: (1) the number of experiments (sample points) required for constructing or training the proxy, and (2) the proxy's capability to replicate the nonlinear behavior of the forward model.

3.5.1 Design of Experiments

In the context of proxy modeling, the objective of the design of experiments is to acquire the maximum information about the behavior of the dynamical system using the minimum possible number of executions of the forward model. Increasing the number of experiments typically results in improvements in the quality of the proxy; however, it diminishes the advantage of using proxies compared to using the full-featured forward model. Therefore, the selection of a proper set of experiments has a major impact on the performance of the process.

Bahrami et al. [23] divide the strategies for sampling experimental points into two broad classes: stationary and sequential (adaptive). In the stationary strategy, a predefined number of experiments is obtained and used to build the proxy. In the sequential strategy, a limited set of experiments is used to construct an initial proxy model, which is then refined through the selection of additional experimental points based on the progress of the optimization or sampling process.

Factorial and Plackett-Burman designs are among the most commonly used methods in the context of stationary sampling. In these cases, each

parameter assumes a prime number of values, known as levels (e.g., 2, 3, 5, etc.). Typically, the number of levels (L) is restricted to either 2 (representing minimum and maximum) or 3 (representing minimum, maximum, and middle values). A full factorial design entails conducting L^{N_m} experiments. For instance, with $N_m = 3$ parameters, this results in $2^3 = 8$ or $3^3 = 27$ experiments, while with $N_m = 10$ parameters, it requires $2^{10} = 1024$ or $3^{10} = 59,049$ experiments. Clearly, these designs scale poorly with N_m .

Fractional factorial designs are a subset of full factorials, where the number of experiments is determined by $L^{N_m-p} > N_m$, with p indicating the degree of reduction. When constructing such designs, it is crucial to maintain balance and orthogonality. In a balanced design, each level of a particular parameter must be used an equal number of times, while orthogonality aims to eliminate correlations between main effects and interaction terms.

Other commonly employed methods for selecting experimental points to construct proxies include random Monte Carlo sampling, Latin hypercube sampling (LHS) [301], and orthogonal array sampling (OAS) [340]. LHS is a stratified sampling method that divides the parameter space into a predefined number of intervals with the same probability value. OAS, on the other hand, is an extension of LHS that uniformly distributes sample points in the dimensional projection of the parameter domain. A recent review of the main methods used for experimental design is presented by Garud et al. [161].

3.5.2 Standard Proxy Models

The number of proxy models developed in the context of reservoir data assimilation is quite extensive, with numerous publications in this area, including a few recent review papers [211, 23, 309]. This topic is too broad to be covered as a section in a book on ensemble methods. Nevertheless, in the following section, we discuss the formulation of three commonly used proxies and provide references to other popular methods.

3.5.2.1 Quadratic Polynomials

Quadratic polynomials were among the first proxies used in the context of reservoir data assimilation [96] and are still employed in various commercial software for data assimilation and uncertainty quantification. A quadratic polynomial, denoted as $\mathcal{Q}(\mathbf{m})$, can be expressed as follows:

$$\begin{aligned}
 \mathcal{Q}(\mathbf{m}) &= a + \mathbf{b}^\top \mathbf{m} + \mathbf{m}^\top \mathbf{C} \mathbf{m} \\
 &= a + \sum_{i=1}^{N_m} b_i m_i + \sum_{i=1}^{N_m} \sum_{j=1}^{N_m} c_{ij} m_i m_j,
 \end{aligned} \tag{3.50}$$

where $c_{ij} = c_{ji}$. $\mathcal{Q}(\mathbf{m})$ is used to replace a quantity of interest, $\mathcal{F}(\mathbf{m})$, obtained through the execution of the forward model. For example, $\mathcal{F}(\mathbf{m})$ may represent the data mismatch objective function, $\mathcal{F}(\mathbf{m}) = \mathcal{O}_d(\mathbf{m})$. Alternatively, $\mathcal{F}(\mathbf{m})$ may correspond to a specific datum predicted by the forward model, in which case we may have $\mathcal{F}_i(\mathbf{m}) = g_i(\mathbf{m})$ for $i = 1, \dots, N_d$. In this situation, we can fit a set of quadratics $\mathcal{Q}_i(\mathbf{m}) \approx g_i(\mathbf{m})$ and use these proxies to compute $\mathcal{O}_d(\mathbf{m})$.

To fully determine all coefficients of $\mathcal{Q}(\mathbf{m})$, it is necessary to conduct $(N_m + 1)(N_m + 2)/2$ experiments. Consequently, when dealing with many model parameters, the computational demand becomes impractical. For example, $N_m = 10$ requires 66 experiments while $N_m = 100$ necessitates 5151 experiments. If we have more than $(N_m + 1)(N_m + 2)/2$ experiments, we have an overdetermined problem (see Appendix B, Section B.3). In such cases, we can compute a least-squares solution to determine the coefficients of the polynomial.

Sometimes, a simplified version of $\mathcal{Q}(\mathbf{m})$ omitting the interaction terms is adopted to reduce the number of experiments. In this scenario, only the diagonal elements of $\mathbf{C}$ are computed, and the quadratic reduces to:

$$\tilde{\mathcal{Q}}(\mathbf{m}) = a + \sum_{i=1}^{N_m} b_i m_i + \sum_{i=1}^{N_m} c_{ii} m_i^2. \tag{3.51}$$

This simplified version requires only $2N_m + 1$ experiments.

3.5.2.2 Kriging

Variations of kriging, or more generically Gaussian processes, are very common techniques used in proxy modeling. Here, we assume that the quantity of interest, $\mathcal{F}(\mathbf{m})$, can be modeled as

$$\mathcal{F}(\mathbf{m}) \approx \mathcal{Q}(\mathbf{m}) + \mathcal{Z}(\mathbf{m}), \tag{3.52}$$

where $\mathcal{Q}(\mathbf{m})$ is a quadratic model and $\mathcal{Z}(\mathbf{m})$ is a kriging model. This formulation uses $\mathcal{Q}(\mathbf{m})$ to predict the trend and $\mathcal{Z}(\mathbf{m})$ to predict the fluctuations of the quantity $\mathcal{F}(\mathbf{m})$.

The goal is to construct $\mathcal{Z}(\mathbf{m})$ using a collection of samples, $z_i = \mathcal{Z}(\mathbf{m}_i)$ for $i = 1, \dots, N_s$. In kriging, the value at an unsampled location, denoted as $z_0 = \mathcal{Z}(\mathbf{m}_0)$, is estimated as a linear weighted sum of the N_s sample points, i.e.,

$$\begin{aligned}
 z_0 &= \sum_{i=1}^{N_s} w_i z_i \\
 &= \mathbf{w}^\top \mathbf{z},
 \end{aligned} \tag{3.53}$$

where z_i is treated as a random variable. Here we assume that $\mathbb{E}[z_i] = 0^6$ for $i = 0, 1, \dots, N_s$, where $i = 0$ means the unsampled location. We also assume stationary covariance

$$\mathbb{C}[\mathbf{z}] = \mathbb{E}[\mathbf{z}\mathbf{z}^\top] = \mathbf{C}_\mathbf{z}. \tag{3.54}$$

The goal is to compute an estimate of z_0 that minimizes the expected squared error

$$\begin{aligned}
 \sigma_K^2 &= \mathbb{E} \left[(z_0 - \mathbf{w}^\top \mathbf{z})^2 \right] \\
 &= \mathbb{E} [z_0^2] - 2\mathbf{w}^\top \underbrace{\mathbb{E} [z_0 \mathbf{z}]}_{\mathbf{c}_0} + \mathbf{w}^\top \underbrace{\mathbb{E} [\mathbf{z}\mathbf{z}^\top]}_{\mathbf{C}_\mathbf{z}} \mathbf{w} \\
 &= \sigma_z^2 - 2\mathbf{w}^\top \mathbf{c}_0 + \mathbf{w}^\top \mathbf{C}_\mathbf{z} \mathbf{w},
 \end{aligned} \tag{3.55}$$

in which case suffices to compute $\nabla \sigma_K^2 = \mathbf{0}$

$$\nabla \sigma_K^2 = -2\mathbf{c}_0 + 2\mathbf{C}_\mathbf{z} \mathbf{w} = \mathbf{0}, \tag{3.56}$$

and solve for $\mathbf{w}$. This procedure leads to

$$\mathbf{w} = \mathbf{C}_\mathbf{z}^{-1} \mathbf{c}_0. \tag{3.57}$$

The kriging error variance becomes

$$\begin{aligned}
 \sigma_K^2 &= \sigma_z^2 - 2\mathbf{w}^\top \mathbf{c}_0 + \mathbf{w}^\top \mathbf{C}_\mathbf{z} \mathbf{w} \\
 &= \sigma_z^2 - 2\mathbf{c}_0^\top \mathbf{C}_\mathbf{z} \mathbf{c}_0.
 \end{aligned} \tag{3.58}$$

Note that the kriging estimate depends on the choice of a covariance structure. The typical assumption is that $\mathbf{C}_\mathbf{z}$ is a function of the distance between pairs of the vectors $\mathbf{m}_i$ and $\mathbf{m}_j$. Therefore

$$\mathbf{C}_\mathbf{z} = \begin{bmatrix} \mathcal{C}(\mathbf{m}_1, \mathbf{m}_1) & \cdots & \mathcal{C}(\mathbf{m}_1, \mathbf{m}_{N_s}) \\ \vdots & \ddots & \vdots \\ \mathcal{C}(\mathbf{m}_{N_s}, \mathbf{m}_1) & \cdots & \mathcal{C}(\mathbf{m}_{N_s}, \mathbf{m}_{N_s}) \end{bmatrix} \tag{3.59}$$

and

⁶ We assumed $\mathbb{E}[z_i] = 0$ because we want to model the fluctuations of the quantity of interest, in which case the mean of $\mathcal{F}(\mathbf{m})$ is modeled by the quadratic model. To derive the kriging estimate with $\mathbb{E}[z_i] = \text{const}$, we need to impose $\sum_{i=1}^{N_s} w_i = 1$ to ensure that the estimator is unbiased.

$$\mathbf{c}_0 = \begin{bmatrix} \mathcal{C}(\mathbf{m}_0, \mathbf{m}_1) \\ \vdots \\ \mathcal{C}(\mathbf{m}_0, \mathbf{m}_{N_s}) \end{bmatrix}, \quad (3.60)$$

where $\mathcal{C}(\mathbf{m}_i, \mathbf{m}_j)$ is a covariance function typically selected as

$$\mathcal{C}(\mathbf{m}_i, \mathbf{m}_j) = \sigma_z^2 \exp \left[- \left(\frac{\|\mathbf{m}_i - \mathbf{m}_j\|}{L} \right)^p \right]. \quad (3.61)$$

The covariance function (3.61) has three parameters: $\sigma_z > 0$, $L > 0$, and $p \in [1, 2]$. The exponent p is typically selected as $p = 2$ to ensure a smooth estimate. The standard deviation σ_z can be estimated from the set $\{z_i\}_{i=1}^{N_s}$ or other information about the behavior of $\mathcal{F}(\mathbf{m})$. The correlation range L can also be estimated from the set. For example, one can use a cross-validation scheme to test the ability of the kriging to predict z_i . Note that $\|\mathbf{m}_i - \mathbf{m}_j\|$ represents the distance between models $\mathbf{m}_i$ and $\mathbf{m}_j$. However, because the components of $\mathbf{m}$ may represent reservoir properties of different orders of magnitude, it is recommended to normalize the components before building the proxy.

3.5.2.3 Support Vector Regression

Support vector regression (SVR) is a particular type of support vector machine (SVM) [45] specialized in nonlinear regression analysis. Both SVR and SVM are methods that belong to a broader class of machine learning methods. Here, we present a particular version called least-squares SVR (LS-SVR) introduced by Suykens and Vandewalle [423] and applied in reservoir data assimilation by Guo et al. [181]⁷.

LS-SVR approximates the quantity of interest as

$$\mathcal{F}(\mathbf{m}) \approx \mathbf{w}^\top \boldsymbol{\varphi}(\mathbf{m}) + b, \quad (3.62)$$

where $\mathbf{w}$ is a vector of weights, $\boldsymbol{\varphi}(\mathbf{m})$ is a nonlinear mapping to a so-called “feature space,” and b is a bias term. $\mathbf{w}$ and b are obtained by solving a regularized least-squares problem based on a training set with N_s experiments

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^{N_s} e_i^2 \\ \text{s.t.} \quad & e_i = y_i - \mathbf{w}^\top \boldsymbol{\varphi}(\mathbf{m}_i) - b, \quad \text{for } i = 1, \dots, N_s. \end{aligned} \quad (3.63)$$

This constrained problem can be solved using Lagrange multipliers by defining the function

⁷ A more complete review on SVR can be found in [403].

$$\mathfrak{L}(\mathbf{w}, b, \mathbf{e}, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^{N_s} e_i^2 - \sum_{i=1}^{N_s} \lambda_i (\mathbf{w}^\top \boldsymbol{\varphi}(\mathbf{m}_i) + b + e_i - d_i), \quad (3.64)$$

where $d_i = \mathcal{F}(\mathbf{m}_i)$ denotes the predicted quantity of interest obtained with the training point $\mathbf{m}_i$. Setting $\nabla \mathfrak{L}(\mathbf{w}, b, \mathbf{e}, \boldsymbol{\lambda}) = \mathbf{0}$ and solving the resulting expression for $\mathbf{w}$, b , $\mathbf{e}$, and $\boldsymbol{\lambda}$ leads to

$$\nabla_{\mathbf{w}} \mathfrak{L}(\mathbf{w}, b, \mathbf{e}, \boldsymbol{\lambda}) = \mathbf{0} \Rightarrow \mathbf{w} = \sum_{i=1}^{N_s} \lambda_i \boldsymbol{\varphi}(\mathbf{m}_i), \quad (3.65)$$

$$\frac{\partial \mathfrak{L}(\mathbf{w}, b, \mathbf{e}, \boldsymbol{\lambda})}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N_s} \lambda_i = 0, \quad (3.66)$$

$$\frac{\partial \mathfrak{L}(\mathbf{w}, b, \mathbf{e}, \boldsymbol{\lambda})}{\partial e_i} = 0 \Rightarrow \lambda_i = \gamma e_i, \quad \text{for } i = 1, \dots, N_s, \quad (3.67)$$

and

$$\frac{\partial \mathfrak{L}(\mathbf{w}, b, \mathbf{e}, \boldsymbol{\lambda})}{\partial \lambda_i} = 0 \Rightarrow \mathbf{w}^\top \boldsymbol{\varphi}(\mathbf{m}_i) + b + e_i - y_i = 0, \quad \text{for } i = 1, \dots, N_s. \quad (3.68)$$

Using (3.65) and (3.67) in (3.68) results in

$$d_i = \sum_{j=1}^{N_s} \lambda_j \boldsymbol{\varphi}(\mathbf{m}_j)^\top \boldsymbol{\varphi}(\mathbf{m}_i) + b + \frac{\lambda_j}{\gamma}, \quad \text{for } i = 1, \dots, N_s, \quad (3.69)$$

which can be written in a matrix form as

$$\mathbf{d} = b \mathbf{1} + \left(\boldsymbol{\Psi} + \frac{1}{\gamma} \mathbf{I} \right) \boldsymbol{\lambda}, \quad (3.70)$$

where $\mathbf{1} = [1 \ 1 \cdots 1]^\top$ and $\boldsymbol{\Psi}$ is the $N_s \times N_s$ matrix with (i, j) th entry computed as $\boldsymbol{\varphi}(\mathbf{m}_i)^\top \boldsymbol{\varphi}(\mathbf{m}_j)$. Combining (3.70) and (3.66) into a single system of linear equations leads to

$$\begin{bmatrix} 0 & \mathbf{1}^\top \\ \mathbf{1} \left(\boldsymbol{\Psi} + \frac{1}{\gamma} \mathbf{I} \right) \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{d} \end{bmatrix}, \quad (3.71)$$

whose solution is

$$b = \frac{\mathbf{1}^\top \left(\boldsymbol{\Psi} + \frac{1}{\gamma} \mathbf{I} \right)^{-1} \mathbf{d}}{\mathbf{1}^\top \left(\boldsymbol{\Psi} + \frac{1}{\gamma} \mathbf{I} \right)^{-1} \mathbf{1}}, \quad (3.72)$$

and

$$\boldsymbol{\lambda} = \left(\boldsymbol{\Psi} + \frac{1}{\gamma} \mathbf{I} \right)^{-1} \mathbf{d} - b \left(\boldsymbol{\Psi} + \frac{1}{\gamma} \mathbf{I} \right)^{-1} \mathbf{1}. \quad (3.73)$$

The quantity of interest evaluated at a new point $\mathbf{m}$ is estimated using

$$\mathcal{F}(\mathbf{m}) \approx \sum_{i=1}^{N_s} \lambda_i \boldsymbol{\varphi}(\mathbf{m}_i)^\top \boldsymbol{\varphi}(\mathbf{m}) + b. \quad (3.74)$$

The resulting expressions only require inner products, $\boldsymbol{\varphi}(\mathbf{m}_i)^\top \boldsymbol{\varphi}(\mathbf{m})$, making it unnecessary to explicitly compute the mapping $\boldsymbol{\varphi}(\mathbf{m})$. This concept is known as the kernel trick [403]. In practice, $\boldsymbol{\varphi}(\mathbf{m}_i)^\top \boldsymbol{\varphi}(\mathbf{m})$ is computed using a kernel function. Various kernel options exist, but the Gaussian kernel, also known as the radial basis function, is commonly used in practical applications:

$$\mathcal{K}(\mathbf{m}_i, \mathbf{m}) = \boldsymbol{\varphi}(\mathbf{m}_i)^\top \boldsymbol{\varphi}(\mathbf{m}) = \exp \left(-\frac{\|\mathbf{m}_i - \mathbf{m}\|_2^2}{\sigma^2} \right). \quad (3.75)$$

Hence,

$$\mathcal{F}(\mathbf{m}) \approx \sum_{i=1}^{N_s} \lambda_i \mathcal{K}(\mathbf{m}_i, \mathbf{m}) + b. \quad (3.76)$$

LS-SVR has two free parameters, γ and σ , which have to be determined beforehand. Sousa and Reynolds [408] claim that for $\gamma > 200$ the training procedure is nearly insensitive to this parameter. The value σ controls the smoothness of the SVR model. Sousa and Reynolds [408] propose to normalize the components of $\mathbf{m}$ and $\mathbf{d}$ in the range $[0, 1]$ and set

$$\sigma = \varrho \sqrt{N_m}, \quad (3.77)$$

with $\varrho \in (0, 1]$, say $\varrho = 0.2$.

3.5.2.4 Example: Proxy of a Nonlinear Function

Fig. 3.3 presents plots of proxies derived from a quadratic function, a combination of quadratic and kriging models, and LS-SVR. These proxies are trained using only four sample points from the function:

$$\mathcal{F}(m) = m \sin(m) + 0.1m^2. \quad (3.78)$$

A quadratic model can effectively capture the general trend but falls short in representing the oscillatory behavior of $\mathcal{F}(m)$. Employing a quadratic model in combination with kriging and LS-SVR yields models that can partially capture the local variations of $\mathcal{F}(m)$. An additional benefit of kriging is the estimation of the standard deviation, which can be utilized to identify areas that may require more samples.

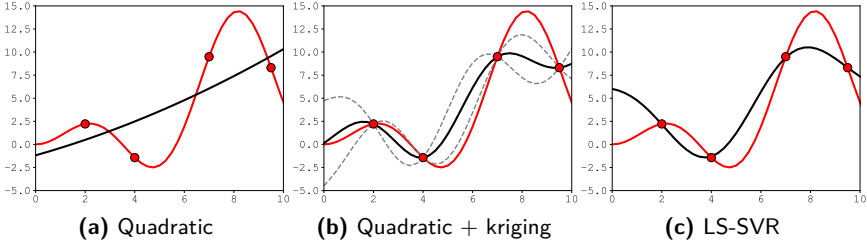


Fig. 3.3: Proxies for the function defined in Eq. (3.78). The red curve in each plot is the actual function, the solid black curve is the proxy, and the red circles are the four sample points. The dashed curves in (b) correspond to the quadratic + kriging $\pm$ one standard deviation. Example 3.5.2.4.

3.5.3 Other Frequently-Adopted Proxy Models

Numerous methods have been proposed as proxies in the context of reservoir data assimilation. Notable examples include variations of shallow neural networks [487, 391, 93, 48, 292] and deep neural networks [486, 427, 426, 385], physics-informed neural networks [190, 24, 95, 474], as well as polynomial chaos expansion [117, 54].

Another group of strategies for proxy modeling involves simplifying the forward model. This can be achieved through various techniques such as up-scaling methods [462, 10], reduced-order modeling [227, 189, 468], streamline simulation [458, 231], multiscale methods [102, 103], among others.

Remark 3.3: Challenges on Proxy Modeling

Despite the numerous applications reported in the literature, the use of proxies for reservoir data assimilation presents several challenges, particularly in balancing accuracy with computational efficiency.

Although proxies are designed to approximate the behavior of high-fidelity models at reduced computational cost, their ability to capture the full complexity of reservoir dynamics can be limited. Proxies often struggle to represent nonlinear interactions between geological heterogeneities and flow processes, which can result in inaccuracies in model predictions. This problem is exacerbated by the “curse of dimensionality.” As the number of model parameters increases, constructing accurate proxies becomes exceedingly complex.

Additionally, the training process requires selecting appropriate datasets to represent the range of reservoir conditions, and this selection can introduce bias if not carefully managed. Moreover, as reservoir conditions evolve over time, proxies may require frequent updates or recalibration, further complicating their use in long-term forecasts or complex data assimilation problems.

3.6 Sensitivity Analysis

Practical reservoir data assimilation studies frequently incorporate a screening or sensitivity analysis (SA) step to identify the parameters that have the greatest influence on the predicted data. One fundamental reason for conducting SA is to reduce the number of parameters by eliminating those with low influence in predicting data. Additionally, SA is useful for enhancing the understanding of the reservoir’s behavior in relation to the main uncertainty parameters, offering deeper insights into the dynamics of the system.

The simplest SA method involves perturbing each model parameter individually while keeping the other parameters at a base value. Typically, two perturbations per parameter are considered—one at its lower value and another at its upper value. By comparing the variations in the predicted data with a base simulation case, we can rank the parameters according to their importance, typically presented in the form of a tornado plot (Fig. 3.4). The total number of simulations required in this approach is $2N_m + 1$. Besides the obvious computational cost for problems with a large number of parameters, this method has the limitation of neglecting the combined effects of the model parameters.

There are more advanced SA methods beyond the standard approach of evaluating the forward model by varying each parameter between its lower

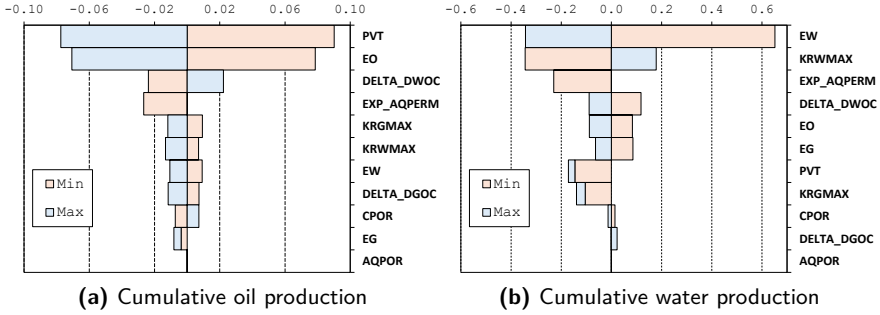


Fig. 3.4: Example of tornado plots from a one-at-a-time SA for a modified version of the PUNQ-S3 problem [150]. The vertical axis labels represent the selected model parameters. Each bar shows the relative change in the quantity of interest when the corresponding parameter is set to its minimum or maximum value, compared to the base case.

and upper bounds while holding the others fixed. These methods, such as the Morris [307] and Sobol [406] methods, provide a global assessment of parameter influence by accounting for interactions and nonlinear effects across the entire input space. The Morris method, based on perturbing one parameter at a time, offers an efficient screening tool for identifying important variables. Sobol analysis is a variance-based method that estimates sensitivity coefficients by decomposing the output variance into contributions from individual inputs and their interactions. A key advantage of Sobol analysis is its ability to handle nonlinear and non-additive models without requiring specific assumptions about model structure. However, its implementation can be computationally demanding, particularly for high-dimensional models, as it typically requires a large number of model evaluations.

3.6.1 Sensitivity Analysis Based on Monte Carlo Simulations

In practice, SA is often performed using Monte Carlo simulations, where the goal is to estimate the sensitivity coefficients of a quantity of interest, $\mathcal{F}(\mathbf{m})$, with respect to $\mathbf{m}$, based on an ensemble of N_e random samples. This procedure is particularly convenient in the context of ensemble methods, where the sensitivity coefficients can be readily estimated from the prior ensemble simulations. Several approaches exist for this purpose, but a particularly useful method is to estimate the sensitivities using the correlation coefficients between $\mathbf{m}$ and $\mathcal{F}(\mathbf{m})$.

3.6.1.1 Pearson Correlation

Let m represent a model parameter and d the quantity of interest to be analyzed. One of the simplest methods for SA is to estimate the sensitivity values using the Pearson's correlation coefficient:

$$\rho_{md} = \frac{\sum_{j=1}^{N_e} (m_j - \bar{m})(d_j - \bar{d})}{\sqrt{\sum_{j=1}^{N_e} (m_j - \bar{m})^2} \sqrt{\sum_{j=1}^{N_e} (d_j - \bar{d})^2}}, \quad (3.79)$$

where $\rho_{md} \in [-1; 1]$ provides a measure of the strength of the linear relationship between m and d . $\rho_{md} > 0$ means that both m and d are increasing or decreasing together, while $\rho_{md} < 0$ means that m and d move in opposite directions. $|\rho_{md}| \approx 1$ corresponds to a linear relationship, while $\rho_{md} \approx 0$ indicates that there is no linear relationship between m and d .

3.6.1.2 Spearman Correlation

The Spearman correlation between two variables is equivalent to the Pearson correlation applied to the rank values of those variables. Instead of assessing only linear relationships like the Pearson correlation, the Spearman correlation evaluates monotonic relationships, whether linear or nonlinear. The Spearman correlation is less sensitive to outliers because it operates on the ranked values, meaning the actual differences between values do not influence the calculation.

3.6.1.3 Partial Correlation

Partial correlations aim to estimate the correlation between two random variables, removing the effect of other relevant variables. The partial correlation between m and d given a set of removed variables $\mathbf{m}_0$, denoted $\rho_{md, \mathbf{m}_0}$, can be computed as the correlation between the residuals r_m and r_d resulting from a linear regression of m with $\mathbf{m}_0$ and d with $\mathbf{m}_0$, respectively.

The linear regression between m and $\mathbf{m}_0$ can be written as

$$\begin{aligned} m &= w_0 + w_1 m_1 + w_2 m_2 + \dots + w_{N_m-1} m_{N_m-1} \\ &= \mathbf{x}^\top \mathbf{w}, \end{aligned} \quad (3.80)$$

where $\mathbf{x}^\top = [1 \ \mathbf{m}_0^\top]$, i.e., $\mathbf{x}$ is the vector containing $\mathbf{m}_0$ augmented with a 1 in the first entry to allow for a constant term in the regression. The coefficients of the linear regression can be computed by solving the least-squares problem:

$$\begin{aligned}
\mathbf{w}_m^* &= \arg \min_{\mathbf{w}} \sum_{j=1}^{N_e} (m_j - \mathbf{x}_j^\top \mathbf{w})^2 \\
&= \arg \min_{\mathbf{w}} \|\mathbf{m} - \mathbf{X}\mathbf{w}\|_2^2
\end{aligned} \tag{3.81}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & m_{1,1} & \cdots & m_{N_m,1} \\ \vdots & & & \vdots \\ 1 & m_{1,N_e} & \cdots & m_{N_m,N_e} \end{bmatrix}. \tag{3.82}$$

If $N_e \geq N_m$ and $\text{rank}(\mathbf{X}) = N_m$, the solution of the LS problem is

$$\mathbf{w}_m^* = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}\mathbf{m}. \tag{3.83}$$

Similarly for d , we write

$$\mathbf{w}_d^* = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}\mathbf{d}. \tag{3.84}$$

The residuals for the j th samples of m , d and $\mathbf{X}$ are given by

$$r_{m_j} = m_j - \mathbf{x}_j^\top \mathbf{w}_m^* \tag{3.85}$$

and

$$r_{d_j} = d_j - \mathbf{x}_j^\top \mathbf{w}_d^*. \tag{3.86}$$

The partial correlation is computed with the standard formula

$$\rho_{md, \mathbf{m}_0} = \frac{\sum_{j=1}^{N_e} r_{m_j} r_{d_j}}{\sqrt{\sum_{j=1}^{N_e} r_{m_j}^2} \sqrt{\sum_{j=1}^{N_e} r_{d_j}^2}}. \tag{3.87}$$

3.6.1.4 Regression Coefficients

Another popular strategy is to use the regression coefficients of linear or quadratic polynomials as sensitivity coefficients. Note that because model parameters and data may present different magnitudes, it is important to normalize before regression. For example, we can write

$$\hat{m} = \frac{m - \bar{m}}{\sigma_m} \tag{3.88}$$

and

$$\hat{d} = \frac{d - \bar{d}}{\sigma_d} \tag{3.89}$$

and perform the regression with $\hat{m}$ and $\hat{d}$.

3.6.1.5 Example: Sensitivity Analysis in the PUNQ-S3 Case

Fig. 3.5 presents tornado plots of cumulative oil production for a modified version of the PUNQ-S3 case [150]. These plots were generated using Monte Carlo sampling with an ensemble of 100 realizations. The four methods considered yielded similar rankings of model parameters, with the top three most influential parameters remaining consistent across all plots. These parameters also align with the results of the one-at-a-time SA shown in Fig. 3.4.

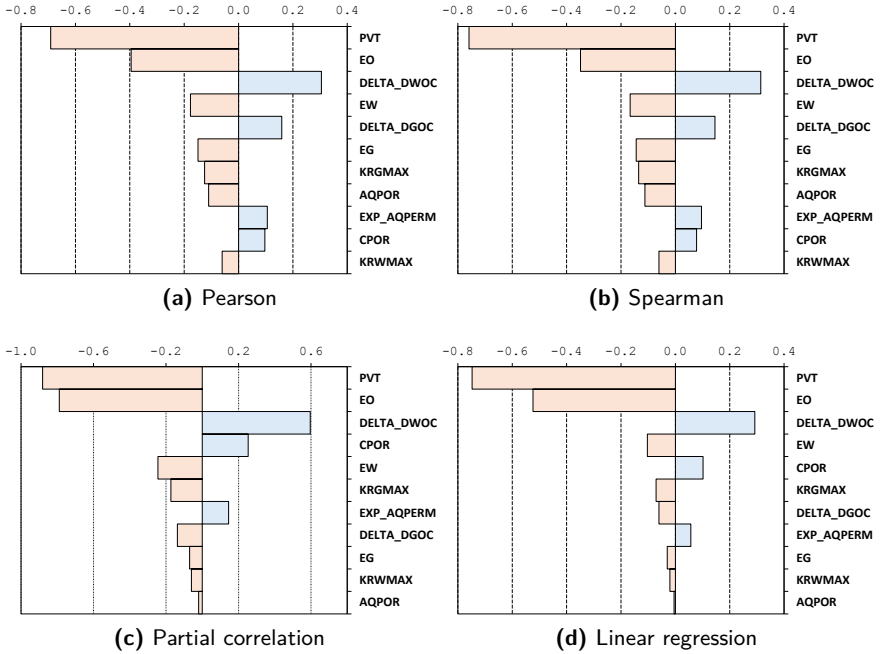


Fig. 3.5: Tornado plots from a SA of cumulative oil production for a modified version of the PUNQ-S3 problem. The labels on the vertical axis represent the selected model parameters, while each bar represents the sensitivity coefficient estimated for each parameter. Example 3.6.1.5.

3.7 Parametrization

One of the primary challenges in applying optimization methods to reservoir data assimilation is dealing with a large number of model parameters. When every parameter in a reservoir model, such as the porosity and permeability of each gridblock, is allowed to vary independently, the resulting model space

becomes exceedingly vast. Expecting an optimization method to efficiently explore the entire $\mathbb{R}^{N_m}$ may seem impractical. However, there are mitigating factors. Physical constraints, like porosity values falling within the range $[0, 0.4]$, play a role. Additionally, geological models possess internal structures that suggest the existence of model manifolds—a connected region in space where each point represents a conceivable model. In simpler terms, the presence of a model manifold implies that we can reasonably approximate points within it using a small number of dimensions within a higher-dimensional space (Fig. 3.6). This concept opens up possibilities for parametrizations to explore the model space more efficiently.

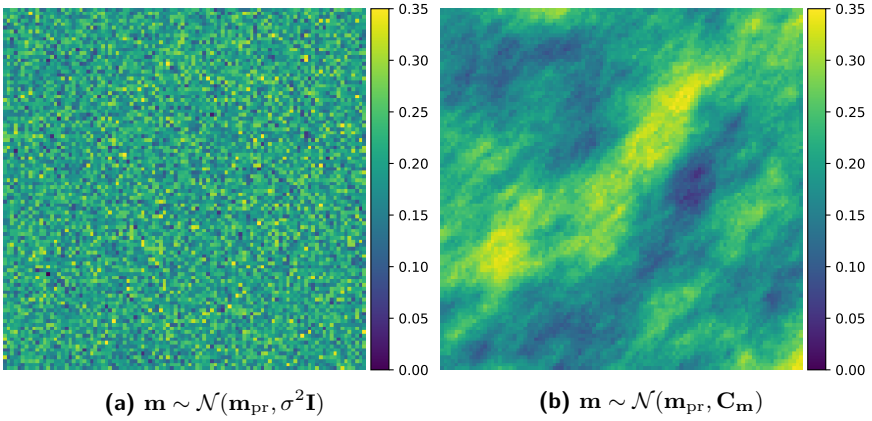


Fig. 3.6: Two random realizations of porosity distribution in a model. Panel (a) shows a realization generated by independently sampling each gridblock, resulting in a noisy realization. While there is a non-zero probability of obtaining a plausible model using this procedure, it never occurs in practice. Panel (b) shows a realization generated with an underlying covariance structure, illustrating that the models of interest occupy only a fraction of the $\mathbb{R}^{N_m}$ space.

While the parameter space can be extremely large, the dynamic data often carries limited information. This limitation arises from the system's diffusive nature, the fact that data are collected at limited numbers of wells sparsely distributed in the reservoir, the presence of measurement errors, and the inherently approximate nature of reservoir models. Consequently, the number of model parameters that can be reliably determined from the available data remains relatively small.

The fundamental concept behind parametrization methods is the reduction of the model space's dimension. As discussed in Oliver et al. [335], parametrization methods can be described as expressing changes in the model as a linear combination of a limited set of basis vectors, represented by the equation:

$$\delta \mathbf{m} = \mathbf{P} \mathbf{z}. \quad (3.90)$$

Here, $\mathbf{P}$ is a matrix with dimensions $N_m \times N_z$, and its columns constitute a basis for a vector space of dimension N_z . Meanwhile, $\mathbf{z}$ is a vector containing coefficients corresponding to each basis vector. The primary objective is to choose $\mathbf{P}$ in such a way that $N_z \ll N_m$ while ensuring it spans a subspace where $\mathbf{m}$ can be accurately represented.

Remark 3.4: Reduced Parametrization and Uncertainty

The choice of a reduced parametrization involves a trade-off. On the one hand, representing the model with fewer basis vectors helps alleviate the problem's ill-conditioning, thereby improving the performance of optimization algorithms. On the other hand, reducing the parametrization confines the solution search to a limited subspace, which can lead to an underestimation of uncertainty.

The preface of “Inverse Theory for Petroleum Characterization and History Matching,” [334] highlights this issue by stating: “*It is impossible to correctly estimate all the parameters of a model from inaccurate, insufficient, and inconsistent data, but reducing the number of parameters in order to get low levels of uncertainty is misleading.*”

3.7.1 Zonation

Zonation is one of the oldest and most straightforward parametrization methods, and it is still widely used in practice. In this approach, the basis vectors represent constant values within specific reservoir regions—zones—and are zero elsewhere. These zones are usually predefined and selected based on geological interpretations and engineering judgment. While zonation typically results in an initial significant reduction in the objective function, the final mismatch tends to be larger than desired due to the limited number of parameters (degrees of freedom) and suboptimal zone selections.

Critiques of zonation primarily revolve around the potentially arbitrary nature of zone delineation, which may not accurately capture the geological characteristics of the reservoir. Additionally, zonation can introduce discontinuities in rock properties between zones, impacting the representation of subsurface features.

3.7.1.1 Example: Manual History Matching

Manual history matching is a widely practiced method in petroleum reservoir engineering for data assimilation involving direct manipulation of the reservoir model. In this approach, reservoir engineers typically apply modifications, such as zonal multiplication factors, to improve the model's agreement with historical data. The process is often conducted in a trial-and-error manner, drawing on the engineer's experience and close collaboration with geoscientists involved in the model's construction. This method can be time-consuming and subjective, as it heavily depends on the engineer's judgment and expertise.

Fig. 3.7a shows an example of a permeability field obtained through a manual history matching approach for an oilfield in the Campos Basin. In this model, the field engineer modified the initial permeability by applying multiplication factors to specific areas to match water cut data at oil-producing wells. These areas can be easily identified by the rectangular shapes of the zones; for example, see the region around well P-209. For comparison, Fig. 3.7b shows the mean permeability obtained using an ensemble-based data assimilation method (ES-MDA, Chapter 5) for the same field, which exhibits a more natural permeability distribution, without the discontinuities observed in the manual history matched model.

One limitation of the zonation parametrization and manual approaches is the difficulty in generating multiple solutions. It is interesting to note that in the example shown in Fig. 3.7, the reservoir engineer achieved a data match by reducing the permeability around well P-209. For the same problem, the ensemble method found solutions that, on average, have high permeabilities around the same well. This result illustrates the multiplicity of solutions in a data assimilation problem.

3.7.2 Pilot Point

In the pilot-point method [101, 250], reservoir parameter values are determined at a small number of specific locations known as pilot points. The property values for the remaining gridblocks within the model are then computed using interpolation techniques, often employing kriging. Various strategies for selecting the locations of these pilot points have been discussed in the literature [357, 8].

The parametrization introduced by the pilot-point method can be expressed in the form of Eq. (3.90), with $\mathbf{P}$ containing the columns of $\mathbf{C}_m$ corresponding to the positions of the pilot points, and $\mathbf{z}$ containing the coefficients of the linear combinations. It has been reported in the literature [333, 266] that this parametrization tends to yield extreme parameter values at the pilot points, leading to nonphysical artifacts in the rock properties.

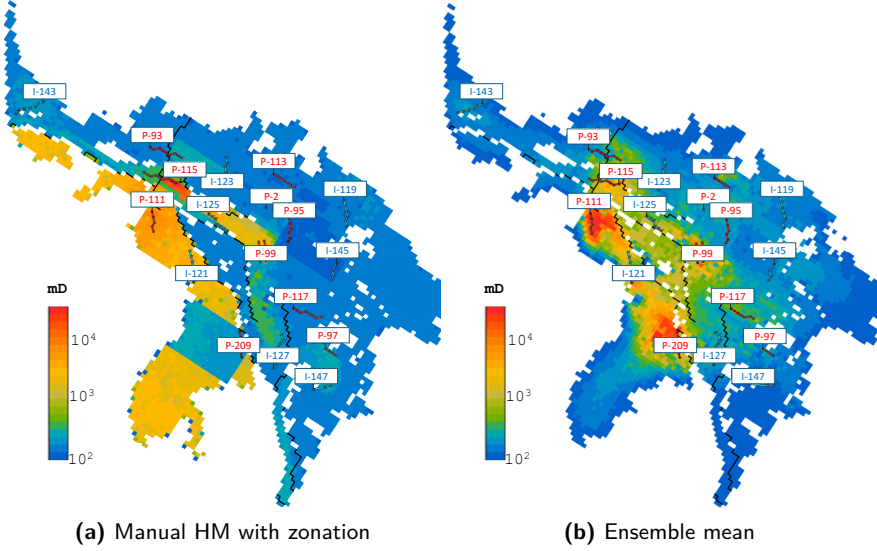


Fig. 3.7: Permeability for a model of an oilfield in the Campos Basin. Panel (a) shows the permeability after a manual history matching using a zonation parametrization. Panel (b) shows the ensemble mean permeability obtained with an ensemble-based data assimilation. Example 3.7.1.1.

Consequently, it may be necessary to introduce some form of regularization to control the property values at these pilot point locations.

3.7.3 Gradual Deformation

Gradual deformation is designed to gradually deform a geological model until it honors the measurements [368, 203]. It operates by expressing the model as a linear combination of independent Gaussian prior models, $\{\mathbf{m}_j\}_{j=1}^{N_s}$, where $\mathbf{m}_j \sim \mathcal{N}(\mathbf{m}_{pr}, \mathbf{C}_m)$:

$$\hat{\mathbf{m}} = \sum_{j=1}^{N_s} z_j (\mathbf{m}_j - \mathbf{m}_{pr}) + \mathbf{m}_{pr}. \quad (3.91)$$

The coefficients must be selected such that

$$\sum_{j=1}^{N_s} z_j^2 = 1. \quad (3.92)$$

Note that for $\sum_{j=1}^{N_s} z_j^2 = 1$ the expectation and covariance of $\hat{\mathbf{m}}$ become

$$\begin{aligned}
 \mathbb{E}[\hat{\mathbf{m}}] &= \mathbb{E} \left[\sum_{j=1}^{N_s} z_j (\mathbf{m}_j - \mathbf{m}_{\text{pr}}) + \mathbf{m}_{\text{pr}} \right] \\
 &= \mathbf{m}_{\text{pr}} + \sum_{j=1}^{N_s} z_j \mathbb{E}[\mathbf{m}_j - \mathbf{m}_{\text{pr}}] \\
 &= \mathbf{m}_{\text{pr}},
 \end{aligned} \tag{3.93}$$

and

$$\begin{aligned}
 \mathbb{C}[\hat{\mathbf{m}}] &= \mathbb{E} \left[\left(\sum_{i=1}^{N_s} z_i (\mathbf{m}_i - \mathbf{m}_{\text{pr}}) \right) \left(\sum_{j=1}^{N_s} z_j (\mathbf{m}_j - \mathbf{m}_{\text{pr}}) \right)^\top \right] \\
 &= \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} z_i z_j \mathbb{E}[(\mathbf{m}_i - \mathbf{m}_{\text{pr}})(\mathbf{m}_j - \mathbf{m}_{\text{pr}})^\top] \\
 &= \sum_{j=1}^{N_s} z_j^2 \mathbf{C}_{\mathbf{m}} \\
 &= \mathbf{C}_{\mathbf{m}}.
 \end{aligned} \tag{3.94}$$

These results mean that selecting $\{z_j\}_{j=1}^{N_s}$ such that $\sum_{j=1}^{N_s} z_j^2 = 1$ yields to $\hat{\mathbf{m}}$ to be a sample from the prior PDF. Imposing this condition, however, is somewhat contradictory if we want to generate conditional realizations, i.e., samples from the posterior. The most basic gradual deformation algorithm combines models in pairs and use

$$\hat{\mathbf{m}}(\nu) = \mathbf{m}_1 \cos(\pi\nu) + \mathbf{m}_2 \sin(\pi\nu), \tag{3.95}$$

with ν ranging between 0 and 2. Note that in this case, we have $z_1 = \cos(\pi\nu)$, $z_2 = \sin(\pi\nu)$, and $z_1^2 + z_2^2 = 1$. We then minimize the objective function with respect to a single parameter, ν . If an acceptable data match is achieved, the process stops. Otherwise, $\mathbf{m}_1$ is replaced with the best $\hat{\mathbf{m}}(\nu)$, $\mathbf{m}_2$ is sampled from the prior, and the optimization is repeated. Several refinements to this method exist; for instance, Hu et al. [204] developed a procedure for locally deforming the models.

3.7.4 Parametrizations Based on PCA

Principal component analysis (PCA) or Karhunen-Loève expansion is a technique used to decompose the spatial-covariance of an image into a series of

orthogonal basis. PCA of $\mathbf{C}_m$ was one of the first parametrization methods used to describe permeability fields in reservoir data assimilation [164, 363].

For PCA, we need to solve the following eigenvalue problem

$$\mathbf{C}_m \mathbf{x} = \lambda \mathbf{x}, \quad (3.96)$$

with λ denoting the eigenvalue of $\mathbf{C}_m$ corresponding to the eigenvector $\mathbf{x}$.

The idea is to use a small set of eigenvectors corresponding to the largest eigenvalues as a basis to parameterize the model. In this sense, PCA parametrization tends to be very effective. However, one problem is that if $\mathbf{C}_m$ is large, the eigenvalue problem may not be computationally feasible to solve.

One alternative is to replace $\mathbf{C}_m$ with a low-rank approximation computed from a set of N_e prior realizations

$$\begin{aligned} \tilde{\mathbf{C}}_m &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{m}_j - \bar{\mathbf{m}}) (\mathbf{m}_j - \bar{\mathbf{m}})^\top \\ &= \Delta \mathbf{M} \Delta \mathbf{M}^\top, \end{aligned} \quad (3.97)$$

where

$$\Delta \mathbf{M} = \frac{1}{\sqrt{N_e - 1}} [\mathbf{m}_1 - \bar{\mathbf{m}}, \dots, \mathbf{m}_{N_e} - \bar{\mathbf{m}}]. \quad (3.98)$$

Applying SVD (Appendix A, Section A.4) to $\Delta \mathbf{M}$ and keeping the $N_r \leq \min\{N_m, N_e - 1\}$ largest singular values and the corresponding singular vectors

$$\Delta \mathbf{M} = \mathbf{U}_r \Sigma_r \mathbf{V}_r^\top. \quad (3.99)$$

It is easy to show that we can write new realizations as

$$\hat{\mathbf{m}} = \bar{\mathbf{m}} + \mathbf{U}_r \Sigma_r \mathbf{z}, \quad (3.100)$$

with $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Note that because we select $N_e \ll N_m$, we have $N_r \ll N_m$.

3.7.4.1 Example: PCA with Low-Rank Prior Covariance

Consider the example from Section 2.4.1.3, which showed unconditional realizations of porosity on a 100×100 uniform grid generated using the Cholesky decomposition of the prior covariance $\mathbf{C}_m$. Using the same procedure, we generated an ensemble with $N_e = 1,000$ realizations. Fig. 3.8 displays the first three realizations from this ensemble.

Here, we use this ensemble to construct a low-rank approximation of $\mathbf{C}_m$ and employ Eq. (3.100) to sample a new realization of $\mathbf{m}$ by sampling $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Since $\mathbf{z} \in \mathbb{R}^{N_r}$ and $N_r \leq N_e - 1 < N_m$, $\mathbf{z}$ serves as a reduced parametrization of $\mathbf{m}$, which can be further refined by reducing the number of retained singular values. A common approach to determine N_r is to calculate the number of singular values needed to retain a predefined fraction, ξ , of the

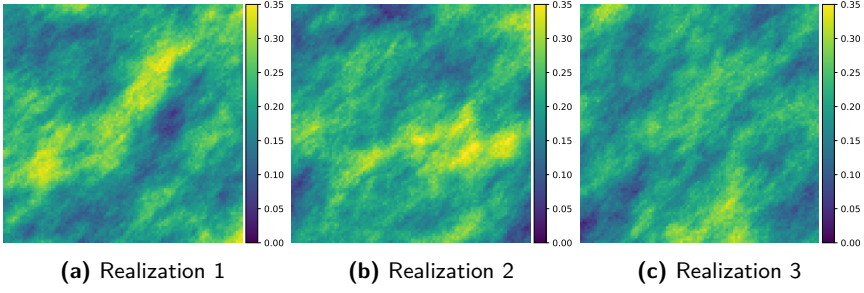


Fig. 3.8: Three realizations of porosity generated with the Cholesky decomposition of $\mathbf{C}_m$. Example 3.7.4.1.

total sum of singular values. In other words, N_r is selected as the smallest integer such that

$$\sum_{i=1}^{N_r} \sigma_i \geq \xi, \quad \text{for } \xi \in (0, 1]. \quad (3.101)$$

Eq. (3.101) assumes that the singular values, σ_i , are sorted in decreasing order. Fig. 3.9 presents a plot of the singular values of $\tilde{\mathbf{C}}_m$. Given that we used 1,000 realizations and subtracted the average vector to form $\Delta\mathbf{M}$, the total number of non-zero singular values is 999.

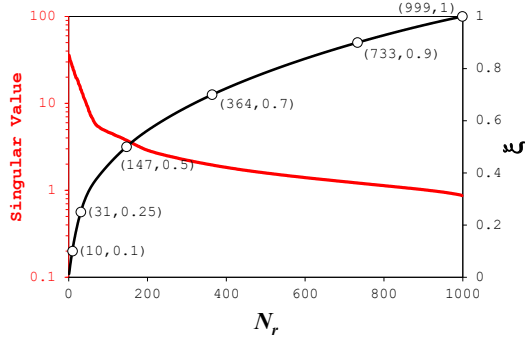


Fig. 3.9: Singular values of $\tilde{\mathbf{C}}_m$. The values besides each white circle indicate the number of singular values and corresponding energy, ξ . Example 3.7.4.1.

Fig. 3.10 presents a realization generated with PCA at different levels of truncation energy. The realization generated while retaining all singular values (Fig. 3.10a) presents similar features of the prior realizations generated using the Cholesky decomposition of the full $\mathbf{C}_m$. As the number of singular values decreases, high-frequency components are discarded, resulting

in smoother realizations. Until $\xi = 0.7$, the differences are barely noticeable. However, further reduction in the number of singular vectors leads to overly smooth realizations. An important characteristic of PCA is that it provides an optimal basis in a least-squares sense, meaning it offers the best representation of $\mathbf{m}$ within the space spanned by the N_e prior realizations.

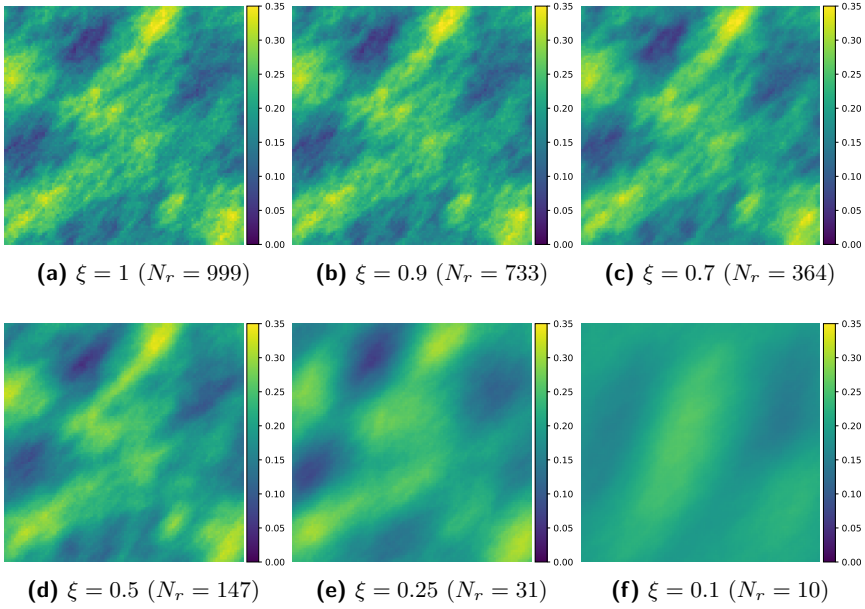


Fig. 3.10: PCA realization generated with different levels of truncation energy. Example 3.7.4.1.

3.7.5 Parametrizations Based on Data Sensitivity

PCA parametrization uses only prior information to define the basis vectors, which remain fixed throughout the data assimilation process. However, an “optimal parametrization” should ideally consider both the type and amount of data utilized. This can be achieved by incorporating data sensitivity (or a related measure) into the parametrization. An early example of such a parametrization is the gradzone method [40]. In this method, the primary eigenvectors of the matrix $\mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G}$ are employed to identify regions (gradzones) that have the most significant impact on reducing the objective function. It is worth noting that instead of computing the entire sensitivity

matrix, the original method calculates derivatives at a few spaced locations and then interpolates the values for the rest of the model.

Rodrigues [367] proposed a parametrization method based on the dominant right singular values of the dimensionless sensitivity matrix, $\mathbf{G}_d = \mathbf{C}_e^{-1/2} \mathbf{G} \mathbf{C}_m^{1/2}$. Rodrigues also introduced a technique for computing the singular vectors associated with the largest singular values without explicitly constructing the matrix $\mathbf{G}_d$ by employing the Lanczos algorithm [170]. Building on these ideas, Tavakoli and Reynolds [429] extended the methodology and provided a theoretical justification, arguing that parametrization using the leading right singular vectors of $\mathbf{G}_d$ creates an optimal basis by discarding directions that contribute minimally to uncertainty reduction. Further developments and enhancements of this approach are discussed in subsequent works [430, 393, 395].

3.7.6 Parametrizations for Facies

The typical workflow for constructing reservoir models includes a facies modeling step, in which the reservoir is divided into regions with similar rock properties. Facies, in this context, are categorical variables representing distinct rock types. Incorporating facies into data assimilation tends to be challenging due to their categorical nature.

The literature presents a wide variety of methods for reparameterizing facies models in data assimilation. Among the earliest approaches is the use of level-set functions [64, 272, 305, 304, 347]. Another class of methods includes the dynamic updating of probability maps, which are incorporated as soft constraints in geostatistical simulations [212, 432, 65, 252, 388]. Additional techniques include the discrete cosine transform [213], kernel PCA [378], optimization-based PCA [453, 120], and PCA with truncation based on the cumulative density function of the prior [69, 70, 160].

To date, truncated Gaussian [296] and truncated plurigaussian (TPG) methods [157, 17, 35] are the most utilized and effective approaches for parameterizing facies models, especially in the context of ensemble-based data assimilation [267, 484, 387, 386]. These methods represent facies using one or more Gaussian latent vectors along with a rock-type rule. During the data assimilation process, these latent vectors are iteratively updated to generate new facies distributions.

Fig. 3.11 illustrates a facies realization generated by truncating two latent vectors, $\mathbf{z}_1$ and $\mathbf{z}_2$. In this example, $\mathbf{z}_1$ models channelized structures, while $\mathbf{z}_2$ models a secondary depositional process. During data assimilation, the latent vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ are updated and used to generate new facies.

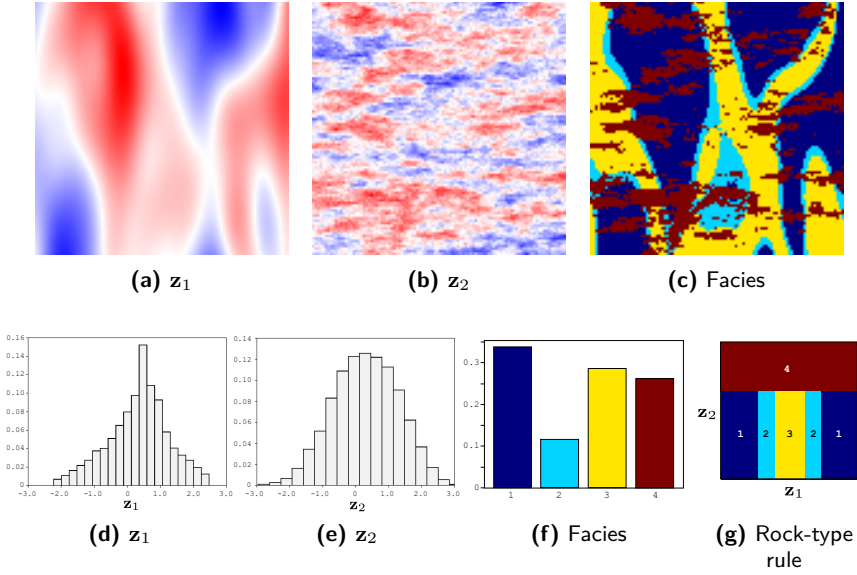


Fig. 3.11: Model with four facies generated by truncating two latent Gaussian vectors.

3.7.7 Parametrizations for Complex Geological Models

Parametrizations based on TPG are widely used in practical reservoir data assimilation applications [75, 123]. However, TPG may not adequately represent facies with complex geometries resulting from object-based modeling [107, 106], multiple-point statistics [179, 417, 290], or rule-based modeling [352]. Consequently, considerable research is underway to develop parametrization methods capable of representing complex facies distributions. Recent publications [308, 63, 57] suggest that deep learning methods show promise as alternative approaches for re-parameterizing facies models. Similar to TPG, some of these methods represent facies using Gaussian latent variables, allowing the prior PDF to be considered Gaussian. This approach aligns with the objective function formulation in Chapter 2, making these methods particularly suitable for use with ensemble smoothers.

3.7.8 Big-Loop Parametrization

The concept of “big-loop” workflow, as described by Webb et al. [461], involves integrating various components of the reservoir modeling process, such as structural modeling, facies and petrophysical modeling, upscaling, and reservoir simulation, during the data assimilation step.

In a conventional data assimilation problem, the goal is to sample the distribution $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$, where $\mathbf{m}$ contains, for example, values of porosity and permeability at the gridblocks of the model. This problem is represented by Bayes' rule as:

$$\underbrace{p(\mathbf{m}|\mathbf{d}_{\text{obs}})}_{\text{posterior}} \propto \underbrace{p(\mathbf{d}_{\text{obs}}|\mathbf{m})}_{\text{likelihood}} \times \underbrace{p(\mathbf{m})}_{\text{prior}}. \quad (3.102)$$

In a big-loop workflow, we may also want to estimate hyper-parameters, such as variogram ranges and azimuth, which are used to generate porosity and permeability realizations through geostatistical methods. This extended problem can be described using hierarchical priors in Bayes' rule

$$\underbrace{p(\mathbf{m}|\mathbf{d}_{\text{obs}})}_{\text{posterior}} \propto \underbrace{p(\mathbf{d}_{\text{obs}}|\mathbf{m})}_{\text{likelihood}} \times \underbrace{p(\mathbf{m}|\mathbf{z})}_{\text{conditional prior}} \times \underbrace{p(\mathbf{z})}_{\text{hierarchical prior}}, \quad (3.103)$$

where $\mathbf{z}$ contains the unknown hyper-parameters. This approach is more challenging due to the “double stochasticity”: $\mathbf{m}$ depends on both the random vector $\mathbf{z}$ and the random seed r used to initiate the geostatistical algorithm. Fig. 3.12 illustrates this issue. In a standard “small loop” problem, solving with a derivative-based method requires computing derivatives of the predictions with respect to $\mathbf{m}$. However, in the big-loop case, we must employ the chain rule due to the dependency of $\mathbf{m}$ on $\mathbf{z}$ and account for the random seed r . This involves marginalizing the random seed by averaging the derivatives—repeating the derivative calculation for different random seeds and computing the mean. This procedure, while necessary, significantly increases the computational cost of the process.

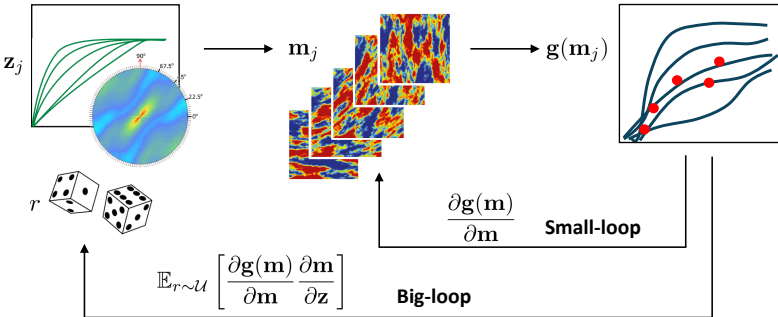


Fig. 3.12: Illustration of the small- and big-loop model updating schemes.

Ensemble Kalman Filter

Abstract: *This chapter reviews the ensemble Kalman filter (EnKF), a widely used data assimilation method across various geophysical fields. It addresses key topics such as the Kalman filter, sequential data assimilation, ensemble covariances, parameter-state consistency, and square root formulations. This chapter also discusses how to integrate the EnKF with reservoir simulators.*

4.1 Introduction

Reservoir data assimilation is typically framed as a parameter-estimation problem, aiming to estimate the vector of model parameters, $\mathbf{m}$, given a set of observations, $\mathbf{d}_{\text{obs}}$. However, this chapter adopts a different approach by considering the sequential assimilation of data over time. In this context, it is convenient to formulate the problem in terms of a combined parameter-state vector. While parameters refer to the “static” uncertainty properties of the reservoir, states refer to the “dynamic” variables that evolve over time, such as pressure, temperature, phase saturation, and fluid composition.

The primary reason for expressing the problem in terms of both parameters and states is to enable the forward model to be initiated from any intermediate time step during the sequential data assimilation process. This approach assumes that the forward model is a first-order Markov process, meaning that the system’s state at time step t_n , denoted as $\mathbf{x}^n$, depends solely on $\mathbf{m}$ and the state at the previous time step, t_{n-1} , and not on any earlier instances, i.e.,

$$\mathbf{x}^n = \mathbf{f}(\mathbf{m}, \mathbf{x}^{n-1}), \quad (4.1)$$

where $\mathbf{f}(\cdot)$ represents the nonlinear forward model.

The primary objective of this chapter is to discuss the ensemble Kalman filter (EnKF), a sequential data assimilation method introduced by Evensen [136] that has become widely adopted in oceanographic and atmospheric

sciences. EnKF is an ensemble-based adaptation of the renowned Kalman filter, designed to assimilate data in nonlinear dynamical systems.

Let $\mathbf{y}^n \in \mathbb{R}^{N_y}$ be the combined parameter-state vector at time t_n defined as:

$$\mathbf{y}^n = \begin{bmatrix} \mathbf{m} \\ \mathbf{x}^n \end{bmatrix} = \begin{bmatrix} \mathbf{m} \\ \mathbf{f}(\mathbf{m}, \mathbf{x}^{n-1}) \end{bmatrix}. \quad (4.2)$$

Also, let $\mathbf{d}_{\text{obs}}^n$ denote the set of observations available at time t_n ($\mathbf{d}_{\text{obs}}^n \in \mathbb{R}^{N_n}$)¹. Given $\mathbf{d}_{\text{obs}}^n$, we can define three problems of interest:

- **Filtering:** find the distribution of the current parameter-state vector given the current observations, $p(\mathbf{y}^n | \mathbf{d}_{\text{obs}}^n)$.
- **Smoothing:** find the distribution of a past parameter-state vector given the current observations, $p(\mathbf{y}^k | \mathbf{d}_{\text{obs}}^n)$ for $k < n$.
- **Forecasting:** find the distribution of a future parameter-state vector given the current observations, $p(\mathbf{y}^k | \mathbf{d}_{\text{obs}}^n)$ for $k > n$.

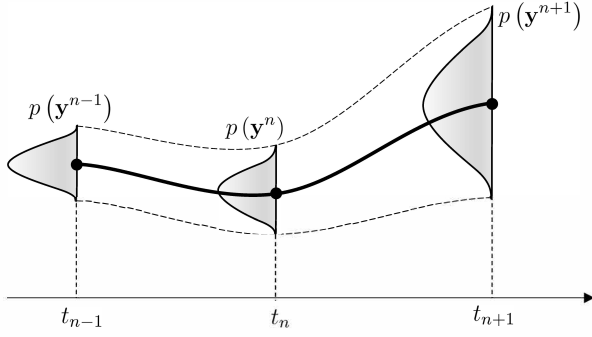
Fig. 4.1 illustrates the evolution of the distribution of a state variable for three times steps before (prior) and after (posterior) the assimilation of a datum at time t_n . The posterior distributions $p(\mathbf{y}^{n-1} | \mathbf{d}_{\text{obs}}^n)$, $p(\mathbf{y}^n | \mathbf{d}_{\text{obs}}^n)$, and $p(\mathbf{y}^{n+1} | \mathbf{d}_{\text{obs}}^n)$ correspond to the smoothing, filtering, and forecasting solutions, respectively².

4.2 Kalman Filter

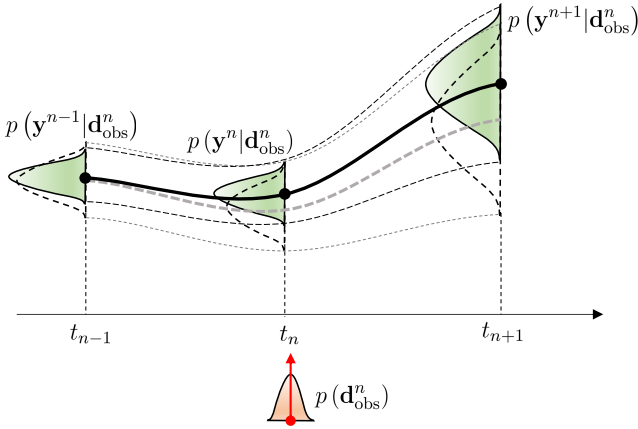
Before delving into the EnKF, it is important to discuss its famous predecessor: the Kalman filter (KF), named after Rudolf E. Kálmán, one of its primary developers. The KF finds widespread applications in guidance, navigation, and vehicle control systems, as well as in time series analysis, signal processing, and econometrics.

¹ Note that N_n denotes the number of data points at time t_n , which is different from our previous notation where N_d denotes the total number of data points, including all time steps.

² The distinction between filtering, smoothing, and forecasting is not very common in the reservoir data assimilation literature, but these terms are used frequently in other fields. Moreover, this terminology helps us to understand the names given to some of the methods discussed in this chapter. The attentive reader may have noticed that we used the term “forecasting” instead of “prediction.” It appears that these words are used interchangeably in the literature. In fact, the original Kalman filter paper by Kalman [228] uses the word “prediction.” However, we are going to make a distinction between the two words. We use “prediction” as a more general term; for example, we use “prediction” for a state generated from a model. In this sense, we can use a model to predict a past state, sometimes referred to as “hindcasting” [99]. We can also use the model to predict the present, which is referred to as “nowcasting” [338]. We reserve the word “forecasting” for the prediction of future states.



(a) Prior



(b) Posterior

Fig. 4.1: Illustration of the distribution of a state variable before (prior) and after (posterior) the assimilation of a datum at t_n . The posterior distributions indicated in (b), $p(\mathbf{y}^{n-1} | \mathbf{d}_{\text{obs}}^n)$, $p(\mathbf{y}^n | \mathbf{d}_{\text{obs}}^n)$, and $p(\mathbf{y}^{n+1} | \mathbf{d}_{\text{obs}}^n)$ correspond to the solutions of the smoothing, filtering and forecasting problems, respectively.

From a theoretical standpoint, the main assumptions of the KF are that the underlying dynamical system is linear and that all error terms follow multivariate Gaussian distributions.

4.2.1 Kalman Filter as a Minimum Variance Estimator

In this section, the KF is derived as the best linear unbiased estimate for the combined parameter-state vector $\mathbf{y}$. To reduce wording, we will simply refer to this vector as the state vector. First, we assume that the unknown true state at time t_n , denoted by $\mathbf{y}_{\text{true}}^n$, is given by:

$$\mathbf{y}_{\text{true}}^n = \mathbf{F}_n \mathbf{y}_{\text{true}}^{n-1} + \mathbf{e}_y^n. \quad (4.3)$$

Here, $\mathbf{F}_n$ is an $N_y \times N_y$ matrix representing the linear forward model, and $\mathbf{e}_y^n$ is a Gaussian model error. In the following, we will either neglect the model error $\mathbf{e}_y^n$ or assume an error term $\mathbf{e}^n$ that combines the effects of both model and measurement errors, as discussed in Chapter 2.

Similarly, let the vector of observations at t_n be represented by

$$\mathbf{d}_{\text{obs}}^n = \mathbf{H}_n \mathbf{y}_{\text{true}}^n + \mathbf{e}^n, \quad (4.4)$$

where $\mathbf{H}_n$ is an $N_n \times N_y$ matrix and $\mathbf{e}^n \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{e}^n})$ is the Gaussian data-error vector. $\mathbf{H}_n$ is often referred to as an observation matrix.

The forecast state vector at t_n is written as

$$\mathbf{y}^{n,f} = \mathbf{F}_n \mathbf{y}^{n-1,a}, \quad (4.5)$$

where the superscripts f and a denote forecast and analysis, respectively. Note that Eq. (4.5) indicates that the forecast state at time t_n depends on the analyzed (or updated) state at time t_{n-1} only.

We now define the innovation vector as

$$\delta \mathbf{d}^n = \mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \mathbf{y}^{n,f}, \quad (4.6)$$

and the forecast error vector as

$$\delta \mathbf{y}^{n,f} = \mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}. \quad (4.7)$$

Before we continue, it is convenient to show that the forecast state is unbiased, i.e., the expectation of the forecast error is zero

$$\mathbb{E} [\delta \mathbf{y}^{n,f}] = \mathbb{E} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}] = \mathbf{0}. \quad (4.8)$$

To show (4.8), it is necessary to assume that at t_0 the true state is a sample from the prior, $\mathbf{y}_{\text{true}}^0 \sim \mathcal{N}(\mathbf{y}^{0,a}, \mathbf{C}_{\mathbf{y}^0}^a)$. This is the same as assuming that we have an unbiased prior estimate of the truth. At time t_1 the true and forecast estimates are

$$\mathbf{y}_{\text{true}}^1 = \mathbf{F}_1 \mathbf{y}_{\text{true}}^0, \quad (4.9)$$

and

$$\mathbf{y}^{1,f} = \mathbf{F}_1 \mathbf{y}^{0,a}. \quad (4.10)$$

Hence

$$\begin{aligned}
 \mathbb{E} [\delta \mathbf{y}^{1,f}] &= \mathbb{E} [\mathbf{y}_{\text{true}}^1 - \mathbf{y}^{1,f}] \\
 &= \mathbb{E} [\mathbf{F}_1 (\mathbf{y}_{\text{true}}^0 - \mathbf{y}^{0,a})] \\
 &= \mathbf{F}_1 (\mathbb{E} [\mathbf{y}_{\text{true}}^0] - \mathbb{E} [\mathbf{y}^{0,a}]) \\
 &= \mathbf{F}_1 (\mathbf{y}_{\text{true}}^0 - \mathbf{y}_{\text{true}}^0) = \mathbf{0}.
 \end{aligned} \tag{4.11}$$

Now by mathematical induction³, we assume that the same is valid for t_{n-1} and conclude that

$$\begin{aligned}
 \mathbb{E} [\delta \mathbf{y}^{n,f}] &= \mathbb{E} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}] \\
 &= \mathbb{E} [\mathbf{F}_n (\mathbf{y}_{\text{true}}^{n-1} - \mathbf{y}^{n-1,a})] \\
 &= \mathbf{F}_n (\mathbb{E} [\mathbf{y}_{\text{true}}^{n-1}] - \mathbb{E} [\mathbf{y}^{n-1,a}]) \\
 &= \mathbf{F}_n (\mathbf{y}_{\text{true}}^{n-1} - \mathbf{y}_{\text{true}}^{n-1}) = \mathbf{0},
 \end{aligned} \tag{4.12}$$

which completes the proof that the forecasts state is unbiased.

We can also show that the expectation of the innovation vector is zero

$$\begin{aligned}
 \mathbb{E} [\delta \mathbf{d}^n] &= \mathbb{E} [\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \mathbf{y}^{n,f}] \\
 &= \mathbb{E} [\mathbf{H}_n \mathbf{y}_{\text{true}}^n + \mathbf{e}^n - \mathbf{H}_n \mathbf{y}^{n,f}] \\
 &= \mathbf{H}_n \mathbb{E} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}] + \mathbb{E} [\mathbf{e}^n] = \mathbf{0}.
 \end{aligned} \tag{4.13}$$

Now we define the analyzed state vector as

$$\mathbf{y}^{n,a} = \mathbf{y}^{n,f} + \mathbf{K}_n \delta \mathbf{d}^n, \tag{4.14}$$

where $\mathbf{K}_n$ is the so-called Kalman gain.

We define the analyzed error vector as

$$\delta \mathbf{y}^{n,a} = \mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,a}. \tag{4.15}$$

Note that the analysis (Eq. (4.14)) is unbiased because

$$\begin{aligned}
 \mathbb{E} [\delta \mathbf{y}^{n,a}] &= \mathbb{E} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,a}] \\
 &= \mathbb{E} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f} - \mathbf{K}_n \delta \mathbf{d}^n] \\
 &= \mathbb{E} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}] - \mathbf{K}_n \mathbb{E} [\delta \mathbf{d}^n] = \mathbf{0}.
 \end{aligned} \tag{4.16}$$

The forecast error-covariance is defined as

$$\begin{aligned}
 \mathbb{C} [\delta \mathbf{y}^{n,f}] &= \mathbb{E} [(\delta \mathbf{y}^{n,f} - \mathbb{E} [\delta \mathbf{y}^{n,f}]) (\delta \mathbf{y}^{n,f} - \mathbb{E} [\delta \mathbf{y}^{n,f}])^\top] \\
 &= \mathbb{E} [\delta \mathbf{y}^{n,f} (\delta \mathbf{y}^{n,f})^\top].
 \end{aligned} \tag{4.17}$$

³ Mathematical induction is a technique used to prove statements about ordered sets. In mathematical induction, we first show that the statement holds for $n = 1$, then we assume it holds for $n - 1$ and show it holds for n .

Note that this error-covariance is the same as the covariance of the forecast state because

$$\begin{aligned}
\mathbb{C} [\delta \mathbf{y}^{n,f}] &= \mathbb{C} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}] \\
&= \mathbb{E} [(\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}]) \times \\
&\quad (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}])^\top] \\
&= \mathbb{E} [(\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f} - \mathbf{y}_{\text{true}}^n + \mathbb{E} [\mathbf{y}^{n,f}]) \times \\
&\quad (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f} - \mathbf{y}_{\text{true}}^n + \mathbb{E} [\mathbf{y}^{n,f}])^\top] \\
&= \mathbb{E} [(\mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}^{n,f}]) (\mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}^{n,f}])^\top] \\
&= \mathbb{C} [\mathbf{y}^{n,f}] \\
&= \mathbf{C}_{\mathbf{y}^n}^f.
\end{aligned} \tag{4.18}$$

Similarly, the analyzed state/error-covariance is defined as

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}^n}^a &= \mathbb{C} [\delta \mathbf{y}^{n,a}] \\
&= \mathbb{E} [\delta \mathbf{y}^{n,a} (\delta \mathbf{y}^{n,a})^\top].
\end{aligned} \tag{4.19}$$

Finally, we can write the covariance of the innovation vector using

$$\begin{aligned}
\mathbf{C}_{\delta \mathbf{d}^n} &= \mathbb{C} [\delta \mathbf{d}^n] \\
&= \mathbb{E} [(\delta \mathbf{d}^n - \mathbb{E} [\delta \mathbf{d}^n]) (\delta \mathbf{d}^n - \mathbb{E} [\delta \mathbf{d}^n])^\top] \\
&= \mathbb{E} [\delta \mathbf{d}^n (\delta \mathbf{d}^n)^\top] \\
&= \mathbb{E} [(\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \mathbf{y}^{n,f}) (\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \mathbf{y}^{n,f})^\top] \\
&= \mathbb{E} [(\mathbf{H}_n \mathbf{y}_{\text{true}}^n + \mathbf{e}^n - \mathbf{H}_n \mathbf{y}^{n,f}) (\mathbf{H}_n \mathbf{y}_{\text{true}}^n + \mathbf{e}^n - \mathbf{H}_n \mathbf{y}^{n,f})^\top] \\
&= \mathbb{E} [(\mathbf{H}_n (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}) + \mathbf{e}^n) (\mathbf{H}_n (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}) + \mathbf{e}^n)^\top] \\
&= \mathbb{E} [(\mathbf{H}_n (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}) + \mathbf{e}^n) ((\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f})^\top \mathbf{H}_n^\top + (\mathbf{e}^n)^\top)] \\
&= \mathbb{E} [\mathbf{H}_n (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}) (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f})^\top \mathbf{H}_n^\top] + \\
&\quad \mathbb{E} [\mathbf{H}_n (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}) (\mathbf{e}^n)^\top] + \\
&\quad \mathbb{E} [\mathbf{e}^n (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f})^\top] + \mathbb{E} [\mathbf{e}^n (\mathbf{e}^n)^\top] \\
&= \mathbf{H}_n \mathbb{E} [(\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}) (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f})^\top] \mathbf{H}_n^\top + \mathbb{E} [\mathbf{e}^n (\mathbf{e}^n)^\top] \\
&= \mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top + \mathbf{C}_{\mathbf{e}^n}.
\end{aligned} \tag{4.20}$$

We seek for a matrix $\mathbf{K}_n$ that minimizes the expectation of the norm of the analyzed error vector, $\|\delta\mathbf{y}^{n,a}\|_2^2$, which is given by

$$\begin{aligned}
\mathbb{E} [\|\delta\mathbf{y}^{n,a}\|_2^2] &= \mathbb{E} \left[(\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,a})^\top (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,a}) \right] \\
&= \mathbb{E} \left[\sum_{i=1}^{N_y} (y_{\text{true},i}^n - y_i^{n,a})^2 \right] \\
&= \sum_{i=1}^{N_y} \mathbb{V} [y_{\text{true},i}^n - y_i^{n,a}] \\
&= \text{tr} (\mathbf{C}_{\mathbf{y}^n}^a). \tag{4.21}
\end{aligned}$$

Therefore, we need to write an expression for the analyzed covariance matrix $\mathbf{C}_{\mathbf{y}^n}^a$

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}^n}^a &= \mathbb{C} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,a}] \\
&= \mathbb{C} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f} - \mathbf{K}_n (\mathbf{H}_n \mathbf{y}_{\text{true}}^n + \mathbf{e}^n - \mathbf{H}_n \mathbf{y}^{n,f})] \\
&= \mathbb{C} [(\mathbf{I} - \mathbf{K}_n \mathbf{H}_n) (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}) - \mathbf{K}_n \mathbf{e}^n]. \tag{4.22}
\end{aligned}$$

Noting that the random vectors $\mathbf{e}^n$ and $(\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f})$ are independent and that the expectations of both these terms are zero, we can demonstrate, using the usual technique of expanding expectations that

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}^n}^a &= \mathbb{E} \left[(\mathbf{I} - \mathbf{K}_n \mathbf{H}_n) (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}) (\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f})^\top (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n)^\top \right] + \\
&\quad \mathbb{E} \left[\mathbf{K}_n \mathbf{e}^n (\mathbf{e}^n)^\top \mathbf{K}_n^\top \right] \\
&= (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n) \mathbb{C} [\mathbf{y}_{\text{true}}^n - \mathbf{y}^{n,f}] (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n)^\top + \mathbf{K}_n \mathbb{C} [\mathbf{e}^n] \mathbf{K}_n^\top \\
&= (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n) \mathbf{C}_{\mathbf{y}^n}^f (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n)^\top + \mathbf{K}_n \mathbf{C}_{\mathbf{e}^n} \mathbf{K}_n^\top \\
&= (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n) \left(\mathbf{C}_{\mathbf{y}^n}^f - \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top \mathbf{K}_n^\top \right) + \mathbf{K}_n \mathbf{C}_{\mathbf{e}^n} \mathbf{K}_n^\top \\
&= \mathbf{C}_{\mathbf{y}^n}^f - \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top \mathbf{K}_n^\top - \mathbf{K}_n \mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f + \mathbf{K}_n \mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top \mathbf{K}_n^\top + \mathbf{K}_n \mathbf{C}_{\mathbf{e}^n} \mathbf{K}_n^\top \\
&= \mathbf{C}_{\mathbf{y}^n}^f - \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top \mathbf{K}_n^\top - \mathbf{K}_n \mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f + \mathbf{K}_n \left(\mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top + \mathbf{C}_{\mathbf{e}^n} \right) \mathbf{K}_n^\top \\
&= \mathbf{C}_{\mathbf{y}^n}^f - \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top \mathbf{K}_n^\top - \mathbf{K}_n \mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f + \mathbf{K}_n \mathbf{C}_{\delta \mathbf{d}^n} \mathbf{K}_n^\top. \tag{4.23}
\end{aligned}$$

Taking the derivative⁴ of $\text{tr} (\mathbf{C}_{\mathbf{y}^n}^a)$ with respect to $\mathbf{K}_n$ gives

⁴ Appendix A (Section A.3) presents a basic review of matrix calculus.

$$\begin{aligned}
\frac{\partial \text{tr}(\mathbf{C}_{\mathbf{y}^n}^a)}{\partial \mathbf{K}_n} &= -\mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top - \left(\mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \right)^\top + 2\mathbf{K}_n \mathbf{C}_{\delta \mathbf{d}^n} \\
&= -2\mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top + 2\mathbf{K}_n \mathbf{C}_{\delta \mathbf{d}^n}.
\end{aligned} \tag{4.24}$$

Setting this derivative equal to zero and solving for $\mathbf{K}_n$ leads to

$$\begin{aligned}
\mathbf{K}_n &= \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top \mathbf{C}_{\delta \mathbf{d}^n}^{-1} \\
&= \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top \left(\mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top + \mathbf{C}_{\mathbf{e}^n} \right)^{-1}.
\end{aligned} \tag{4.25}$$

Eq. (4.25) is the Kalman gain matrix originally derived by Kalman [228]. In statistics, this type of estimator is known as BLUE (best linear unbiased estimator), with “best” referring to the estimator that results in the lowest variance of the estimate. The literature on the KF is extensive, given its wide range of applications across various fields. The derivation and discussions presented here are confined to the key concepts necessary to introduce and motivate the development of ensemble-based variants.

Remark 4.1: BLUE

BLUE (best linear unbiased estimator) is a statistical concept used to describe an estimator that satisfies the following three properties:

1. **Linearity:** The estimator is a linear function of the observed data.
2. **Unbiasedness:** The expected value of the estimator equals the true parameter being estimated. In other words, on average, the estimator is correct.
3. **Best:** Among all linear and unbiased estimators, the BLUE has the minimum variance.

4.2.1.1 Compact Notation

In the Kalman gain, the forecast state covariance matrix, $\mathbf{C}_{\mathbf{y}^n}^f$, appears multiplied by the transpose of the observation matrix, $\mathbf{H}_n$. We can show that this product is equivalent to the cross-covariance between the state vector and predicted data, $\mathbf{C}_{\mathbf{y} \mathbf{d}^n}^f$,

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}\mathbf{d}^n}^f &= \mathbb{E} \left[(\mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}^{n,f}]) (\mathbf{d}^{n,f} - \mathbb{E} [\mathbf{d}^{n,f}])^\top \right] \\
&= \mathbb{E} \left[(\mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}^{n,f}]) (\mathbf{H}_n \mathbf{y}^{n,f} - \mathbb{E} [\mathbf{H}_n \mathbf{y}^{n,f}])^\top \right] \\
&= \mathbb{E} \left[(\mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}^{n,f}]) (\mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}^{n,f}])^\top \mathbf{H}_n^\top \right] \\
&= \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top.
\end{aligned} \tag{4.26}$$

Similarly, the product $\mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top$ is equivalent to the auto-covariance of predicted data $\mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f$ because

$$\begin{aligned}
\mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f &= \mathbb{E} \left[(\mathbf{d}^{n,f} - \mathbb{E} [\mathbf{d}^{n,f}]) (\mathbf{d}^{n,f} - \mathbb{E} [\mathbf{d}^{n,f}])^\top \right] \\
&= \mathbf{H}_n \mathbb{E} \left[(\mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}^{n,f}]) (\mathbf{y}^{n,f} - \mathbb{E} [\mathbf{y}^{n,f}])^\top \right] \mathbf{H}_n^\top \\
&= \mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top.
\end{aligned} \tag{4.27}$$

Therefore, we can write the KF analysis in a very compact and convenient form

$$\mathbf{y}^{n,a} = \mathbf{y}^{n,f} + \mathbf{C}_{\mathbf{y}\mathbf{d}^n}^f \left(\mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f + \mathbf{C}_{\mathbf{e}^n} \right)^{-1} (\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \mathbf{y}^{n,f}). \tag{4.28}$$

The KF can be summarized as the recursive application of two steps: *forecast* and *analysis*, which is indicated in Pseudo-code 4.1. Fig. 4.2 illustrates the process.

Pseudo-code 4.1: Kalman Filter

1. **Forecast:** given the best estimate of the true state at time t_{n-1} , denoted by $\mathbf{y}^{n-1,a}$, with uncertainty represented by the covariance $\mathbf{C}_{\mathbf{y}^{n-1}}^a$, we advance the linear dynamical system to the next data assimilation time step based on the forward model using

$$\mathbf{y}^{n,f} = \mathbf{F}_n \mathbf{y}^{n-1,a}, \quad (4.29)$$

and

$$\mathbf{C}_{\mathbf{y}^n}^f = \mathbf{F}_n \mathbf{C}_{\mathbf{y}^{n-1}}^a \mathbf{F}_n^\top. \quad (4.30)$$

2. **Analysis:** given the a set of observations at time t_n , we update our estimates of the state and covariance using

$$\mathbf{y}^{n,a} = \mathbf{y}^{n,f} + \mathbf{K}_n (\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \mathbf{y}^{n,f}), \quad (4.31)$$

and

$$\mathbf{C}_{\mathbf{y}^n}^a = (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n) \mathbf{C}_{\mathbf{y}^n}^f, \quad (4.32)$$

where

$$\mathbf{K}_n = \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top (\mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top + \mathbf{C}_{\mathbf{e}^n})^{-1}. \quad (4.33)$$

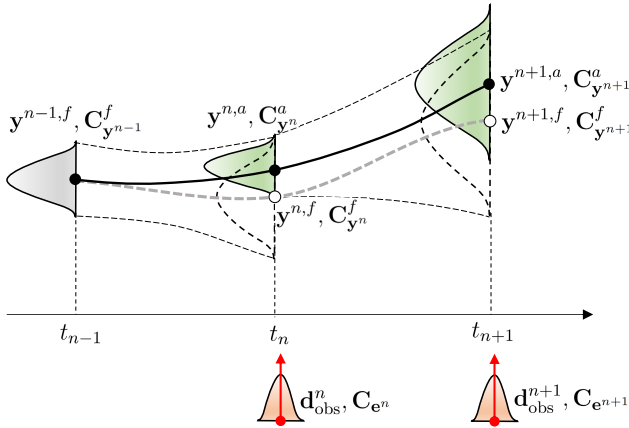


Fig. 4.2: Illustration of the data assimilation sequence of the Kalman filter. In the KF the PDF of the state vector at each time step is Gaussian. In this figure, the dashed PDFs correspond to the forecast and the solid PDF the analysis.

4.2.1.2 Example: Porosity of a Rock Sample

Consider the same problem discussed in Section 2.3.1.1, where the objective is to estimate the porosity of a rock sample by combining prior information, $\phi \sim \mathcal{N}(\phi_{\text{pr}}, \sigma_{\text{pr}}^2)$, with a measurement, $d_{\text{obs}} \sim \mathcal{N}(\phi_{\text{obs}}, \sigma_e^2)$.

We can solve this problem by simply applying the KF equations. However, for the sake of the exercise, let's derive the KF equations directly for this problem. First, we express our best prior estimate of the porosity as

$$\phi_f = \phi_{\text{pr}} = \phi_{\text{true}} + \delta\phi_f, \quad (4.34)$$

where $\delta\phi_f$ represents the forecast error. We express the observed data as

$$\phi_{\text{obs}} = \phi_{\text{true}} + e, \quad (4.35)$$

where e is the measurement error.

We do not know the values of $\delta\phi_f$ and e , but we make the following assumptions about their statistics:

1. $\mathbb{E}[\delta\phi_f] = 0$: unbiased prior estimate.
2. $\mathbb{E}[e] = 0$: unbiased measurement errors.
3. $\mathbb{E}[\delta\phi_f e] = 0$: independent errors.
4. $\mathbb{V}[\delta\phi_f] = \sigma_f^2$: we know the prior variance.
5. $\mathbb{V}[e] = \sigma_e^2$: we know the variance of the measurement error.

We are looking for a linear estimator

$$\phi_a = k_1\phi_f + k_2\phi_{\text{obs}}, \quad (4.36)$$

which is unbiased

$$\begin{aligned} \mathbb{E}[\phi_a] &= \mathbb{E}[k_1\phi_f + k_2\phi_{\text{obs}}] = \phi_{\text{true}} \\ &= \mathbb{E}[k_1(\phi_{\text{true}} + \delta\phi_f) + k_2(\phi_{\text{true}} + e)] = \phi_{\text{true}} \\ &= \mathbb{E}[k_1\phi_{\text{true}} + k_2\phi_{\text{true}}] = \phi_{\text{true}}. \end{aligned} \quad (4.37)$$

Therefore

$$k_1 + k_2 = 1. \quad (4.38)$$

Our linear estimator becomes

$$\phi_a = (1 - k_2)\phi_f + k_2\phi_{\text{obs}}, \quad (4.39)$$

which is clearly a weighted average between ϕ_f and ϕ_{obs} . An expression for the error associated with the new estimate can be derived by noting that

$$\begin{aligned} \phi_a &= \phi_{\text{true}} + \delta\phi_a \\ \phi_{\text{true}} + \delta\phi_a &= (1 - k_2)(\phi_{\text{true}} + \delta\phi_f) + k_2(\phi_{\text{true}} + e) \\ \delta\phi_a &= \delta\phi_f + k_2(e - \delta\phi_f). \end{aligned} \quad (4.40)$$

This estimate is also unbiased

$$\mathbb{E}[\delta\phi_a] = \mathbb{E}[\delta\phi_f + k_2(e - \delta\phi_f)] = 0. \quad (4.41)$$

The variance of the error in the estimate is

$$\begin{aligned} \mathbb{V}[\delta\phi_a] &= \mathbb{E}[(\delta\phi_a - \mathbb{E}[\delta\phi_a])^2] \\ &= \mathbb{E}[\delta\phi_a^2] \\ &= \mathbb{E}[(\delta\phi_f + k_2(e - \delta\phi_f))^2] \\ &= \mathbb{E}[\delta\phi_f^2 - 2k_2\delta\phi_f^2 + k_2^2(e^2 + \delta\phi_f^2) + 2k_2\delta\phi_f e - 2k_2^2 e\delta\phi_f] \\ &= \mathbb{E}[\delta\phi_f^2] - 2k_2\mathbb{E}[\delta\phi_f^2] + k_2^2\mathbb{E}[(e^2 + \delta\phi_f^2)] + 2k_2\mathbb{E}[\delta\phi_f e] - \\ &\quad 2k_2^2\mathbb{E}[e\delta\phi_f] \\ &= \sigma_f^2 - 2k_2\sigma_f^2 + k_2^2(\sigma_e^2 + \sigma_f^2) \\ &= \sigma_a^2. \end{aligned} \quad (4.42)$$

To minimize the variance of the new estimate, we simply compute the derivative and set to zero

$$\frac{d\sigma_a^2}{dk_2} = -2\sigma_f^2 + 2k_2(\sigma_e^2 + \sigma_f^2) = 0. \quad (4.43)$$

Now solving for k_2 , which corresponds to the Kalman gain, we obtain

$$k_2 = \frac{\sigma_f^2}{\sigma_e^2 + \sigma_f^2}. \quad (4.44)$$

The porosity estimate becomes

$$\phi_a = \phi_f + \frac{\sigma_f^2}{\sigma_e^2 + \sigma_f^2}(\phi_{\text{obs}} - \phi_f), \quad (4.45)$$

with variance given by

$$\sigma_a^2 = \sigma_f^2 - \frac{\sigma_f^4}{\sigma_e^2 + \sigma_f^2}. \quad (4.46)$$

4.2.2 Kalman Filter as a Bayesian Estimator

We presented the derivation of the KF as a minimum variance estimator. However, the KF equations can also be obtained from Bayes' rule. In fact, when expressing the KF analysis equations in terms of the vector of model parameters, $\mathbf{m}$, and utilizing the Kalman gain definition, we arrive at the following equations:

$$\mathbf{m}^{n,a} = \mathbf{m}^{n,f} + \mathbf{C}_{\mathbf{m}^n}^f \mathbf{G}_n^\top \left(\mathbf{G}_n \mathbf{C}_{\mathbf{m}^n}^f \mathbf{G}_n^\top + \mathbf{C}_{\mathbf{e}^n} \right)^{-1} (\mathbf{d}_{\text{obs}}^n - \mathbf{G}_n \mathbf{m}^{n,f}) \quad (4.47)$$

and

$$\mathbf{C}_{\mathbf{m}^n}^a = \mathbf{C}_{\mathbf{m}^n}^f - \mathbf{C}_{\mathbf{m}^n}^f \mathbf{G}_n^\top \left(\mathbf{G}_n \mathbf{C}_{\mathbf{m}^n}^f \mathbf{G}_n^\top + \mathbf{C}_{\mathbf{e}^n} \right)^{-1} \mathbf{G}_n \mathbf{C}_{\mathbf{m}^n}^f. \quad (4.48)$$

We replaced $\mathbf{H}_n$ with $\mathbf{G}_n$ to denote the sensitivity matrix of the predicted data with respect to the vector of model parameters, $\mathbf{m}$, to be consistent with our previous notation. Noting that our best estimate of $\mathbf{m}$ at time zero is the prior mean, i.e., $\mathbf{m}^{0,f} = \mathbf{m}_{\text{pr}}$, and the covariance is also the prior, $\mathbf{C}_{\mathbf{m}^0}^f = \mathbf{C}_{\mathbf{m}}$, we obtain:

$$\mathbf{m}^{0,a} = \mathbf{m}_{\text{map}} \quad (4.49)$$

and

$$\mathbf{C}_{\mathbf{m}^0}^a = \mathbf{C}_{\mathbf{m}_c}. \quad (4.50)$$

Assuming that measurement and model errors are uncorrelated in time and that the model's evolution is a first-order Markov process, we can show that sequential and simultaneous data assimilation are equivalent, as discussed in Section 4.2.3. In this case, the result presented above holds for any time t_n .

4.2.2.1 Example: Porosity of a Rock Sample

We can solve the same problem of Example 4.2.1.2 starting from the Bayes' rule. In this case, we assume a Gaussian prior distribution for porosity

$$p(\phi) \propto \exp \left[-\frac{1}{2} \frac{(\phi - \phi_{\text{pr}})^2}{\sigma_{\text{pr}}^2} \right] \quad (4.51)$$

and a Gaussian Likelihood

$$\mathcal{L}(\phi) \propto \exp \left[-\frac{1}{2} \frac{(\phi_{\text{obs}} - \phi)^2}{\sigma_e^2} \right]. \quad (4.52)$$

Using Bayes' rule, the posterior becomes

$$\begin{aligned}
p(\phi|\phi_{\text{obs}}) &\propto \mathcal{L}(\phi)p(\phi) \\
&\propto \exp\left[-\frac{1}{2}\frac{(\phi_{\text{obs}} - \phi)^2}{\sigma_e^2}\right] \exp\left[-\frac{1}{2}\frac{(\phi - \phi_{\text{pr}})^2}{\sigma_{\text{pr}}^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\frac{(\phi_{\text{obs}} - \phi)^2}{\sigma_e^2} + \frac{(\phi - \phi_{\text{pr}})^2}{\sigma_{\text{pr}}^2}\right)\right] \tag{4.53}
\end{aligned}$$

We can expand the squared terms

$$\begin{aligned}
p(\phi|\phi_{\text{obs}}) &\propto \exp\left[-\frac{1}{2}\left(\frac{\phi_{\text{obs}}^2 - 2\phi\phi_{\text{obs}} + \phi^2}{\sigma_e^2} + \frac{\phi^2 - 2\phi\phi_{\text{pr}} + \phi_{\text{pr}}^2}{\sigma_{\text{pr}}^2}\right)\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\frac{\phi_{\text{obs}}^2\sigma_{\text{pr}}^2 - 2\phi\phi_{\text{obs}}\sigma_{\text{pr}}^2 + \phi^2\sigma_{\text{pr}}^2 + \phi^2\sigma_e^2 - 2\phi\phi_{\text{pr}}\sigma_e^2 + \phi_{\text{pr}}^2\sigma_e^2}{\sigma_e^2\sigma_{\text{pr}}^2}\right)\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\frac{\phi^2(\sigma_{\text{pr}}^2 + \sigma_e^2)}{\sigma_e^2\sigma_{\text{pr}}^2} - \frac{2\phi(\phi_{\text{obs}}\sigma_{\text{pr}}^2 + \phi_{\text{pr}}\sigma_e^2)}{\sigma_e^2\sigma_{\text{pr}}^2} + \frac{\phi_{\text{obs}}^2\sigma_{\text{pr}}^2 + \phi_{\text{pr}}^2\sigma_e^2}{\sigma_e^2\sigma_{\text{pr}}^2}\right)\right]. \tag{4.54}
\end{aligned}$$

Because the prior Gaussian and the model is linear, the posterior is also Gaussian. Therefore

$$\begin{aligned}
p(\phi|\phi_{\text{obs}}) &\propto \exp\left[-\frac{1}{2}\frac{(\phi - \phi_{\text{map}})^2}{\sigma_{\phi_c}^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\frac{\phi^2 - 2\phi\phi_{\text{map}} + \phi_{\text{map}}^2}{\sigma_{\phi_c}^2}\right)\right]. \tag{4.55}
\end{aligned}$$

We have two expressions for $p(\phi|\phi_{\text{obs}})$. In order for these expressions to be equivalent, the coefficients multiplying ϕ^2 must be the same. Hence

$$\frac{1}{\sigma_{\phi_c}^2} = \frac{\sigma_{\text{pr}}^2 + \sigma_e^2}{\sigma_e^2\sigma_{\text{pr}}^2}, \tag{4.56}$$

which lead to

$$\sigma_{\phi_c}^2 = \frac{\sigma_e^2\sigma_{\text{pr}}^2}{\sigma_{\text{pr}}^2 + \sigma_e^2}. \tag{4.57}$$

Previously, we derived the following expression using the KF

$$\begin{aligned}
\sigma_a^2 &= \sigma_f^2 - \frac{\sigma_f^4}{\sigma_e^2 + \sigma_f^2} \\
&= \frac{\sigma_f^2 (\sigma_e^2 + \sigma_f^2) - \sigma_f^4}{\sigma_e^2 + \sigma_f^2} \\
&= \frac{\sigma_f^2 \sigma_e^2}{\sigma_e^2 + \sigma_f^2}.
\end{aligned} \tag{4.58}$$

Therefore, writing $\sigma_f = \sigma_{\text{pr}}$, we have $\sigma_{\phi_c} = \sigma_a$. Repeating the same procedure for the coefficients multiplying 2ϕ gives

$$\frac{\phi_{\text{map}}}{\sigma_{\phi_c}} = \frac{\phi_{\text{obs}}\sigma_{\text{pr}}^2 + \phi_{\text{pr}}\sigma_e^2}{\sigma_e^2\sigma_{\text{pr}}^2}. \tag{4.59}$$

Hence,

$$\begin{aligned}
\phi_{\text{map}} &= \frac{\phi_{\text{obs}}\sigma_{\text{pr}}^2 + \phi_{\text{pr}}\sigma_e^2}{\sigma_e^2 + \sigma_{\text{pr}}^2} \\
&= \frac{\phi_{\text{obs}}\sigma_{\text{pr}}^2 + \phi_{\text{pr}}\sigma_e^2 + \phi_{\text{pr}}\sigma_{\text{pr}}^2 - \phi_{\text{pr}}\sigma_{\text{pr}}^2}{\sigma_e^2 + \sigma_{\text{pr}}^2} \\
&= \phi_{\text{pr}} + \frac{\sigma_{\text{pr}}^2}{\sigma_e^2 + \sigma_{\text{pr}}^2} (\phi_{\text{obs}} - \phi_{\text{pr}}).
\end{aligned} \tag{4.60}$$

Again, writing $\phi_{\text{pr}} = \phi_f$ and $\sigma_f = \sigma_{\text{pr}}$, we have $\phi_{\text{map}} = \phi_a$. We have arrived to at the same expressions starting from the assumption of BLUE and from Bayes' rule with Gaussian prior and likelihood.

4.2.2.2 Example: Porosity Distribution in a Core Sample

Consider the same problem discussed in Section 2.3.1.4, where the objective is to estimate the porosity distribution in a rock core discretized in 50 gridblocks. Before, we obtained the MAP estimate and RML samples by assimilating 10 data points simultaneously. Now, we use the KF and assimilate data sequentially, mimicking “pseudo time steps.” Fig. 4.3 shows the evolution of the data assimilation process, indicating the best estimate after the assimilation of each datum and the corresponding uncertainty range. The estimated mean and standard deviation after the assimilation of the 10th data point are exactly the same as presented in Fig. 2.2b.

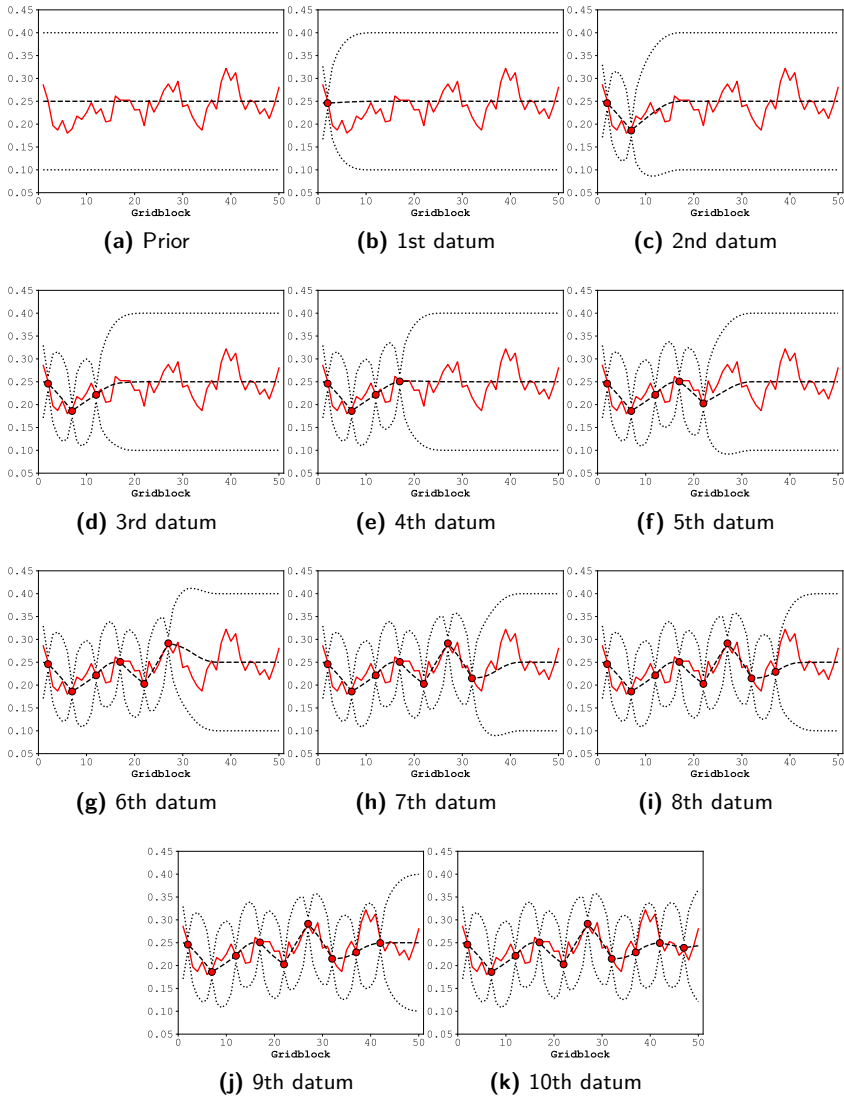


Fig. 4.3: KF applied to estimate the porosity in a core sample. The solid red line represents the ground truth. The red circles represent the observed data points. The dashed lines represent the best estimate, and the dotted lines represent the error bands corresponding to three standard deviations. Example 4.2.2.2.

4.2.3 Sequential Data Assimilation

Sequential data assimilation (SDA) refers to the process of integrating measured data into a dynamical system sequentially over a series of discrete time steps. The fundamental condition for SDA is that the system's evolution can be modeled as a first-order Markov process, meaning that the state at time t_n depends solely on the state at the previous time step, i.e.,

$$\mathbf{y}^n = \mathbf{f}(\mathbf{y}^{n-1}, \mathbf{y}^{n-2}, \dots, \mathbf{y}^0) = \mathbf{f}(\mathbf{y}^{n-1}). \quad (4.61)$$

In fact, when we derived the KF, we wrote the forecast state vector as

$$\mathbf{y}^{n,f} = \mathbf{F}_n \mathbf{y}^{n-1,a}. \quad (4.62)$$

Note that the first-order Markov condition can hold even for nonlinear problems⁵.

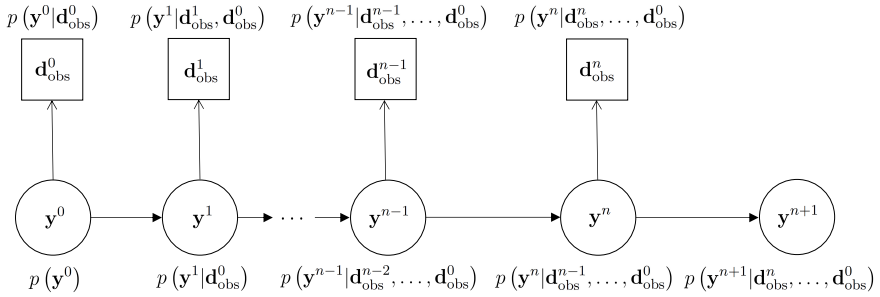


Fig. 4.4: Sequential Bayesian data assimilation. The horizontal arrows indicate the evolution of the dynamical system, and the vertical arrows the data-conditioning steps.

SDA fits in the Bayesian framework as illustrated in Fig. 4.4, which depicts the main elements. In this process, sequential and simultaneous data assimilation are equivalent. In order to show this equivalence, first assume that at the time t_{n-1} , we have updated the distribution of $\mathbf{y}^n$ by incorporating all data available up to the time t_{n-1} simultaneously. In this case, we write the posterior PDF as

$$p(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1}) \propto p(\mathbf{y}^n) p(\hat{\mathbf{d}}_{\text{obs}}^{n-1} | \mathbf{y}^n), \quad (4.63)$$

where $p(\mathbf{y}^n)$ represents our prior knowledge about the distribution of $\mathbf{y}$ at time t_n . In the above expression, we use a hat over the observed data vector

⁵ It is worth noting that the first-order Markov condition holds in reservoir simulations, where the state (say, pressure and saturation) at time t_n is solely dependent on static parameters and the state at the preceding time step, t_{n-1} . This property is precisely why we can employ restarts in reservoir simulators.

to denote that it contains all the observations up to time t_{n-1} , i.e.,

$$\hat{\mathbf{d}}_{\text{obs}}^{n-1} = \begin{bmatrix} \mathbf{d}_{\text{obs}}^{n-1} \\ \mathbf{d}_{\text{obs}}^{n-2} \\ \vdots \\ \mathbf{d}_{\text{obs}}^0 \end{bmatrix}. \quad (4.64)$$

Now if we want to incorporate a new set of observations at time t_n , we could start from scratch and write

$$\underbrace{p(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^n)}_{\text{post}} \propto \underbrace{p(\mathbf{y}^n)}_{\text{prior}} \underbrace{p(\hat{\mathbf{d}}_{\text{obs}}^n | \mathbf{y}^n)}_{\text{likelihood}}. \quad (4.65)$$

Eq. (4.65) is the standard Bayes' rule for the PDF of state $\mathbf{y}$ at time t_n conditioned to all observations from time zero until (and including) t_n . In other words, this is the posterior PDF obtained by simultaneous data assimilation.

Alternatively, we could claim that just before time t_n , the knowledge about $\mathbf{y}$ is summarized in the distribution $p(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1})$. Hence, we can just use $p(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1})$ as the distribution prior to the assimilation of $\mathbf{d}_{\text{obs}}^n$ and update the PDF of $\mathbf{y}$ as

$$\underbrace{p_s(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^n)}_{\text{sequential post}} \propto \underbrace{p(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1})}_{\text{"new" prior}} \underbrace{p(\mathbf{d}_{\text{obs}}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1}, \mathbf{y}^n)}_{\text{"new" likelihood}}. \quad (4.66)$$

We added a subscript "s" to denote that this is the posterior PDF obtained by assimilating data from t_0 to t_{n-1} simultaneously, and assimilating the data at t_n sequentially. We want to show that (4.65) and (4.66) are equivalent, which means

$$p_s(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^n) = p(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^n). \quad (4.67)$$

First, we write

$$\begin{aligned} p_s(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^n) &\propto p(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1}) p(\mathbf{d}_{\text{obs}}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1}, \mathbf{y}^n) \\ &\propto p(\mathbf{y}^n) p(\hat{\mathbf{d}}_{\text{obs}}^{n-1} | \mathbf{y}^n) p(\mathbf{d}_{\text{obs}}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1}, \mathbf{y}^n). \end{aligned} \quad (4.68)$$

Now, assuming that measurement and model errors are uncorrelated in time is equivalent to saying that

$$p(\mathbf{d}_{\text{obs}}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1}, \mathbf{y}^n) = p(\mathbf{d}_{\text{obs}}^n | \mathbf{y}^n) \quad (4.69)$$

and

$$p\left(\hat{\mathbf{d}}_{\text{obs}}^{n-1}|\mathbf{y}^n\right)p\left(\mathbf{d}_{\text{obs}}^n|\mathbf{y}^n\right)=p\left(\hat{\mathbf{d}}_{\text{obs}}^n|\mathbf{y}^n\right). \quad (4.70)$$

Using (4.70) in (4.68) results in

$$\begin{aligned} p_s\left(\mathbf{y}^n|\hat{\mathbf{d}}_{\text{obs}}^n\right) &\propto p\left(\mathbf{y}^n|\hat{\mathbf{d}}_{\text{obs}}^{n-1}\right)p\left(\mathbf{d}_{\text{obs}}^n|\hat{\mathbf{d}}_{\text{obs}}^{n-1},\mathbf{y}^n\right) \\ &\propto p\left(\mathbf{y}^n\right)p\left(\hat{\mathbf{d}}_{\text{obs}}^n|\mathbf{y}^n\right) \\ &= p\left(\mathbf{y}^n|\hat{\mathbf{d}}_{\text{obs}}^n\right). \end{aligned} \quad (4.71)$$

This completes the derivation of the equivalence between sequential and simultaneous data assimilation. Note that the demonstration was limited to time t_n , but it obviously holds for any time by mathematical induction.

Returning to the KF, in which case we assume that all PDFs are Gaussian, we can write that at time t_{n-1} , our knowledge about the state vector is summarized by the PDF

$$p\left(\mathbf{y}^{n-1,a}|\hat{\mathbf{d}}_{\text{obs}}^{n-1}\right)=\mathcal{N}\left(\mathbf{y}^{n-1,a},\mathbf{C}_{\mathbf{y}^{n-1}}^a\right). \quad (4.72)$$

Therefore, the two steps of the KF can be stated in the following Pseudo-code 4.2.

Pseudo-code 4.2: Kalman Filter (Bayesian Version)

1. **Forecast:** apply the forward model to obtain the PDF of the state at time t_n given the observations up to time t_{n-1} :

$$p\left(\mathbf{y}^{n,f}|\hat{\mathbf{d}}_{\text{obs}}^{n-1}\right)=p\left(\mathbf{F}_n\mathbf{y}^{n-1,a}|\hat{\mathbf{d}}_{\text{obs}}^{n-1}\right). \quad (4.73)$$

2. **Analysis:** update the PDF using Bayes' rule:

$$\underbrace{p\left(\mathbf{y}^{n,a}\right)}_{\text{post}} \equiv \underbrace{p\left(\mathbf{y}^{n,f}|\hat{\mathbf{d}}_{\text{obs}}^n\right)}_{\text{prior}} \propto \underbrace{p\left(\mathbf{d}_{\text{obs}}^n|\mathbf{y}^{n,f}\right)}_{\text{likelihood}}. \quad (4.74)$$

4.3 Ensemble Kalman Filter

In the KF, the mean and covariance of the state vector are updated whenever new data become available. However, even for linear problems, updating the covariance matrix using the KF equations can become computationally unfeasible when dealing with high-dimensional state vectors. Additionally, if the problem is nonlinear, the KF cannot be applied directly, in which case the extended Kalman filter (EKF) serves as an alternative. The EKF uses

linearizations of the model and observation equations around the estimated mean of the state. However, in highly nonlinear scenarios, these linearizations may lead to unbounded instabilities in the covariance updates [139, Chap. 4]. Furthermore, the EKF still requires updating the entire state covariance matrix, posing computational challenges, particularly in high-dimensional problems. For example, contemporary reservoir models consist of $O(10^6)$ gridblocks, resulting in a $\mathbf{C}_y$ matrix with a staggering $O(10^{12})$ entries, which is impractical to store in memory.

Evensen [136] introduced the ensemble Kalman filter (EnKF) as an alternative to address the issues encountered with the EKF in high-dimensional nonlinear dynamical systems. The EnKF is a Monte Carlo method in which an ensemble of states is updated over time and used to estimate the mean and covariance. While the EnKF was originally introduced in [136], the method was later refined in [51, 197], by introducing the concept of updating each state vector in the ensemble with independently perturbed observations, leading to the current implementation of the EnKF.

In the EnKF, uncertainty is represented and propagated using an ensemble of states. This contrasts with the KF and the EKF, where uncertainty is propagated through updates to the state covariance matrix. Furthermore, in practical applications of the EnKF, the number of state vectors in the ensemble is typically much smaller than the number of unknowns, making the ensemble update computationally more efficient than updating the state covariance matrix.

Since its introduction, the body of literature on EnKF has expanded significantly. The EnKF has found applications in several areas spanning disciplines such as oceanography [34, 236], atmospheric modeling [464, 200], numerical weather prediction [199, 425], hydrology [360, 82, 264, 418], and petroleum reservoir engineering [3, 331].

Reservoir data assimilation is typically defined as a parameter estimation problem. Therefore, in order to use EnKF in this context is convenient to write the equations in terms of the combined parameter-state vector $\mathbf{y}$. As before, hereafter, we refer to $\mathbf{y}$ simply as the state vector.

Assuming that the prior PDF for the state vector at time t_n is Gaussian and that the data errors also follow a Gaussian distribution, we can write

$$p(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^{n-1}) \propto \exp \left\{ -\frac{1}{2} (\mathbf{y}^n - \mathbf{y}^{n,f})^\top (\mathbf{C}_{\mathbf{y}^n}^f)^{-1} (\mathbf{y}^n - \mathbf{y}^{n,f}) \right\} \quad (4.75)$$

and

$$\mathcal{L}(\mathbf{y}^n | \mathbf{d}_{\text{obs}}^n) \propto \exp \left\{ -\frac{1}{2} (\mathbf{d}_{\text{obs}}^n - \mathbf{h}(\mathbf{y}^n))^\top \mathbf{C}_{\mathbf{e}^n}^{-1} (\mathbf{d}_{\text{obs}}^n - \mathbf{h}(\mathbf{y}^n)) \right\}, \quad (4.76)$$

where $\mathbf{h}(\cdot)$ denotes the nonlinear relationship between state and data, meaning that $\mathbf{h}(\mathbf{y}^n)$ is the predicted data vector at t_n . Hence, the posterior PDF becomes

$$p(\mathbf{y}^n | \hat{\mathbf{d}}_{\text{obs}}^n) \propto \exp\{-\mathcal{O}(\mathbf{y}^n)\}, \quad (4.77)$$

where

$$\begin{aligned} \mathcal{O}(\mathbf{y}^n) = & \frac{1}{2} (\mathbf{y}^n - \mathbf{y}^{n,f})^\top (\mathbf{C}_{\mathbf{y}^n}^f)^{-1} (\mathbf{y}^n - \mathbf{y}^{n,f}) + \\ & \frac{1}{2} (\mathbf{d}_{\text{obs}}^n - \mathbf{h}(\mathbf{y}^n))^\top \mathbf{C}_{\mathbf{e}^n}^{-1} (\mathbf{d}_{\text{obs}}^n - \mathbf{h}(\mathbf{y}^n)). \end{aligned} \quad (4.78)$$

However, because the state-data relationship $\mathbf{h}(\mathbf{y}^n)$ is nonlinear, an analytical minimum of $\mathcal{O}(\mathbf{y}^n)$ cannot be computed.

One way to derive the EnKF analysis equation is by introducing a “trick” to deal with the nonlinearity in $\mathcal{O}(\mathbf{y}^n)$ ⁶. The trick consists of augmenting the state vector by including the predicted data

$$\hat{\mathbf{y}}^n = \begin{bmatrix} \mathbf{y}^n \\ \mathbf{h}(\mathbf{y}^n) \end{bmatrix} = \begin{bmatrix} \mathbf{m} \\ \mathbf{x}^n \\ \mathbf{h}(\mathbf{y}^n) \end{bmatrix}. \quad (4.79)$$

Then, we define the observation matrix, $\mathbf{H}_n$, as

$$\mathbf{H}_n = [\mathbf{O}_{N_m} \quad \mathbf{O}_{N_x} \quad \mathbf{I}_{N_n}], \quad (4.80)$$

where $\mathbf{O}_{N_m}$ and $\mathbf{O}_{N_x}$ are null matrices with dimensions $N_n \times N_m$ and $N_n \times N_x$, respectively, while $\mathbf{I}_{N_n}$ is the $N_n \times N_n$ identity matrix. Using the augmented state vector, the predicted data vector can be written as

$$\mathbf{h}(\mathbf{y}^n) = \mathbf{H}_n \hat{\mathbf{y}}^n. \quad (4.81)$$

Now, we assume that the prior PDF for the augmented state vector is Gaussian

$$p(\hat{\mathbf{y}}^n | \mathbf{d}_{\text{obs}}^n) \propto \exp\{-\mathcal{O}(\hat{\mathbf{y}}^n)\}, \quad (4.82)$$

where

$$\begin{aligned} \mathcal{O}(\hat{\mathbf{y}}^n) = & \frac{1}{2} (\hat{\mathbf{y}}^n - \hat{\mathbf{y}}^{n,f})^\top (\mathbf{C}_{\hat{\mathbf{y}}^n}^f)^{-1} (\hat{\mathbf{y}}^n - \hat{\mathbf{y}}^{n,f}) + \\ & \frac{1}{2} (\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \hat{\mathbf{y}}^n)^\top \mathbf{C}_{\mathbf{e}^n}^{-1} (\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \hat{\mathbf{y}}^n). \end{aligned} \quad (4.83)$$

Because we wrote the predicted data vector as a linear function of $\hat{\mathbf{y}}^n$, we can find the posterior mean by setting $\nabla \mathcal{O}(\hat{\mathbf{y}}^n) = \mathbf{0}$ and solving for $\hat{\mathbf{y}}^n$. This procedure leads to

⁶ This trick disguises the problem of nonlinearities, but it does not avoid it [256].

$$\begin{aligned}
\hat{\mathbf{y}}^{n,a} &= \hat{\mathbf{y}}^{n,f} + \left[\left(\mathbf{C}_{\hat{\mathbf{y}}^n}^f \right)^{-1} + \mathbf{H}_n^\top \mathbf{C}_{\mathbf{e}^n}^{-1} \mathbf{H}_n \right]^{-1} \mathbf{H}_n^\top \mathbf{C}_{\mathbf{e}^n}^{-1} (\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \hat{\mathbf{y}}^{n,f}) \\
&= \hat{\mathbf{y}}^{n,f} + \mathbf{C}_{\hat{\mathbf{y}}^n}^f \mathbf{H}_n^\top \left(\mathbf{H}_n \mathbf{C}_{\hat{\mathbf{y}}^n}^f \mathbf{H}_n^\top + \mathbf{C}_{\mathbf{e}^n} \right)^{-1} (\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \hat{\mathbf{y}}^{n,f}) \\
&= \hat{\mathbf{y}}^{n,f} + \mathbf{K}_n (\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \hat{\mathbf{y}}^{n,f}). \tag{4.84}
\end{aligned}$$

Note that these expressions are equivalent to the expressions derived for the KF and the MAP estimate. However, here we are interested in sampling the posterior PDF. This can be accomplished using RML, in which case to obtain the j th sample of $p(\hat{\mathbf{y}}^n | \mathbf{d}_{\text{obs}}^n)$, we replace $\hat{\mathbf{y}}^{n,f}$ in the expression for $\mathcal{O}(\hat{\mathbf{y}}^n)$ by $\hat{\mathbf{y}}_j^{n,f} \sim \mathcal{N}(\hat{\mathbf{y}}^{n,f}, \mathbf{C}_{\hat{\mathbf{y}}^n}^f)$, and $\mathbf{d}_{\text{obs}}^n$ by $\mathbf{d}_{\text{obs},j}^n \sim \mathcal{N}(\mathbf{d}_{\text{obs}}^n, \mathbf{C}_{\mathbf{e}^n})$ and minimize the resulting objective function

$$\begin{aligned}
\mathcal{O}_r(\hat{\mathbf{y}}^n) &= \frac{1}{2} (\hat{\mathbf{y}}^n - \hat{\mathbf{y}}_j^{n,f})^\top \left(\mathbf{C}_{\hat{\mathbf{y}}^n}^f \right)^{-1} (\hat{\mathbf{y}}^n - \hat{\mathbf{y}}_j^{n,f}) + \\
&\quad \frac{1}{2} (\mathbf{d}_{\text{obs},j}^n - \mathbf{H}_n \hat{\mathbf{y}}^n)^\top \mathbf{C}_{\mathbf{e}^n}^{-1} (\mathbf{d}_{\text{obs},j}^n - \mathbf{H}_n \hat{\mathbf{y}}^n). \tag{4.85}
\end{aligned}$$

Minimizing $\mathcal{O}_r(\hat{\mathbf{y}}^n)$ for $\hat{\mathbf{y}}^n$ and calling the resulting state vector as $\hat{\mathbf{y}}_j^{n,a}$ leads to

$$\begin{aligned}
\hat{\mathbf{y}}_j^{n,a} &= \hat{\mathbf{y}}_j^{n,f} + \left[\left(\mathbf{C}_{\hat{\mathbf{y}}^n}^f \right)^{-1} + \mathbf{H}_n^\top \mathbf{C}_{\mathbf{e}^n}^{-1} \mathbf{H}_n \right]^{-1} \mathbf{H}_n^\top \mathbf{C}_{\mathbf{e}^n}^{-1} (\mathbf{d}_{\text{obs},j}^n - \mathbf{H}_n \hat{\mathbf{y}}_j^{n,f}) \\
&= \hat{\mathbf{y}}_j^{n,f} + \mathbf{C}_{\hat{\mathbf{y}}^n}^f \mathbf{H}_n^\top \left(\mathbf{H}_n \mathbf{C}_{\hat{\mathbf{y}}^n}^f \mathbf{H}_n^\top + \mathbf{C}_{\mathbf{e}^n} \right)^{-1} (\mathbf{d}_{\text{obs},j}^n - \mathbf{H}_n \hat{\mathbf{y}}_j^{n,f}) \\
&= \hat{\mathbf{y}}_j^{n,f} + \mathbf{K}_n (\mathbf{d}_{\text{obs},j}^n - \mathbf{H}_n \hat{\mathbf{y}}_j^{n,f}). \tag{4.86}
\end{aligned}$$

Repeating this process for $j = 1, 2, \dots, N_e$ gives an ensemble of N_e conditional realizations. Eq. (4.86) forms the basis of the EnKF method.

To derive the final EnKF analysis equation, we need to develop expressions for the covariance matrix $\mathbf{C}_{\hat{\mathbf{y}}^n}^f$. First, we note that $\mathbf{C}_{\hat{\mathbf{y}}^n}^f$ can be partitioned as

$$\mathbf{C}_{\hat{\mathbf{y}}^n}^f = \begin{bmatrix} \mathbf{C}_{\text{mm}^n}^f & \mathbf{C}_{\text{mx}^n}^f & \mathbf{C}_{\text{md}^n}^f \\ \mathbf{C}_{\text{xm}^n}^f & \mathbf{C}_{\text{xx}^n}^f & \mathbf{C}_{\text{xd}^n}^f \\ \mathbf{C}_{\text{dm}^n}^f & \mathbf{C}_{\text{dx}^n}^f & \mathbf{C}_{\text{dd}^n}^f \end{bmatrix}. \tag{4.87}$$

However, we only need the products

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top &= \begin{bmatrix} \mathbf{C}_{\mathbf{m}\mathbf{m}^n}^f & \mathbf{C}_{\mathbf{m}\mathbf{x}^n}^f & \mathbf{C}_{\mathbf{m}\mathbf{d}^n}^f \\ \mathbf{C}_{\mathbf{x}\mathbf{m}^n}^f & \mathbf{C}_{\mathbf{x}\mathbf{x}^n}^f & \mathbf{C}_{\mathbf{x}\mathbf{d}^n}^f \\ \mathbf{C}_{\mathbf{d}\mathbf{m}^n}^f & \mathbf{C}_{\mathbf{d}\mathbf{x}^n}^f & \mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f \end{bmatrix} \begin{bmatrix} \mathbf{O}_{N_m} \\ \mathbf{O}_{N_x} \\ \mathbf{I}_{N_n} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{C}_{\mathbf{m}\mathbf{d}^n}^f \\ \mathbf{C}_{\mathbf{x}\mathbf{d}^n}^f \\ \mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f \end{bmatrix} \\
&= \mathbf{C}_{\mathbf{y}\mathbf{d}^n}^f
\end{aligned} \tag{4.88}$$

and

$$\mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top = \mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f. \tag{4.89}$$

Using these last two results in (4.86) gives

$$\begin{aligned}
\hat{\mathbf{y}}_j^{n,a} &= \hat{\mathbf{y}}_j^{n,f} + \begin{bmatrix} \mathbf{C}_{\mathbf{m}\mathbf{d}^n}^f \\ \mathbf{C}_{\mathbf{x}\mathbf{d}^n}^f \\ \mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f \end{bmatrix} \left(\mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f + \mathbf{C}_{\mathbf{e}^n} \right)^{-1} \left(\mathbf{d}_{\text{obs},j}^n - \mathbf{H}_n \hat{\mathbf{y}}_j^{n,f} \right) \\
&= \hat{\mathbf{y}}_j^{n,f} + \mathbf{C}_{\mathbf{y}\mathbf{d}^n}^f \left(\mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f + \mathbf{C}_{\mathbf{e}^n} \right)^{-1} \left(\mathbf{d}_{\text{obs},j}^n - \mathbf{H}_n \hat{\mathbf{y}}_j^{n,f} \right) \\
&= \hat{\mathbf{y}}_j^{n,f} + \mathbf{K}_n \left(\mathbf{d}_{\text{obs},j}^n - \mathbf{h} \left(\mathbf{y}_j^{n,f} \right) \right).
\end{aligned} \tag{4.90}$$

Under the assumptions we have made, application of Eq. (4.90) for $j = 1, 2, \dots, N_e$ provides a sampling for the posterior PDF $p(\hat{\mathbf{y}}^n | \mathbf{d}_{\text{obs}}^n)$. Note that the trick of augmenting the state vector was done only to derive the EnKF equation with a form similar to the KF. In terms of computational implementation, however, we do not need to augment the state vector and Eq. (4.90) can be written directly in terms of $\mathbf{y}^n$.

4.3.1 Perturbed Observations

Eq. (4.90) was derived from RML replacing $\mathbf{d}_{\text{obs}}^n$ by a sample $\mathbf{d}_{\text{obs},j}^n \sim \mathcal{N}(\mathbf{d}_{\text{obs}}^n, \mathbf{C}_{\mathbf{e}^n})$, which is equivalent to adding a random vector $\mathbf{e}_j^n \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{e}^n})$ to the observations

$$\begin{aligned}
\mathbf{d}_{\text{obs},j}^n &= \mathbf{d}_{\text{obs}}^n + \mathbf{e}_j^n \\
&= \mathbf{d}_{\text{obs}}^n + \mathbf{C}_{\mathbf{e}^n}^{1/2} \mathbf{z}^n,
\end{aligned} \tag{4.91}$$

where $\mathbf{z}^n \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. This procedure is known as perturbed observation scheme in the ensemble data assimilation literature and it was independently introduced by Burgers et al. [51] and Houtekamer and Mitchell [197] in the EnKF formulation.

The perturbed observations are necessary to ensure that EnKF results in the correct posterior covariance in the linear case with infinite ensemble size. To show the role of $\mathbf{e}_j^n$, consider the EnKF equation written as

$$\mathbf{y}_j^{n,a} = \mathbf{y}_j^{n,f} + \mathbf{K}_n \left(\mathbf{d}_{\text{obs}}^n + \mathbf{e}_j^n - \mathbf{H}_n \mathbf{y}_j^{n,f} \right). \quad (4.92)$$

The corresponding equation for updating the ensemble mean is

$$\bar{\mathbf{y}}^{n,a} = \bar{\mathbf{y}}^{n,f} + \mathbf{K}_n \left(\mathbf{d}_{\text{obs}}^n - \mathbf{H}_n \bar{\mathbf{y}}^{n,f} \right), \quad (4.93)$$

where it has been used the fact that $\bar{\mathbf{e}}^n = \mathbf{0}$. Hence

$$\begin{aligned} \mathbf{y}_j^{n,a} - \bar{\mathbf{y}}^{n,a} &= \mathbf{y}_j^{n,f} - \bar{\mathbf{y}}^{n,f} + \mathbf{K}_n \left(\mathbf{e}_j^n - \mathbf{H}_n \left(\mathbf{y}_j^{n,f} - \bar{\mathbf{y}}^{n,f} \right) \right) \\ &= \delta \mathbf{y}_j^{n,f} + \mathbf{K}_n \left(\mathbf{e}_j^n - \mathbf{H}_n \delta \mathbf{y}_j^{n,f} \right). \end{aligned} \quad (4.94)$$

The expression for the posterior covariance of $\mathbf{y}^n$ is

$$\begin{aligned} \mathbf{C}_{\mathbf{y}^n}^a &= \mathbb{E} \left[\left(\mathbf{y}_j^{n,a} - \bar{\mathbf{y}}^{n,a} \right) \left(\mathbf{y}_j^{n,a} - \bar{\mathbf{y}}^{n,a} \right)^\top \right] \\ &= \mathbb{E} \left[\left(\delta \mathbf{y}_j^{n,f} + \mathbf{K}_n (\mathbf{e}_j^n - \mathbf{H}_n \delta \mathbf{y}_j^{n,f}) \right) \left(\delta \mathbf{y}_j^{n,f} + \mathbf{K}_n (\mathbf{e}_j^n - \mathbf{H}_n \delta \mathbf{y}_j^{n,f}) \right)^\top \right] \\ &= \underbrace{\mathbb{E} \left[\delta \mathbf{y}_j^{n,f} (\delta \mathbf{y}_j^{n,f})^\top \right]}_{\mathbf{C}_{\mathbf{y}^n}^f} + \underbrace{\mathbb{E} \left[\delta \mathbf{y}_j^{n,f} (\mathbf{K}_n \mathbf{e}_j^n)^\top \right]}_{=0} - \\ &\quad \underbrace{\mathbb{E} \left[\delta \mathbf{y}_j^{n,f} (\delta \mathbf{y}_j^{n,f})^\top \right] \mathbf{H}_n^\top \mathbf{K}_n^\top}_{\mathbf{C}_{\mathbf{y}^n}^f} + \underbrace{\mathbb{E} \left[\mathbf{K}_n \mathbf{e}_j^n (\delta \mathbf{y}_j^{n,f})^\top \right]}_{=0} - \\ &\quad \underbrace{\mathbf{K}_n \mathbf{H}_n \mathbb{E} \left[\delta \mathbf{y}_j^{n,f} (\delta \mathbf{y}_j^{n,f})^\top \right]}_{\mathbf{C}_{\mathbf{y}^n}^f} + \underbrace{\mathbf{K}_n \mathbb{E} \left[\mathbf{e}_j^n (\mathbf{e}_j^n)^\top \right] \mathbf{K}_n^\top}_{\mathbf{C}_{\mathbf{e}^n}} - \\ &\quad \underbrace{\mathbf{K}_n \mathbb{E} \left[\mathbf{e}_j^n (\mathbf{H}_n \delta \mathbf{y}_j^{n,f})^\top \right] \mathbf{K}_n^\top}_{=0} - \underbrace{\mathbf{K}_n \mathbb{E} \left[\mathbf{H}_n \delta \mathbf{y}_j^{n,f} (\mathbf{e}_j^n)^\top \right] \mathbf{K}_n^\top}_{=0} + \\ &\quad \underbrace{\mathbf{K}_n \mathbf{H}_n \mathbb{E} \left[\delta \mathbf{y}_j^{n,f} (\delta \mathbf{y}_j^{n,f})^\top \right] \mathbf{H}_n^\top \mathbf{K}_n^\top}_{\mathbf{C}_{\mathbf{y}^n}^f} \\ &= \mathbf{C}_{\mathbf{y}^n}^f - \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top \mathbf{K}_n^\top - \mathbf{K}_n \mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f + \mathbf{K}_n \mathbf{C}_{\mathbf{e}^n} \mathbf{K}_n^\top + \mathbf{K}_n \mathbf{H}_n \mathbf{C}_{\mathbf{y}^n}^f \mathbf{H}_n^\top \mathbf{K}_n^\top. \end{aligned} \quad (4.95)$$

Therefore, if we neglect the perturbation vector $\mathbf{e}_j^n$ in the EnKF analysis, the term

$$\mathbf{K}_n \mathbb{E} \left[\mathbf{e}_j^n (\mathbf{e}_j^n)^\top \right] \mathbf{K}_n^\top = \mathbf{K}_n \mathbf{C}_{\mathbf{e}^n} \mathbf{K}_n^\top \quad (4.96)$$

present in the expression for $\mathbf{C}_{\mathbf{y}^n}^a$ disappears. Note that $\mathbf{K}_n \mathbf{C}_{\mathbf{e}^n} \mathbf{K}_n^\top$ is at least positive semidefinite. Therefore, the resulting expression for $\mathbf{C}_{\mathbf{y}^n}^a$ has a positive semidefinite term missing. As a result, if we neglect $\mathbf{e}_j^n$ the entries in the matrix $\mathbf{C}_{\mathbf{y}^n}^a$ are underestimated.

The traditional interpretation of perturbed observations treats the observations as random variables. The same concept is employed in RML. However, some authors argue that the correct interpretation involves perturbing the predicted data vector. The rationale behind this perspective is based on the assumption that the true system is a sample from the same distribution as the ensemble members. Since the measurements of the true system are already perturbed by observation errors, it follows that the predictions of each ensemble member should be perturbed in a similar manner [447]. Nevertheless, here, the equations are presented in terms of perturbed observations because this approach is more conventional in the data assimilation literature. It is worth noting that in practice, the actual calculations remain unchanged, whether one considers $\mathbf{e}_j^n$ as a perturbation added to the observations or to the predicted data. However, these two interpretations may yield different results when attempting to extend the methods to account for non-additive errors.

4.3.2 Ensemble-Based Covariance

The last step to obtain the final EnKF method is to introduce the ensemble-based approximations of the covariances $\mathbf{C}_{\mathbf{md}^n}^f$, $\mathbf{C}_{\mathbf{xd}^n}^f$ and $\mathbf{C}_{\mathbf{dd}^n}^f$ as

$$\begin{aligned} \mathbf{C}_{\mathbf{md}^n}^f &\approx \tilde{\mathbf{C}}_{\mathbf{md}^n}^f \\ &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(\mathbf{m}_j^{n,f} - \overline{\mathbf{m}}^{n,f} \right) \left(\mathbf{h} \left(\mathbf{y}_j^{n,f} \right) - \overline{\mathbf{h} \left(\mathbf{y}^{n,f} \right)} \right)^\top \\ &= \Delta \mathbf{M}^{n,f} \left(\Delta \mathbf{D}^{n,f} \right)^\top, \end{aligned} \quad (4.97)$$

$$\begin{aligned} \mathbf{C}_{\mathbf{xd}^n}^f &\approx \tilde{\mathbf{C}}_{\mathbf{xd}^n}^f \\ &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(\mathbf{x}_j^{n,f} - \overline{\mathbf{x}}^{n,f} \right) \left(\mathbf{h} \left(\mathbf{y}_j^{n,f} \right) - \overline{\mathbf{h} \left(\mathbf{y}^{n,f} \right)} \right)^\top \\ &= \Delta \mathbf{X}^{n,f} \left(\Delta \mathbf{D}^{n,f} \right)^\top, \end{aligned} \quad (4.98)$$

and

$$\begin{aligned}
 \mathbf{C}_{\mathbf{d}\mathbf{d}^n}^f &\approx \tilde{\mathbf{C}}_{\mathbf{d}\mathbf{d}^n}^f \\
 &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(\mathbf{h}(\mathbf{y}_j^{n,f}) - \overline{\mathbf{h}(\mathbf{y}^{n,f})} \right) \left(\mathbf{h}(\mathbf{y}_j^{n,f}) - \overline{\mathbf{h}(\mathbf{y}^{n,f})} \right)^\top \\
 &= \Delta \mathbf{D}^{n,f} (\Delta \mathbf{D}^{n,f})^\top.
 \end{aligned} \tag{4.99}$$

The tildes over the covariance matrices were introduced to denote that these matrices are estimated from ensembles. $\Delta \mathbf{M}^{n,f}$, $\Delta \mathbf{X}^{n,f}$ and $\Delta \mathbf{D}^{n,f}$ in Eqs. 4.97, 4.98 and 4.99 are defined as

$$\begin{aligned}
 \Delta \mathbf{M}^{n,f} &= \frac{1}{\sqrt{N_e - 1}} \left(\mathbf{M}^{n,f} - \overline{\mathbf{M}}^{n,f} \right) \\
 &= \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{m}_1^{n,f} - \overline{\mathbf{m}}^{n,f}, \dots, \mathbf{m}_{N_e}^{n,f} - \overline{\mathbf{m}}^{n,f} \right],
 \end{aligned} \tag{4.100}$$

$$\begin{aligned}
 \Delta \mathbf{X}^{n,f} &= \frac{1}{\sqrt{N_e - 1}} \left(\mathbf{X}^{n,f} - \overline{\mathbf{X}}^{n,f} \right) \\
 &= \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{x}_1^{n,f} - \overline{\mathbf{x}}^{n,f}, \dots, \mathbf{x}_{N_e}^{n,f} - \overline{\mathbf{x}}^{n,f} \right],
 \end{aligned} \tag{4.101}$$

and

$$\begin{aligned}
 \Delta \mathbf{D}^{n,f} &= \frac{1}{\sqrt{N_e - 1}} \left(\mathbf{D}^{n,f} - \overline{\mathbf{D}}^{n,f} \right) \\
 &= \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{h}(\mathbf{y}_1^{n,f}) - \overline{\mathbf{h}(\mathbf{y}^{n,f})}, \dots, \mathbf{h}(\mathbf{y}_{N_e}^{n,f}) - \overline{\mathbf{h}(\mathbf{y}^{n,f})} \right],
 \end{aligned} \tag{4.102}$$

where

$$\overline{\mathbf{m}}^{n,f} = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{m}_j^{n,f}, \tag{4.103}$$

$$\overline{\mathbf{x}}^{n,f} = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{x}_j^{n,f}, \tag{4.104}$$

and

$$\overline{\mathbf{h}(\mathbf{y}^{n,f})} = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{h}(\mathbf{y}_j^{n,f}). \tag{4.105}$$

$\overline{\mathbf{M}}^{n,f}$, $\overline{\mathbf{X}}^{n,f}$, and $\overline{\mathbf{D}}^{n,f}$ are matrices with all columns equal to the corresponding ensemble means. A more compact notation can be written defining a centering matrix as

$$\mathbf{A} = \frac{1}{\sqrt{N_e - 1}} \left(\mathbf{I} - \frac{1}{N_e} \mathbf{1}\mathbf{1}^\top \right), \tag{4.106}$$

where $\mathbf{I}$ is the $N_e \times N_e$ identity matrix and $\mathbf{1} = [1 \ 1 \ \dots \ 1]^\top$ is the N_e -dimensional vector with all entries equal to one. Using the centering matrix, we can write

$$\Delta \mathbf{M}^{n,f} = \mathbf{M}^{n,f} \mathbf{A}, \quad (4.107)$$

$$\Delta \mathbf{X}^{n,f} = \mathbf{X}^{n,f} \mathbf{A}, \quad (4.108)$$

and

$$\Delta \mathbf{D}^{n,f} = \mathbf{D}^{n,f} \mathbf{A}. \quad (4.109)$$

The final EnKF analysis equation can be written as

$$\mathbf{y}_j^{n,a} = \mathbf{y}_j^{n,f} + \tilde{\mathbf{C}}_{\mathbf{y}\mathbf{d}^n}^f \left(\tilde{\mathbf{C}}_{\mathbf{d}\mathbf{d}^n}^f + \mathbf{C}_{\mathbf{e}^n} \right)^{-1} \left(\mathbf{d}_{\text{obs},j}^n - \mathbf{h} \left(\mathbf{y}_j^{n,f} \right) \right) \quad (4.110)$$

for $j = 1, 2, \dots, N_e$, where

$$\begin{aligned} \tilde{\mathbf{C}}_{\mathbf{y}\mathbf{d}^n}^f &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{y}_j^{n,f} - \bar{\mathbf{y}}^{n,f}) \left(\mathbf{h} \left(\mathbf{y}_j^{n,f} \right) - \overline{\mathbf{h} \left(\mathbf{y}^{n,f} \right)} \right)^\top \\ &= \Delta \mathbf{Y}^{n,f} \left(\Delta \mathbf{D}^{n,f} \right)^\top \end{aligned} \quad (4.111)$$

and

$$\begin{aligned} \Delta \mathbf{Y}^{n,f} &= \frac{1}{\sqrt{N_e - 1}} \left(\mathbf{Y}^{n,f} - \bar{\mathbf{Y}}^{n,f} \right) \\ &= \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{y}_1^{n,f} - \bar{\mathbf{y}}^{n,f}, \dots, \mathbf{y}_{N_e}^{n,f} - \bar{\mathbf{y}}^{n,f} \right] \\ &= \mathbf{Y}^{n,f} \mathbf{A}. \end{aligned} \quad (4.112)$$

Eq. (4.110) represents the final EnKF analysis equation. It shares a similar form with the original KF equation. However, unlike the KF, we do not directly update the covariance matrix of the state vector. Instead, we update the ensemble of state vectors, which serves as samples from the prior (forecast) distribution, to obtain samples of the posterior (analyzed) distribution.

4.3.2.1 Example: Porosity Distribution in a Core Sample

Consider the same problem introduced in Section 2.3.1.4. We have already computed the MAP estimate (Fig. 2.2), RML samples (Fig. 2.8) and the KF solution (Fig. 4.3) for this problem. Fig. 4.5 presents the results for this problem assimilating the 10 data points sequentially using EnKF with an ensemble of 100 members. The final ensemble from EnKF is very similar to the posterior samples obtained with RML. In fact, for linear-Gaussian

problems, these methods are equivalent up to sampling errors caused by spurious correlations. We are going to discuss this aspect in Chapter 7.

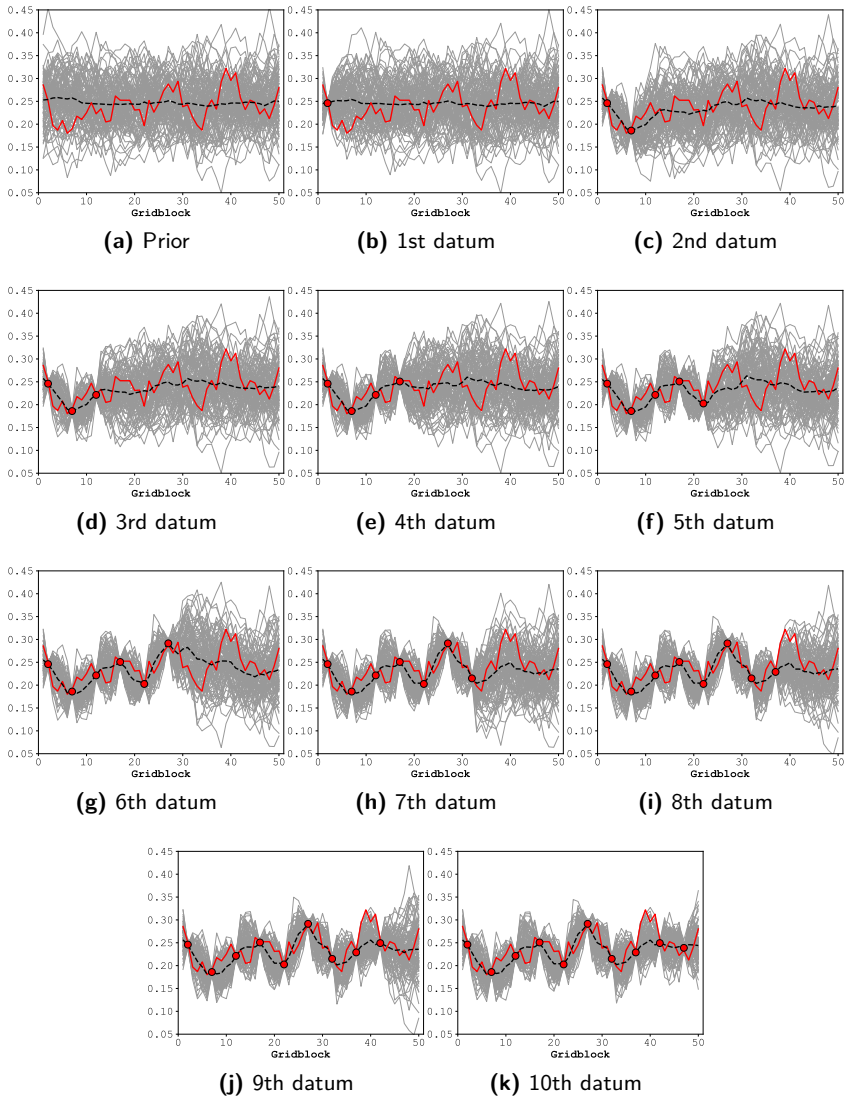


Fig. 4.5: EnKF results with $N_e = 100$ for porosity in a core sample. The solid red line represents the ground truth. The red circles represent the observed data points. The gray lines represent samples obtained by EnKF for different numbers of observations. The dashed lines represent the ensemble mean. Example 4.3.2.1.

4.3.3 Parameters-State Consistency

In practice, the rationale for updating primary variables, which represent the state of the dynamical system, is to circumvent the need to initiate reservoir simulations from time zero (initial reservoir conditions) after each data assimilation time step. This approach operates under the assumption that the updated primary variables are statistically consistent with those that would be obtained by running the reservoir simulator using the updated set of model parameters from time zero. However, this consistency can only be rigorously established for linear-Gaussian problems [435].

However, the equations governing reservoir simulation are typically characterized by moderate to high nonlinearity. Consequently, the assumption of consistency may be significantly violated, potentially undermining the performance of data assimilation and giving rise to convergence challenges in the simulations. In practical terms, following data assimilation at time t_n , reservoir simulations may be restarted with incorrect pressure and saturation values, thus compromising the historical material balance of the field. In extreme cases, the analysis step within the EnKF may yield non-physical values for these variables, such as negative pressures or saturations exceeding one.

Fig. 4.6 illustrates this problem by showing the water saturation profiles at 360 days for the linear two-phase flow discussed in Section 2.4.5.3. This figure includes the updated saturation profiles obtained after applying the EnKF analysis. Although the mean updated water saturation is very close to the true saturation, the updated individual realizations exhibit non-physical values.

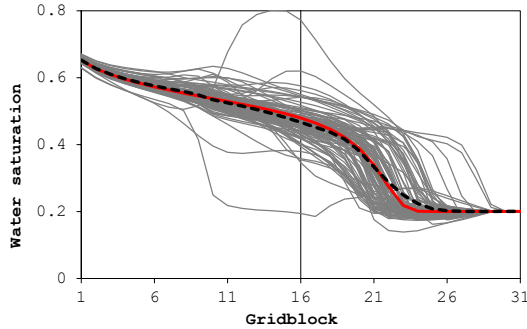


Fig. 4.6: Water saturation profile at 360 days. The gray lines represent the ensemble of saturation profiles updated with EnKF analysis equation. The dashed line represents the ensemble mean and red line represents the ground truth. Reproduced from Emerick and Reynolds [131] with permission from Springer Nature.

To address the parameter-state consistency problem, one straightforward approach is to perform simulations from time zero at each data assimilation

time step. However, this solution can be exceedingly time-consuming. Alternatively, a more efficient strategy involves initiating simulations from time zero only when a large update in the state vector is detected. This method is commonly referred to as the half-iteration EnKF (HI-EnKF) Wang et al. [459].

4.3.4 Coupling EnKF with a Reservoir Simulator

Before we present the EnKF pseudo-code, it might be helpful to examine the components of the state vector from the perspective of reservoir simulation. Reservoir data assimilation is typically framed as a parameter-estimation problem, where the objective is to estimate model parameters based on a set of observations. The vector of model parameters generally comprises “static” rock properties, such as porosity, permeabilities, rock compressibility, and more. However, here the EnKF was introduced as a parameter-state estimation problem. In this scenario, the state encompasses dynamic variables like pressure and phase saturations. In a reservoir model, the state corresponds to the primary variables that are solved for in the fluid transport problem. Hence, the specific variables included in the state vector depend on the formulation of the particular problem at hand. For example, in a standard black-oil formulation for a petroleum reservoir, the state comprises pressure, water and gas saturations, and the gas-oil solubility ratio. In a compositional fluid model, the state may involve the molar fraction of each pseudo-component.

Recall that the parameter-state vector takes the following form:

$$\mathbf{y}^n = \begin{bmatrix} \mathbf{m} \\ \mathbf{x}^n \end{bmatrix}, \quad (4.113)$$

where $\mathbf{m}$ represents the vector of model parameters, and $\mathbf{x}^n$ represents the vector of state variables at time t_n . In the case of a standard black-oil simulator, the parameter-state vector is composed of

$$\mathbf{y}^n = \begin{bmatrix} \overbrace{\phi^\top, (\ln \kappa)^\top, \dots}^{\text{model parameters } (\mathbf{m})}, \underbrace{(\mathbf{p}_r^n)^\top, (\mathbf{s}_w^n)^\top, (\mathbf{s}_g^n)^\top, (\mathbf{r}_s^n)^\top}_{\text{primary variables } (\mathbf{x}^n)} \end{bmatrix}^\top. \quad (4.114)$$

Here, ϕ and $\ln \kappa$ are vectors containing the porosity and log-permeability values for each gridblock in the model, while $\mathbf{p}_r^n$, $\mathbf{s}_w^n$, $\mathbf{s}_g^n$, and $\mathbf{r}_s^n$ are vectors containing the pressure, water saturation, gas saturation, and gas-oil solubility ratio, respectively.

Pseudo-code 4.3 summarizes the EnKF process for reservoir data assimilation. Note that to apply the EnKF analysis at time step, t_n , we need the

forecast state vector, $\mathbf{y}_j^{n,f}$, and the corresponding predicted data vector for each ensemble member. The j th forecast vector of model parameters corresponds to the vector of parameters conditioned to all data up to time t_{n-1} . The forecast vector of states $\mathbf{x}_j^{n,f}$ is obtained by running the forward model with $\mathbf{m}_j^f$ and $\mathbf{x}_j^{n-1,a}$ from t_{n-1} to time t_n , i.e., $\mathbf{x}_j^{n,f} = \mathbf{f}(\mathbf{m}_j^f, \mathbf{x}_j^{n-1,a})$. This means that the execution of the reservoir simulator is restarted at time t_{n-1} with the updated states and forward in time until t_n .

Pseudo-code 4.3: Ensemble Kalman Filter

1. **Initialization:** generate the initial ensemble, $\{\mathbf{y}_j^{0,a}\}_{j=1}^{N_e}$, by sampling the prior distributions of parameters, $\mathbf{m}$, and states, $\mathbf{x}$.
2. For $n = 1$ to N_t :

- **Forecast step:** generate N_e unconditional (forecast) samples of $\mathbf{y}^n$ and the corresponding predicted data, $\mathbf{d}^n$, by running the forward model from time t_{n-1} to t_n for $j = 1, 2, \dots, N_e$:

$$\mathbf{y}_j^{n,f} = \begin{bmatrix} \mathbf{m}_j^f \\ \mathbf{x}_j^{n,f} \end{bmatrix} = \begin{bmatrix} \mathbf{m}_j^f \\ \mathbf{f}(\mathbf{m}_j^f, \mathbf{x}_j^{n-1,a}) \end{bmatrix} \quad (4.115)$$

and

$$\mathbf{d}_j^{n,f} = \mathbf{h}(\mathbf{m}_j^f, \mathbf{x}_j^{n-1,a}), \quad (4.116)$$

where $\mathbf{f}(\cdot)$ and $\mathbf{h}(\cdot)$ denote the nonlinear forward and observation models, respectively.

- Perturb the vector of observations:

$$\mathbf{d}_{\text{obs},j}^n = \mathbf{d}_{\text{obs}}^n + \mathbf{C}_{\mathbf{e}^n}^{1/2} \mathbf{z}_j^n, \quad \text{for } j = 1, 2, \dots, N_e, \quad (4.117)$$

where $\mathbf{z}_j^n \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

- **Analysis step:** update the combined parameter-state vector using the EnKF analysis equation:

$$\mathbf{y}_j^{n,a} = \mathbf{y}_j^{n,f} + \tilde{\mathbf{C}}_{\mathbf{y}\mathbf{d}^n}^f \left(\tilde{\mathbf{C}}_{\mathbf{d}\mathbf{d}^n}^f + \mathbf{C}_{\mathbf{e}^n} \right)^{-1} \left(\mathbf{d}_{\text{obs},j}^n - \mathbf{d}_j^{n,f} \right), \quad (4.118)$$

for $j = 1, 2, \dots, N_e$.

4.3.4.1 Example: Two-Phase Flow in 2D Model

Consider the same synthetic data assimilation problem presented in Section 2.4.5.4. Previously, we presented the MAP estimate and posterior realizations obtained using an adjoint-based implementation of RML. Now, we present data assimilation results using the EnKF with an ensemble of 100 realizations. During data assimilation, we applied a distance-based localization scheme to regularize the EnKF [197]⁷.

Fig. 4.7 illustrates the gradual changes in the log-permeability for the first realization of the ensemble imposed by sequential data assimilation. Fig. 4.8 shows the first three prior and posterior realizations of the ensemble, along with the true log-permeability. The EnKF data assimilation recovered some of the main permeability features observed in the ground truth, such as the high-permeability corridor in the upper part of the model and the low-permeability region in the lower part. Fig. 4.9 presents the predicted water production rate for three wells in the model. The posterior predictions were obtained by running reservoir simulations from time zero with the final ensemble of log-permeability. The results in this figure demonstrate significant improvement in data matching, and the forecast encompasses the predicted data using the true log-permeability.

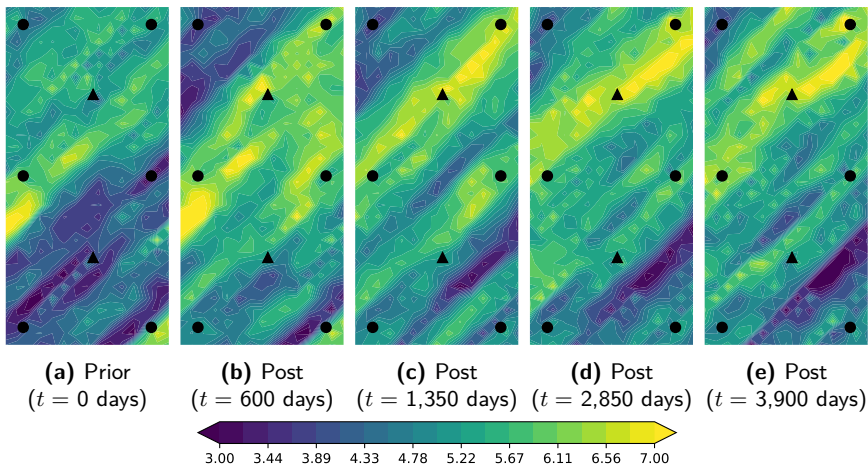


Fig. 4.7: Evolution of the first realization of log-permeability during the sequential data assimilation with EnKF for a 2D model under two-phase flow. Black circles represent the positions of producers, while black triangles denote the positions of water injection wells. Example 4.3.4.1.

⁷ Localization is a standard strategy to address issues arising from using a limited ensemble size to estimate covariances in the EnKF. This topic is discussed in detail in Chapter 7.

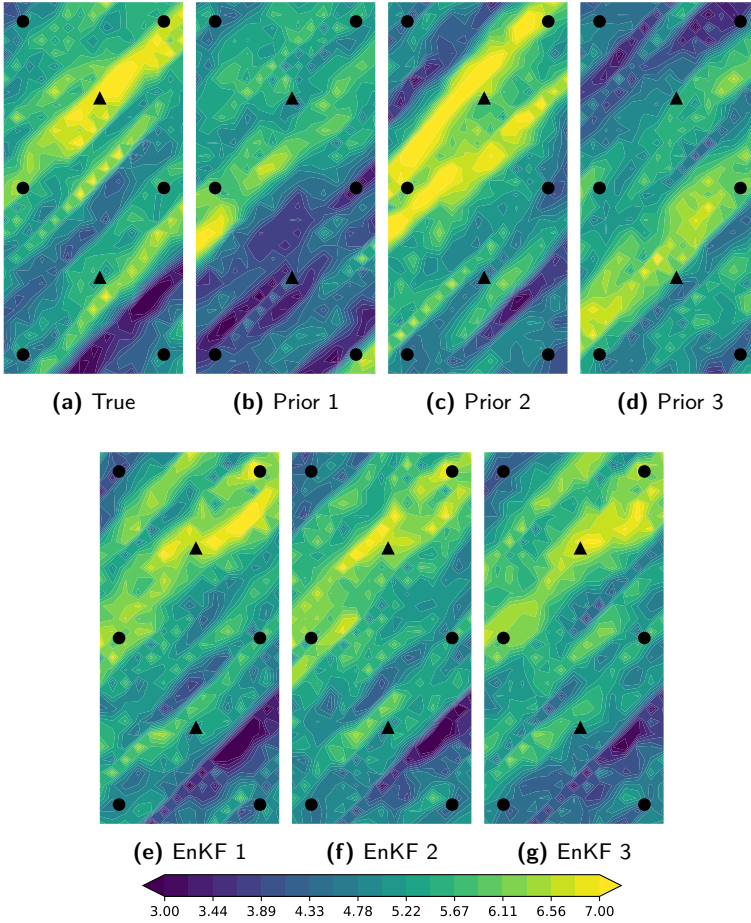


Fig. 4.8: True log-permeability, three prior realizations, and three posterior realizations obtained using EnKF for a 2D model under two-phase flow. Black circles indicate the positions of producers, and black triangles indicate the positions of water injection wells. Example 4.3.4.1.

4.3.5 Filter Divergence

Filter divergence occurs when the EnKF estimates deviate significantly from the true state, as illustrated in Fig. 4.10. Once filter divergence takes hold, it becomes an irreversible situation, and the EnKF is no longer capable of correcting the estimates to converge with the true state. This divergence phenomenon typically arises due to a gradual underestimation of the state covariance, leading to a scenario where $\|\tilde{\mathbf{C}}_{\mathbf{d}\mathbf{d}^n}^f\| \ll \|\mathbf{C}_{\mathbf{e}^n}\|$ [60], causing the filter to begin missing observations. Several factors contribute to filter di-

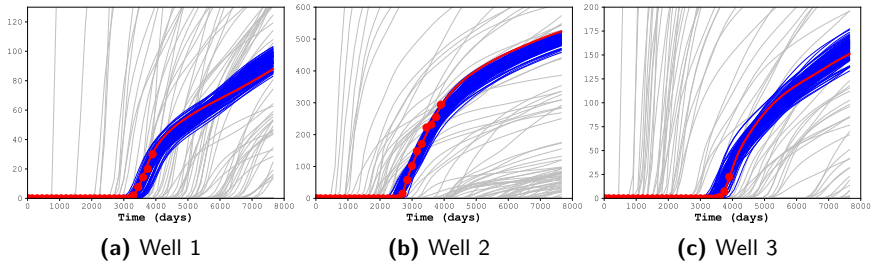


Fig. 4.9: Water production rate in bbl/day for three wells. Red circles are the observations, red line the prediction with the true log-permeability distribution. The gray and blue lines show the predictions from 100 prior and posterior realizations, respectively. The posterior predictions were obtained running the final EnKF realizations from time zero. Example 4.3.4.1.

vergence, including inadequate parametrization, model errors, incorrect (underestimated) specification of the data error covariance $\mathbf{C}_e$, sampling errors, and limitations in degrees of freedom due to the small size of the ensemble.

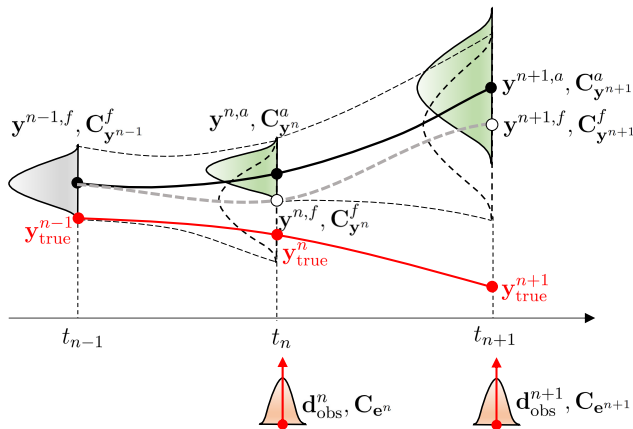


Fig. 4.10: Illustration of filter divergence.

4.3.6 Ensemble Square Root Filters

The KF update equation for the state covariance matrix can be written as

$$\mathbf{C}_{\mathbf{y}^n}^a = (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n) \mathbf{C}_{\mathbf{y}^n}^f. \quad (4.119)$$

According to Whitaker and Hamill [463], because the EnKF employs an updating scheme based on perturbed observations, this equation is satisfied in a statistical sense only. In other words, EnKF satisfies Eq. (4.119) only in the limit of an infinite ensemble size. This limitation can lead to suboptimal filter performance when dealing with small ensembles. To address this issue, several implementations of the analysis equation have been developed, which are based on deterministic updates of the ensemble to ensure the exact satisfaction of Eq. (4.119). These implementations are commonly referred to as ensemble square root filters (EnSRF).

A typical EnSRF follows two main steps depicted in Pseudo-code 4.4.

Pseudo-code 4.4: EnSRF

1. Update the mean:

$$\bar{\mathbf{y}}^{n,a} = \bar{\mathbf{y}}^{n,f} + \tilde{\mathbf{K}}_n \left(\mathbf{d}_{\text{obs}}^n - \bar{\mathbf{h}}(\mathbf{y}^{n,f}) \right). \quad (4.120)$$

2. Update the ensemble deviations:

$$\Delta \mathbf{Y}^{n,a} = \Delta \mathbf{Y}^{n,f} \mathbf{T}_R^n, \quad (4.121)$$

where $\mathbf{T}_R^n$ is called (right-)square root transform matrix.

The matrix $\mathbf{T}_R^n$ is selected such that

$$\begin{aligned} \tilde{\mathbf{C}}_{\mathbf{y}^n}^a &= \Delta \mathbf{Y}^{n,a} (\Delta \mathbf{Y}^{n,a})^\top \\ &= (\Delta \mathbf{Y}^{n,f} \mathbf{T}_R^n) (\Delta \mathbf{Y}^{n,f} \mathbf{T}_R^n)^\top \\ &= (\mathbf{I} - \tilde{\mathbf{K}}_n \mathbf{H}_n) \tilde{\mathbf{C}}_{\mathbf{y}^n}^f. \end{aligned} \quad (4.122)$$

Hence, square root filters are designed to ensure the exact satisfaction of the theoretical covariance update equation. However, this condition does not uniquely define $\mathbf{T}_R^n$. Consequently, several different variants have been proposed in the literature. The procedure described in Pseudo-code 4.4 is often referred to as the right-multiplied EnSRF, as $\mathbf{T}_R^n$ multiplies $\Delta \mathbf{Y}^{n,f}$ from the right. There are also a few left-multiplied schemes where the update equation takes the form

$$\Delta \mathbf{Y}^{n,a} = \mathbf{T}_L^n \Delta \mathbf{Y}^{n,f}. \quad (4.123)$$

Since EnSRF are defined imposing the condition (4.119), they tend to reduce the underestimation of the ensemble variance typically observed in EnKF. There is a vast body of literature on square root ensemble filters in the fields of oceanography and numerical weather prediction; e.g., see [439,

373] and the references therein. Nevertheless, these methods have seldom been employed in the context of reservoir data assimilation applications. One issue with square root filters is that, while they ensure the correct covariance for linear problems, they can sometimes produce highly skewed ensembles in nonlinear cases. Lawson and Hansen [251] demonstrated with one-dimensional nonlinear problems that the EnSRF ensemble may collapse to a single state with a few distant outliers providing the expected variance.

Ensemble Smoother with Multiple Data Assimilation

Abstract: *This chapter reviews the methods ensemble smoother (ES) and the ensemble smoother with multiple data assimilation (ES-MDA). Particular emphasis is placed on alternative derivations and implementations of ES-MDA, with examples applied to reservoir problems.*

5.1 Introduction

The sequential data assimilation scheme of the EnKF is highly appealing for atmospheric science; however, it presents certain challenges when applied to reservoir data assimilation. First, the frequent simulation restarts significantly increase the computational cost of the data assimilation and are particularly challenging when simulations are distributed across a cluster of computers. Additionally, there can be inconsistencies between updated model parameters and updated states, which can introduce convergence problems in the reservoir simulations. Finally, sequential data assimilation is impractical for integrated workflows involving updates in different components of a geomodeling process (e.g., seismic, structural, geological modeling, upscaling, and flow simulation).

One alternative to the EnKF is the ensemble smoother (ES) [448, 400]. In ES, all data are assimilated simultaneously in a single update, eliminating the need for simulation restarts. However, the single update scheme of ES has proven insufficient for properly conditioning reservoir models to dynamic data [78, 132, 133]. This limitation has driven the development of iterative forms of ES [78, 79, 132, 413, 279, 354].

This chapter presents the ensemble smoother with multiple data assimilation (ES-MDA), introduced by Emerick and Reynolds [132] as an iterative form of ES¹. Other iterative ES methods are discussed in Chapter 6.

5.2 Ensemble Smoother

ES was introduced by van Leeuwen and Evensen [448], with its first application in reservoir models presented by Skjervheim et al. [400]. The primary difference between the EnKF and ES is that ES does not assimilate data sequentially over time. Instead, ES computes a global update using all available data simultaneously. This approach eliminates the need to restart reservoir simulations at each data assimilation time step. Additionally, ES addresses the parameter-state consistency problem discussed in Section 4.3.3, as it only requires the estimation of model parameters in reservoir applications.

The ES analysis equation to update a vector of model parameters, $\mathbf{m}$, can be written as

$$\mathbf{m}_{c,j} = \mathbf{m}_j + \tilde{\mathbf{C}}_{\mathbf{md}} \left(\tilde{\mathbf{C}}_{\mathbf{dd}} + \mathbf{C}_e \right)^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}_j)), \quad (5.1)$$

for $j = 1, 2, \dots, N_e$. $\mathbf{m}_{c,j}$ is the j th posterior realization, $\tilde{\mathbf{C}}_{\mathbf{md}}$ is the matrix containing the covariance values between model parameters and predicted data, while $\tilde{\mathbf{C}}_{\mathbf{dd}}$ is the covariance matrix of predicted data. Both matrices are estimated based on the prior ensemble. $\mathbf{d}_{\text{obs},j}$ is the j th realization of the perturbed data vector and $\mathbf{g}(\mathbf{m}_j)$ is the predicted data with the j th prior realization, $\mathbf{m}_j$.

Eq. (5.1) is essentially the same analysis equation of EnKF. The primary difference lies in the update scheme (Fig. 5.1). While ES performs a single linear update, EnKF executes a sequence of smaller updates, enabling it to handle nonlinear dynamics more effectively. As a result, ES typically produces significantly inferior data matches compared to those obtained with EnKF when applied to reservoir data assimilation problems [78, 132, 133].

The superior performance of EnKF compared to ES in terms of data match quality can be explained by interpreting EnKF from a Gauss-Newton (GN) perspective [364]. Recall that the RML objective function can be written as:

¹ In a standard ES-MDA implementation, the update equation is applied a predefined number of times without convergence checking for termination. This has led some researchers to argue that ES-MDA should not be classified as an iterative method [288]. Nonetheless, we still refer to ES-MDA as an iterative form of ES for didactic purposes.

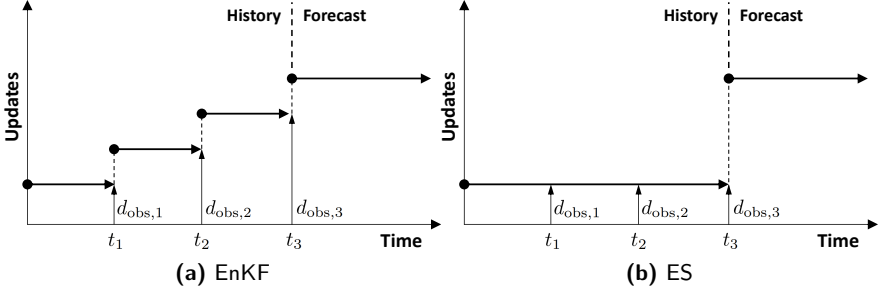


Fig. 5.1: Update schemes of EnKF and ES.

$$\begin{aligned} \mathcal{O}_r(\mathbf{m}) = & \frac{1}{2} (\mathbf{m} - \mathbf{m}_j)^\top \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_j) + \\ & \frac{1}{2} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m})). \end{aligned} \quad (5.2)$$

In GN, we look for a minimizer of (5.2) with an iterative process where the vector of model parameters at ℓ th iteration is updated using

$$\mathbf{m}_j^{\ell+1} = \mathbf{m}_j^\ell + \beta_\ell \delta \mathbf{m}_j^{\ell+1}, \quad (5.3)$$

where β_ℓ is a step size in the direction $\delta \mathbf{m}_j^{\ell+1}$ given by

$$\begin{aligned} \delta \mathbf{m}_j^{\ell+1} = & \mathbf{m}_j - \mathbf{m}_j^\ell + \mathbf{C}_m \mathbf{G}_\ell^\top (\mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top + \mathbf{C}_e)^{-1} \times \\ & (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}_j^\ell) + \mathbf{G}_\ell (\mathbf{m}_j^\ell - \mathbf{m}_j)). \end{aligned} \quad (5.4)$$

Consider the first iteration of GN with a full step, $\beta_0 = 1$, and an initial guess $\mathbf{m}_j^0 = \mathbf{m}_j$, where $\mathbf{m}_j \sim \mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_m)$. Replacing the products $\mathbf{C}_m \mathbf{G}_0^\top$ and $\mathbf{G}_0 \mathbf{C}_m \mathbf{G}_0^\top$ by the ensemble approximations

$$\mathbf{C}_m \mathbf{G}_0^\top \approx \tilde{\mathbf{C}}_{\text{md}} \quad (5.5)$$

and

$$\mathbf{G}_0 \mathbf{C}_m \mathbf{G}_0^\top \approx \tilde{\mathbf{C}}_{\text{dd}} \quad (5.6)$$

results in

$$\mathbf{m}_j^1 = \mathbf{m}_j + \tilde{\mathbf{C}}_{\text{md}} \left(\tilde{\mathbf{C}}_{\text{dd}} + \mathbf{C}_e \right)^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}_j)). \quad (5.7)$$

The last result effectively means that each EnKF analysis is equivalent to a single GN update with an average sensitivity matrix estimated from the ensemble [364]. Building on this finding, Emerick and Reynolds [129] conjectured that sequential assimilation is one of the reasons why EnKF yields acceptable results when assimilating closely spaced production data. In this

scenario, each data assimilation time step corresponds to one GN iteration. Still, due to consecutive production data, multiple GN updates occur. However, in the case of ES, all data points are assimilated in a single update, which may be insufficient to condition all realizations to the observations. Moreover, this process is prone to introduce overcorrection in the parameter values.

Pseudo-code 5.1 summarizes the ES method.

Pseudo-code 5.1: Ensemble Smoother

1. **Initialization:** generate the initial ensemble, $\{\mathbf{m}_j\}_{j=1}^{N_e}$, by sampling the prior distribution of the model parameters.
2. **Forecast step:** run the forward model from time zero until the end of the historical period to compute the vector of predicted data

$$\mathbf{d}_j = \mathbf{g}(\mathbf{m}_j), \quad \text{for } j = 1, 2, \dots, N_e. \quad (5.8)$$

3. Perturb the vector of observations:

$$\mathbf{d}_{\text{obs},j} = \mathbf{d}_{\text{obs}} + \mathbf{C}_e^{1/2} \mathbf{z}_j, \quad \text{for } j = 1, 2, \dots, N_e, \quad (5.9)$$

where $\mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

4. **Analysis step:** update the vectors of model parameters using the ES analysis equation:

$$\mathbf{m}_{c,j} = \mathbf{m}_j + \tilde{\mathbf{C}}_{\text{md}} \left(\tilde{\mathbf{C}}_{\text{dd}} + \mathbf{C}_e \right)^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{d}_j) \quad (5.10)$$

for $j = 1, 2, \dots, N_e$.

5.3 Ensemble Smoother with Multiple Data Assimilation

As mentioned in the introduction of this chapter, the global update scheme is an attractive feature of ES for reservoir data assimilation problems. Unfortunately, ES often fails to deliver acceptable data matches. This shortcoming has led to the proposal of various iterative forms of ES in the literature [78, 132, 79, 413, 279, 354, 140]. Among these methods, ES-MDA stands out due to the simplicity of its formulation and good performance.

The fundamental concept behind ES-MDA is to repeat the ES update equation multiple times, mimicking an iterative process; hence the name multiple data assimilation. However, in ES-MDA, during each repetition of the ES update, the data-error covariance matrix, $\mathbf{C}_e$, is “inflated” (multi-

plied by a coefficient $\alpha > 1$). The rationale behind inflating $\mathbf{C}_e$ is to prevent overweighting the data in the process. This can be illustrated by analyzing the expression derived in Chapter 2 (Section 2.3.1) for the posterior covariance in the linear-Gaussian case. Recall that the posterior covariance has the form

$$\begin{aligned}\mathbf{C}_{m_c} &= \mathbf{C}_m - \mathbf{C}_m \mathbf{G}^\top (\mathbf{G} \mathbf{C}_m \mathbf{G}^\top + \mathbf{C}_e)^{-1} \mathbf{G} \mathbf{C}_m \\ &= \mathbf{C}_m - \Delta \mathbf{C}_m.\end{aligned}\tag{5.11}$$

The matrix $\Delta \mathbf{C}_m$ in Eq. (5.11) is at least positive semidefinite. Therefore, if we assimilate the same data multiple times without modifying $\mathbf{C}_e$, we subtract $\Delta \mathbf{C}_m$ from $\mathbf{C}_m$ multiple times. As a result, we obtain the wrong posterior covariance, and, more importantly, we underestimate the uncertainty in the model parameters.

In fact, MDA was initially introduced in the context of EnKF to assimilate infrequent data, such as production logging and 4D seismic, as outlined in [129]. The inspiration for this approach was drawn from the ideas presented by Rommelse [369]. The integration of MDA with ES was later proposed in [132], which also introduced the general condition for the equivalence of single and multiple data assimilation in the context of linear-Gaussian problems.

5.3.1 MDA for the Linear Case

Emerick and Reynolds [132] derived the general condition to make ES-MDA equivalent to ES for linear-Gaussian problems because ES provides a correct sampling in this case when $N_e \rightarrow \infty$. This condition is described in Proposition 5.1.

Proposition 5.1: MDA in the Linear Case

For the linear-Gaussian case, assimilating the same data multiple times with an inflated covariance matrix of the data errors is equivalent to assimilating data only once with the original data-error covariance matrix as long as the inflation factors satisfy the following condition:

$$\sum_{\ell=1}^{N_a} \frac{1}{\alpha_\ell} = 1 \quad (\text{e.g., } \alpha_\ell = N_a \text{ for } \ell = 1, 2, \dots, N_a).\tag{5.12}$$

To demonstrate Proposition 5.1, we leverage the equivalence between simultaneous and sequential data assimilation. By replicating the entries of the vector $\mathbf{d}_{\text{obs}}$ and matrices $\mathbf{G}$ and $\mathbf{C}_{\mathbf{e}}$ N_a times, we form the following augmented versions

$$\hat{\mathbf{d}}_{\text{obs}} \equiv \begin{bmatrix} \mathbf{d}_{\text{obs}} \\ \vdots \\ \mathbf{d}_{\text{obs}} \end{bmatrix}, \quad (5.13)$$

$$\hat{\mathbf{G}} \equiv \begin{bmatrix} \mathbf{G} \\ \vdots \\ \mathbf{G} \end{bmatrix}, \quad (5.14)$$

and

$$\hat{\mathbf{C}}_{\mathbf{e}} \equiv \begin{bmatrix} \alpha_1 \mathbf{C}_{\mathbf{e}} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \alpha_2 \mathbf{C}_{\mathbf{e}} & \cdots & \mathbf{O} \\ \vdots & & \ddots & \vdots \\ \mathbf{O} & \cdots & & \alpha_{N_a} \mathbf{C}_{\mathbf{e}} \end{bmatrix}, \quad (5.15)$$

where $\mathbf{O}$ is the null matrix.

Using $\hat{\mathbf{G}}$, we can write the augmented predicted data vector $\mathbf{d}$ as

$$\hat{\mathbf{d}}_j = \hat{\mathbf{G}} \mathbf{m}_j, \quad (5.16)$$

where $\mathbf{m}_j \sim \mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_{\mathbf{m}})$, i.e., $\mathbf{m}_j$ is a sample from the prior Gaussian PDF. Now, we formulate the RML objective function for the MDA case as

$$\begin{aligned} \hat{\mathcal{O}}_r(\mathbf{m}) = & \frac{1}{2} (\mathbf{m} - \mathbf{m}_j)^\top \mathbf{C}_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_j) + \\ & \frac{1}{2} \left(\hat{\mathbf{d}}_{\text{obs},j} - \hat{\mathbf{G}} \mathbf{m} \right)^\top \hat{\mathbf{C}}_{\mathbf{e}}^{-1} \left(\hat{\mathbf{d}}_{\text{obs},j} - \hat{\mathbf{G}} \mathbf{m} \right). \end{aligned} \quad (5.17)$$

Requiring the gradient of $\hat{\mathcal{O}}_r(\mathbf{m})$ to vanish, solving for $\mathbf{m}$, and denoting the result as $\hat{\mathbf{m}}$ leads to

$$\hat{\mathbf{m}} = \hat{\mathbf{C}}_{\mathbf{m}_c} \left[\mathbf{C}_{\mathbf{m}}^{-1} \left(\mathbf{m}_{\text{pr}} + \mathbf{C}_{\mathbf{m}}^{1/2} \mathbf{z}_{\mathbf{m}} \right) + \hat{\mathbf{G}}^\top \hat{\mathbf{C}}_{\mathbf{e}}^{-1} \left(\hat{\mathbf{d}}_{\text{obs}} + \hat{\mathbf{C}}_{\mathbf{e}}^{1/2} \hat{\mathbf{z}}_{\mathbf{d}} \right) \right], \quad (5.18)$$

where $\mathbf{z}_{\mathbf{m}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\hat{\mathbf{z}}_{\mathbf{d}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

To establish that $\hat{\mathbf{m}}$ is a sample from the correct posterior PDF, i.e., $\hat{\mathbf{m}} \sim \mathcal{N}(\mathbf{m}_{\text{map}}, \mathbf{C}_{\mathbf{m}_c})$, it suffices to demonstrate the equivalence for the posterior mean and covariance, i.e., $\mathbb{E}[\hat{\mathbf{m}}] = \mathbf{m}_{\text{map}}$ and $\mathbb{C}[\hat{\mathbf{m}}] = \mathbf{C}_{\mathbf{m}_c}$. The equivalence for the posterior covariance follows from

$$\begin{aligned}
 \mathbb{C}[\hat{\mathbf{m}}] &= \left\{ \mathbf{C}_{\mathbf{m}}^{-1} + \hat{\mathbf{G}}^\top \hat{\mathbf{C}}_{\mathbf{e}}^{-1} \hat{\mathbf{G}} \right\}^{-1} \\
 &= \left\{ \mathbf{C}_{\mathbf{m}}^{-1} + [\mathbf{G}^\top \dots \mathbf{G}^\top] \begin{bmatrix} \frac{1}{\alpha_1} \mathbf{C}_{\mathbf{e}}^{-1} & \dots & \mathbf{O} \\ \vdots & \ddots & \vdots \\ \mathbf{O} & \dots & \frac{1}{\alpha_{N_a}} \mathbf{C}_{\mathbf{e}}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{G} \\ \vdots \\ \mathbf{G} \end{bmatrix} \right\}^{-1} \\
 &= \left\{ \mathbf{C}_{\mathbf{m}}^{-1} + \left(\sum_{\ell=1}^{N_a} \frac{1}{\alpha_\ell} \right) \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{G} \right\}^{-1}, \tag{5.19}
 \end{aligned}$$

which is equivalent to $\mathbf{C}_{\mathbf{m}_c}$ if $\sum_{\ell=1}^{N_a} \alpha_\ell^{-1} = 1$.

The equivalence for the posterior mean follows:

$$\begin{aligned}
 \mathbb{E}[\hat{\mathbf{m}}] &= \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} \left(\mathbf{m}_{\text{pr}} + \mathbf{C}_{\mathbf{m}}^{1/2} \mathbb{E}[\mathbf{z}_{\text{m}}] \right) + \hat{\mathbf{G}}^\top \hat{\mathbf{C}}_{\mathbf{e}}^{-1} \left(\hat{\mathbf{d}}_{\text{obs}} + \hat{\mathbf{C}}_{\mathbf{e}}^{1/2} \mathbb{E}[\mathbf{z}_{\text{d}}] \right) \right\} \\
 &= \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_{\text{pr}} + \hat{\mathbf{G}}^\top \hat{\mathbf{C}}_{\mathbf{e}}^{-1} \hat{\mathbf{d}}_{\text{obs}} \right\} \\
 &= \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_{\text{pr}} + [\mathbf{G}^\top \dots \mathbf{G}^\top] \begin{bmatrix} \frac{1}{\alpha_1} \mathbf{C}_{\mathbf{e}}^{-1} & \dots & \mathbf{O} \\ \vdots & \ddots & \vdots \\ \mathbf{O} & \dots & \frac{1}{\alpha_{N_a}} \mathbf{C}_{\mathbf{e}}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{d}_{\text{obs}} \\ \vdots \\ \mathbf{d}_{\text{obs}} \end{bmatrix} \right\} \\
 &= \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_{\text{pr}} + [\mathbf{G}^\top \dots \mathbf{G}^\top] \begin{bmatrix} \frac{1}{\alpha_1} \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs}} \\ \vdots \\ \frac{1}{\alpha_{N_a}} \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs}} \end{bmatrix} \right\} \\
 &= \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_{\text{pr}} + \left(\sum_{\ell=1}^{N_a} \frac{1}{\alpha_\ell} \right) \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs}} \right\} \\
 &= \mathbf{C}_{\mathbf{m}_c} \left\{ \mathbf{C}_{\mathbf{m}}^{-1} \mathbf{m}_{\text{pr}} + \mathbf{G}^\top \mathbf{C}_{\mathbf{e}}^{-1} \mathbf{d}_{\text{obs}} \right\} \\
 &= \mathbf{m}_{\text{map}}. \tag{5.20}
 \end{aligned}$$

The results (5.19) and (5.20) complete the proof of the equivalence for the linear case.

The condition expressed in Proposition 5.1 establishes the equivalence between ES and ES-MDA for linear-Gaussian problems. However, this equivalence does not extend to nonlinear problems. In the context of nonlinear problems, Emerick and Reynolds [132] utilized this condition to regulate the updates applied to the model. The concept involved replacing a single, potentially substantial correction to the model with N_a smaller corrections, as illustrated in Fig. 5.2. In the intervals between these corrections, the forward model is evaluated for all ensemble members to update the covariance matrices, $\tilde{\mathbf{C}}_{\text{md}}^\ell$ and $\tilde{\mathbf{C}}_{\text{dd}}^\ell$, used in the analysis.

The ES-MDA implementation is quite simple requiring very little modification in an ES code. Pseudo-code 5.2 describes the main elements of a standard implementation of ES-MDA.

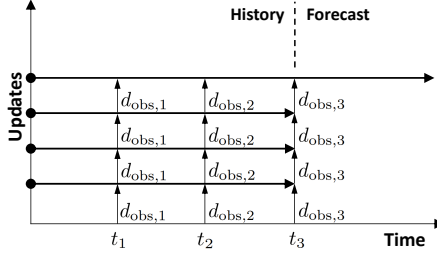


Fig. 5.2: ES-MDA update scheme.

Pseudo-code 5.2: ES-MDA

1. **Initialization:** generate the initial ensemble, $\{\mathbf{m}_j^0\}_{j=1}^{N_e}$, by sampling the prior distribution of parameters. Choose the number of MDA iterations, N_a , and the coefficients $\{\alpha_\ell\}_{\ell=1}^{N_a}$. Set $\ell = 0$.
2. **Forecast step:** run the forward model from time zero until the end of the historical period to compute the vector of predicted data

$$\mathbf{d}_j^\ell = \mathbf{g}(\mathbf{m}_j^\ell), \quad \text{for } j = 1, 2, \dots, N_e, \quad (5.21)$$

3. Perturb the vector of observations:

$$\mathbf{d}_{\text{obs},j}^\ell = \mathbf{d}_{\text{obs}} + \sqrt{\alpha_{\ell+1}} \mathbf{C}_e^{1/2} \mathbf{z}_j^\ell, \quad \text{for } j = 1, 2, \dots, N_e \quad (5.22)$$

where $\mathbf{z}_j^\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

4. **Analysis step:** update the vector of model parameters using

$$\mathbf{m}_j^{\ell+1} = \mathbf{m}_j^\ell + \tilde{\mathbf{C}}_{\text{md}}^\ell \left(\tilde{\mathbf{C}}_{\text{dd}}^\ell + \alpha_{\ell+1} \mathbf{C}_e \right)^{-1} (\mathbf{d}_{\text{obs},j}^\ell - \mathbf{d}_j^\ell), \quad (5.23)$$

for $j = 1, 2, \dots, N_e$.

5. Set $\ell = \ell + 1$.
6. If $\ell = N_a$ then set $\mathbf{m}_{c,j} = \mathbf{m}_j^\ell$ for $j = 1, \dots, N_e$ and stop, else return to step 2.

In ES-MDA, the observed data vector is perturbed using the inflated covariance, meaning that we sample $\mathbf{d}_{\text{obs},j}^\ell \sim \mathcal{N}(\mathbf{d}_{\text{obs}}, \alpha_{\ell+1} \mathbf{C}_e)$. This arises from the derivation, where we expressed $\hat{\mathbf{d}}_{\text{obs},j}^\ell = \hat{\mathbf{d}}_{\text{obs}} + \hat{\mathbf{C}}_e^{1/2} \hat{\mathbf{z}}_d$. This aspect distinguishes ES-MDA from other iterative forms of ES found in the literature. In Pseudo-code 5.2, we resample $\mathbf{d}_{\text{obs},j}^\ell$ at each data assimilation step. The rationale for resampling is rooted in our assumption that $\hat{\mathbf{z}}_d$ is a random vector, which would be violated if we use the same perturbations for

all data assimilations. Additionally, this resampling approach helps mitigate errors arising from matching sampling outliers.

5.3.2 MDA from Bayes' Rule

We derived the condition for the equivalence between single and multiple data assimilation for the linear-Gaussian case using linear algebra. Interestingly, this condition is more general and can also be directly obtained from Bayes' rule. First, we write the posterior PDF as

$$\begin{aligned} p(\mathbf{m}|\mathbf{d}_{\text{obs}}) &= ap(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m}) \\ &= a\mathcal{L}(\mathbf{m})p(\mathbf{m}) \end{aligned} \quad (5.24)$$

where a is a normalizing constant and the likelihood is given by

$$\mathcal{L}(\mathbf{m}) = c \exp(-\mathcal{O}_d(\mathbf{m})) \quad (5.25)$$

with c as a normalizing constant and

$$\mathcal{O}_d(\mathbf{m}) = \frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})). \quad (5.26)$$

Now, we can write a factorization of the likelihood as a product of N_a terms as

$$\mathcal{L}(\mathbf{m}) = \prod_{\ell=1}^{N_a} \mathcal{L}_\ell(\mathbf{m}) \quad (5.27)$$

where

$$\mathcal{L}_\ell(\mathbf{m}) = c_\ell \exp(-\mathcal{O}_d^\ell(\mathbf{m})) \quad (5.28)$$

with

$$\mathcal{O}_d^\ell(\mathbf{m}) = \frac{1}{2} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^\top (\alpha_\ell \mathbf{C}_e)^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})). \quad (5.29)$$

It follows that

$$\begin{aligned} \mathcal{L}(\mathbf{m}) &= \prod_{\ell=1}^{N_a} c_\ell \exp(-\mathcal{O}_d^\ell(\mathbf{m})) \\ &= \prod_{\ell=1}^{N_a} c_\ell \exp\left(-\frac{1}{\alpha_\ell} \mathcal{O}_d(\mathbf{m})\right) \\ &= c \exp\left[-\left(\sum_{\ell=1}^{N_a} \frac{1}{\alpha_\ell}\right) \mathcal{O}_d(\mathbf{m})\right]. \end{aligned} \quad (5.30)$$

The equality in Eq. (5.30) holds if

$$\sum_{\ell=1}^{N_a} \frac{1}{\alpha_\ell} = 1, \quad (5.31)$$

which is the same condition presented in Proposition 5.1 for the linear case. However, no assumptions about the prior or the linearity of data are necessary in this case. This condition arises to ensure that Bayes' rule holds after the factorization of the likelihood. It is important to note, however, that the resulting ES-MDA method still relies on the assumptions of gaussianity and linearity to derive the updating equation based on ES. This derivation also shows that ES-MDA is closely related to annealed importance sampling [315, 316, 414], where the objective is to move to a target distribution via a sequence of tempered transitions.

Remark 5.1: Annealed Importance Sampling

Annealed importance sampling [316] is a Monte Carlo method designed to efficiently sample complex distributions by generating a sequence of intermediate distributions that transition gradually from an initial to the target distribution. The initial distribution, $p_0(\mathbf{m})$, is typically chosen to be easy to sample from, while the intermediate distributions, $p_\ell(\mathbf{m}) = p_0(\mathbf{m})^{\beta_\ell}$, are defined by an annealing parameter $\beta_\ell \in [0, 1]$, often interpreted as a temperature factor, drawing an analogy to thermodynamics. As the annealing process progresses, the intermediate distributions are adjusted using importance sampling and sampled through MCMC methods. The goal is to achieve faster convergence than standard MCMC, particularly when the target distribution has isolated modes, making it difficult to sample directly.

5.3.2.1 Example: Porosity of a Rock Sample

Consider the same example of estimating the porosity of a core sample from a sandstone reservoir described in Section 2.3.1.1. Here, the objective is to compute the posterior variance, $\sigma_{\phi_c}^2$, considering:

- (a) A single data assimilation.
- (b) Assimilating the same observation twice (without the MDA inflation).
- (c) Assimilating the same observation twice (with the MDA inflation).

(a) Single data assimilation: recall the expression derived in Section 2.3.1 for $\mathbf{C}_{\mathbf{m}_c}$ (Eq. (2.46)):

$$\mathbf{C}_{\mathbf{m}_c} = (\mathbf{G}^\top \mathbf{C}_e^{-1} \mathbf{G} + \mathbf{C}_{\mathbf{m}^{-1}})^{-1}.$$

Therefore

$$\frac{1}{\sigma_{\phi_c}^2} = \frac{1}{\sigma_e^2} + \frac{1}{\sigma_{\text{pr}}^2} \quad (5.32)$$

or

$$\sigma_{\phi_c}^2 = \frac{\sigma_e^2 \sigma_{\text{pr}}^2}{\sigma_e^2 + \sigma_{\text{pr}}^2}. \quad (5.33)$$

(b) Assimilation of data twice: after the first assimilation, we have

$$\frac{1}{\sigma_{\phi_1}^2} = \frac{1}{\sigma_e^2} + \frac{1}{\sigma_{\text{pr}}^2} \quad (5.34)$$

or

$$\sigma_{\phi_1}^2 = \frac{\sigma_e^2 \sigma_{\text{pr}}^2}{\sigma_e^2 + \sigma_{\text{pr}}^2} = \sigma_{\phi_c}^2. \quad (5.35)$$

After the second assimilation, we have

$$\begin{aligned} \frac{1}{\sigma_{\phi_2}^2} &= \frac{1}{\sigma_e^2} + \frac{1}{\sigma_{\phi_1}^2} \\ &= \frac{1}{\sigma_e^2} + \frac{1}{\sigma_e^2} + \frac{1}{\sigma_{\text{pr}}^2}. \end{aligned} \quad (5.36)$$

Hence

$$\sigma_{\phi_2}^2 = \frac{\sigma_e^2 \sigma_{\text{pr}}^2}{2\sigma_e^2 + \sigma_{\text{pr}}^2} < \sigma_{\phi_c}^2, \quad (5.37)$$

which means that we underestimate the posterior variance.

(c) Assimilation of data twice with MDA inflation: after the first assimilation, we have

$$\frac{1}{\sigma_{\phi_1}^2} = \frac{1}{\alpha_1 \sigma_e^2} + \frac{1}{\sigma_{\text{pr}}^2}. \quad (5.38)$$

After the second assimilation, we have

$$\begin{aligned} \frac{1}{\sigma_{\phi_2}^2} &= \frac{1}{\alpha_2 \sigma_e^2} + \frac{1}{\sigma_{\phi_1}^2} \\ &= \frac{1}{\alpha_2 \sigma_e^2} + \frac{1}{\alpha_1 \sigma_e^2} + \frac{1}{\sigma_{\text{pr}}^2} \\ &= \underbrace{\left(\frac{1}{\alpha_2} + \frac{1}{\alpha_1} \right)}_{=1} \frac{1}{\sigma_e^2} + \frac{1}{\sigma_{\text{pr}}^2} = \frac{1}{\sigma_e^2} + \frac{1}{\sigma_{\text{pr}}^2}. \end{aligned} \quad (5.39)$$

Hence

$$\sigma_{\phi_2}^2 = \frac{\sigma_e^2 \sigma_{\text{pr}}^2}{\sigma_e^2 + \sigma_{\text{pr}}^2} = \sigma_{\phi_c}^2, \quad (5.40)$$

which means that we recovered the correct posterior variance.

5.3.3 MDA from Regularized Nonlinear Least-Squares

The ES-MDA analysis equation was derived to preserve the ability to correctly sample the multivariate Gaussian posterior that arises in linear settings. However, the same analysis equation can be obtained by approaching the inverse problem from a nonlinear least-squares perspective [209]. In this case, we seek for vector $\mathbf{m}$ that minimizes the following weighted least-squares function

$$\mathcal{O}(\mathbf{m}) = \|\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})\|_{\mathbf{C}_e^{-1}}^2. \quad (5.41)$$

We can derive an iterative procedure for minimizing $\mathcal{O}(\mathbf{m})$ using a first-order Taylor series around the current guess $\mathbf{m}^\ell$ as

$$\begin{aligned} \mathbf{g}(\mathbf{m}) &\approx \mathbf{g}(\mathbf{m}^\ell) + \mathbf{G}_\ell (\mathbf{m} - \mathbf{m}^\ell) \\ &= \mathbf{g}(\mathbf{m}^\ell) + \mathbf{G}_\ell \delta \mathbf{m}^{\ell+1}. \end{aligned} \quad (5.42)$$

Therefore, we can write a linearized version of $\mathcal{O}(\mathbf{m})$ as

$$\mathcal{O}_{\text{ls}}(\delta \mathbf{m}^{\ell+1}) = \|\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}^\ell) - \mathbf{G}_\ell \delta \mathbf{m}^{\ell+1}\|_{\mathbf{C}_e^{-1}}^2. \quad (5.43)$$

Minimization of $\mathcal{O}_{\text{ls}}(\delta \mathbf{m}^{\ell+1})$ is generally ill-posed, so we seek for a regularized solution of the form

$$\mathcal{O}_{\text{rls}}(\delta \mathbf{m}^{\ell+1}) = \|\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}^\ell) - \mathbf{G}_\ell \delta \mathbf{m}^{\ell+1}\|_{\mathbf{C}_e^{-1}}^2 + \alpha \|\delta \mathbf{m}^{\ell+1}\|_{\mathbf{C}_m^{-1}}^2, \quad (5.44)$$

where $\alpha > 0$.

As usual, we set $\nabla \mathcal{O}_{\text{rls}}(\delta \mathbf{m}^{\ell+1}) = \mathbf{0}$, which leads to

$$\begin{aligned} \mathbf{0} &= -\mathbf{G}_\ell^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}^\ell) - \mathbf{G}_\ell \delta \mathbf{m}^{\ell+1}) + \alpha \mathbf{C}_m^{-1} \delta \mathbf{m}^{\ell+1} \\ &= -\mathbf{G}_\ell^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}^\ell)) + (\mathbf{G}_\ell^\top \mathbf{C}_e^{-1} \mathbf{G}_\ell + \alpha \mathbf{C}_m^{-1}) \delta \mathbf{m}^{\ell+1} \\ &= -\mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}^\ell)) (\mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell + \mathbf{C}_m^{-1}) \delta \mathbf{m}^{\ell+1}. \end{aligned} \quad (5.45)$$

Therefore, for $\mathbf{C}_m$ and $\mathbf{C}_e$ positive definite, we have

$$\delta \mathbf{m}^{\ell+1} = (\mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell + \mathbf{C}_m^{-1})^{-1} \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}^\ell)). \quad (5.46)$$

We can show that Eq. (5.46) is equivalent to

$$\delta \mathbf{m}^{\ell+1} = \mathbf{C}_m \mathbf{G}_\ell^\top (\mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top + \alpha \mathbf{C}_e)^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}^\ell)). \quad (5.47)$$

To show this equivalence, it is necessary to show that

$$\begin{aligned} (\mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell + \mathbf{C}_m^{-1})^{-1} \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} = \\ \mathbf{C}_m \mathbf{G}_\ell^\top (\mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top + \alpha \mathbf{C}_e)^{-1}. \end{aligned} \quad (5.48)$$

First note that

$$\begin{aligned} \mathbf{G}_\ell^\top + \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top &= \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} (\alpha \mathbf{C}_e) + \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top \\ &= \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} (\alpha \mathbf{C}_e + \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top). \end{aligned} \quad (5.49)$$

Also

$$\begin{aligned} \mathbf{G}_\ell^\top + \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top &= \mathbf{C}_m^{-1} \mathbf{C}_m \mathbf{G}_\ell^\top + \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top \\ &= (\mathbf{C}_m^{-1} + \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell) \mathbf{C}_m \mathbf{G}_\ell^\top. \end{aligned} \quad (5.50)$$

Since the left-hand side of (5.49) and (5.50) are the same, so are the right-hand side.

$$\mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} (\alpha \mathbf{C}_e + \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top) = (\mathbf{C}_m^{-1} + \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell) \mathbf{C}_m \mathbf{G}_\ell^\top. \quad (5.51)$$

Pre-multiplying both sides of (5.51) by $(\mathbf{C}_m^{-1} + \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell)^{-1}$ gives

$$\begin{aligned} (\mathbf{C}_m^{-1} + \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell)^{-1} \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} (\alpha \mathbf{C}_e + \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top) = \\ \mathbf{C}_m \mathbf{G}_\ell^\top. \end{aligned} \quad (5.52)$$

Pos-multiplying both sides of (5.52) by $(\alpha \mathbf{C}_e + \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top)^{-1}$ gives

$$\begin{aligned} (\mathbf{C}_m^{-1} + \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} \mathbf{G}_\ell)^{-1} \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e)^{-1} = \\ \mathbf{C}_m \mathbf{G}_\ell^\top (\alpha \mathbf{C}_e + \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top)^{-1}, \end{aligned} \quad (5.53)$$

which completes the proof of equivalence between (5.46) and (5.47).

The similarity between Eq. (5.47) and ES-MDA is evident. Using the ensemble approximations $\tilde{\mathbf{C}}_{\text{md}}^\ell \approx \mathbf{C}_m \mathbf{G}_\ell^\top$ and $\tilde{\mathbf{C}}_{\text{dd}}^\ell \approx \mathbf{G}_\ell \mathbf{C}_m \mathbf{G}_\ell^\top$ results in the same update equation of ES-MDA:

$$\mathbf{m}^{\ell+1} = \mathbf{m}^{\ell} + \tilde{\mathbf{C}}_{\text{md}}^{\ell} \left(\tilde{\mathbf{C}}_{\text{dd}}^{\ell} + \alpha \mathbf{C}_{\text{e}} \right)^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}^{\ell})). \quad (5.54)$$

This result reveals the similarity between MDA and regularization methods. This similarity has also practical interest. In fact, it has been used to derive schemes to select the number of data assimilations and the corresponding inflation coefficients [253, 356, 284, 122].

5.3.4 Examples

5.3.4.1 Example: Porosity Distribution in a Core Sample

Consider the same problem introduced in Section 2.3.1.4. Fig. 5.3 presents the results for this problem using ES and ES-MDA with an ensemble of 100 members. For ES-MDA, $N_a = 4$ is used with constant inflation factors. Both methods yield very similar results, closely resembling those obtained with EnKF (Fig. 4.5) and RML (Fig. 2.8), as all these methods converge to the actual Gaussian posterior as the number of samples increases.

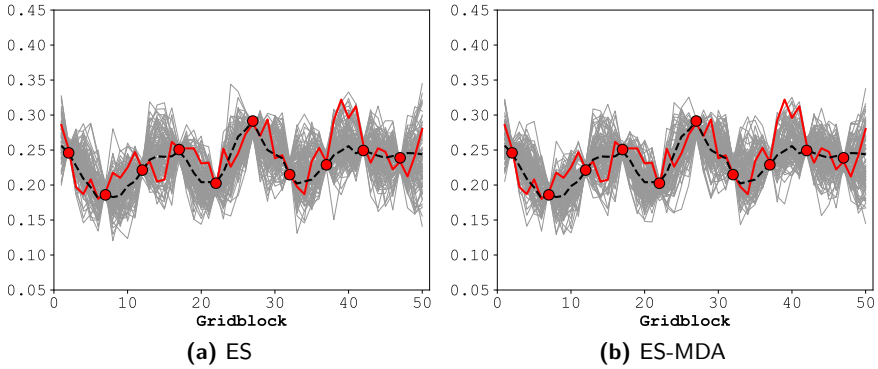


Fig. 5.3: ES and ES-MDA results with $N_e = 100$ for porosity in a core sample. The solid red line is the ground truth. The red circles are the observations. The gray lines are samples from the ES (a) and ES-MDA (b). The dashed lines are the average computed over the 100 samples. Example 5.3.4.1.

5.3.4.2 Example: Nonlinear Univariate Problem

Consider the same univariate nonlinear problem introduced in Section 2.4.5.2. Previously, this problem was solved using MCMC and RML (see Fig. 2.12). Fig. 5.4 presents histograms generated by EnKF and ES-MDA after assim-

ilating one and five data points. Each plot also includes the true posterior distribution for reference. Both cases used an ensemble of 3,000 samples. For ES-MDA, $N_a = 4$ was employed with constant inflation factors.

Unlike MCMC and RML, EnKF and ES-MDA are unable to capture the bimodal nature of the posterior distributions. After assimilating the first data point, both methods tend to cluster samples around the second mode, which is closer to the prior mean. As more data are assimilated, the prominence of the first mode increases. Notably, EnKF struggles to capture this mode, while ES-MDA, though not fully accurate in representing the posterior, achieves significantly better sampling than EnKF.

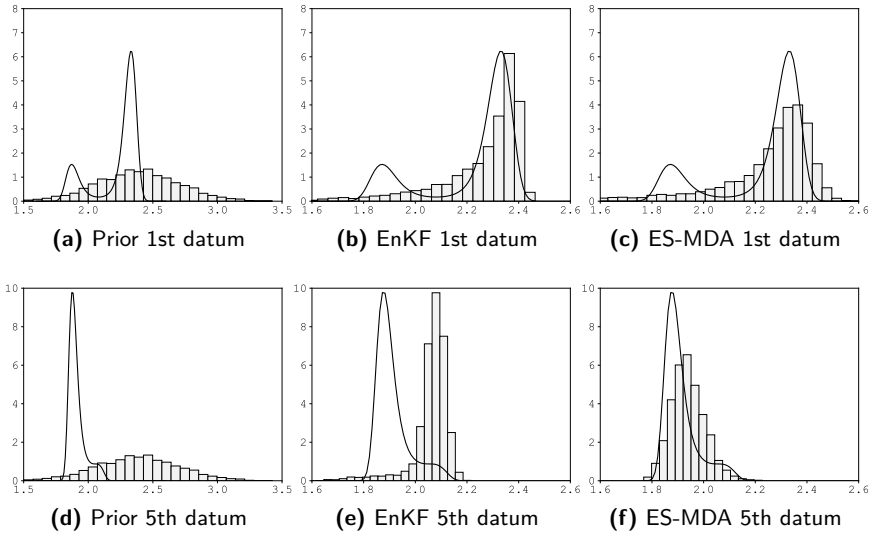


Fig. 5.4: Prior, EnKF and ES-MDA sampling for a nonlinear univariate problem. The black curve in each plot corresponds to the correct posterior distribution. Example 5.3.4.2.

5.3.4.3 Example: Production Logging in a Single Phase Flow

Emerick and Reynolds [132] compared ES and ES-MDA in a single-phase problem involving a reservoir with a single well operating at a constant liquid production rate of 1,000 bbl/day. The objective was to estimate the log-permeability for each of the 40 layers of the model based on liquid rate measurements per layer at a single time step, simulating a production-logging dataset. Fig. 5.5 shows the predicted layer rates obtained with a prior ensemble of 100 realizations and after applying ES and ES-MDA with two and four data assimilations. The results indicate that while the single update from ES led to noticeable improvements in the predicted layer-rate data,

significant mismatches remained. ES-MDA with $N_a = 2$ provided further improvements, and ES-MDA with $N_a = 4$ achieved a near-perfect match with the observations.

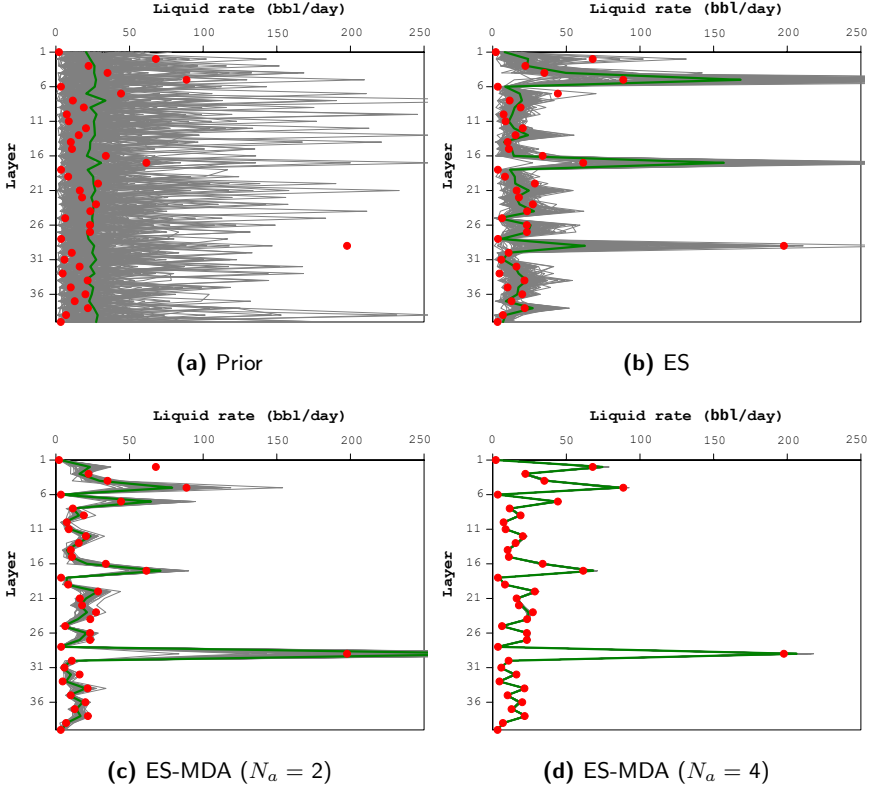


Fig. 5.5: Liquid rate for each reservoir layer. Red circles represent observed data, gray lines show the predicted data, and green lines represent the ensemble mean. Example 5.3.4.3. Reproduced from Emerick and Reynolds [132] with permission from Elsevier.

5.3.4.4 Example: Permeability in a Two-Phase Linear Flow

Consider the same problem discussed in Section 2.4.5.3 corresponding to a two-phase linear flow. Fig. 5.6 presents the percentiles of the posterior marginal distributions of log-permeability obtained with MCMC and ES-MDA, while Fig. 5.7 shows the corresponding distributions obtained for predicted water production rate. ES-MDA results considered an ensemble of 100 realizations with ten iterations and constant inflation factors. The results in Figs. 5.6b and 5.7b are based on a single ensemble, which is different from

[131], where it is shown similar plots combining ten ensembles with $N_e = 100$ each. The results in these figures show that ES-MDA resulted in reasonable estimates of the posterior marginal distributions of log-permeability and water production rate for this problem.

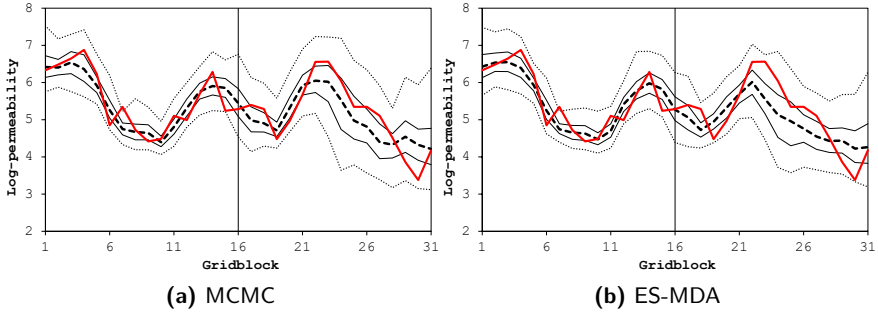


Fig. 5.6: Marginal log-permeability distributions for a two-phase linear flow. The red line represents the ground truth, the dashed line denotes the median, and the solid black lines correspond to the 25th and 75th percentiles. The outer bounding lines encompass the 2nd and 98th percentiles. Example 5.3.4.4. Reproduced from Emerick and Reynolds [131] with permission from Springer Nature.

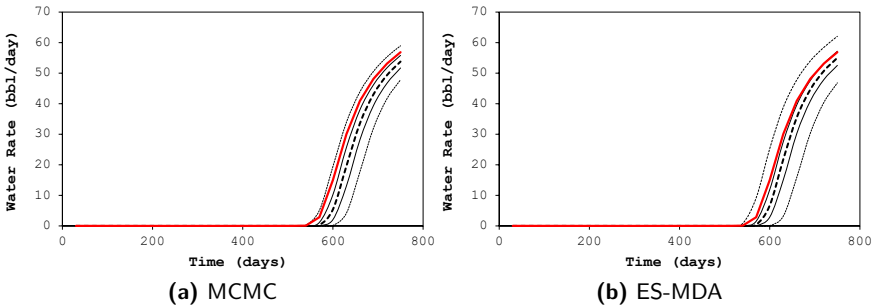


Fig. 5.7: Predicted water production rate. The red line represents the ground truth, the dashed line denotes the median, and the solid black lines correspond to the 25th and 75th percentiles. The outer bounding lines encompass the 2nd and 98th percentiles. Example 5.3.4.4. Reproduced from Emerick and Reynolds [131] with permission from Springer Nature.

5.3.4.5 Example: Two-Phase Flow in 2D Model

Consider the same data assimilation problem of two-phase flow in a 2D model previously discussed in Chapters 2 and 4. Section 2.4.5.4 presented the MAP estimate and RML realizations, while Section 4.3.4.1 presented the EnKF results. Here, we consider the ES-MDA results. The data assimilation used $N_a = 4$ with constant inflation factors to update an ensemble with $N_e = 100$ realizations².

Fig. 5.8 shows the first three prior and posterior realizations of the ensemble, along with the true log-permeability. Similarly to the EnKF, ES-MDA data assimilation recovered the main permeability features of the ground truth. Fig. 5.9 presents the predicted water production rate for three wells, demonstrating improvements in both data matching and production forecast.

5.3.4.6 Example: EnKF, ES and ES-MDA in the Brugge Case

Emerick and Reynolds [132] compared the performance of the EnKF, ES, and ES-MDA (with $N_a = 4$) for assimilating production data in the Brugge case [346]. Fig. 5.10 illustrates a single realization of log-permeability both before and after data assimilation using these methods. All methods resulted in an increase in the permeability of the prior model. Fig. 5.11 presents the water rate data for well P5, highlighting that ES did not achieve a satisfactory improvement in the data matching. The EnKF results are closer to the observations but exhibit a slight bias, whereas ES-MDA provides a good match to the data.

5.3.5 How Many Data Assimilations?

In the original ES-MDA method, the value for N_a must be specified in advance. If, for example, we select a value for N_a and, after completing these N_a iterations, we find that additional iterations are needed, we cannot simply add more iterations as it would violate the condition $\sum_{\ell=1}^{N_a} \alpha_\ell^{-1} = 1$. Instead, a new data assimilation process must be initiated, starting from the prior ensemble with a larger value for N_a (and compatible values for α_ℓ).

Practical experience from synthetic and field cases suggests that setting $N_a = 4$ with $\alpha_1 = \dots \alpha_4 = 4$ is a good initial guess. However, it is essential to note that there are instances where $N_a = 4$ may not be sufficient. For example, the synthetic problem presented in [131] necessitated at least $N_a = 10$ for reasonable results. Additionally, the resemblance between MDA

² As with the EnKF case, the data assimilation with ES-MDA used distance-based localization to regularize the updates; see Chapter 7.

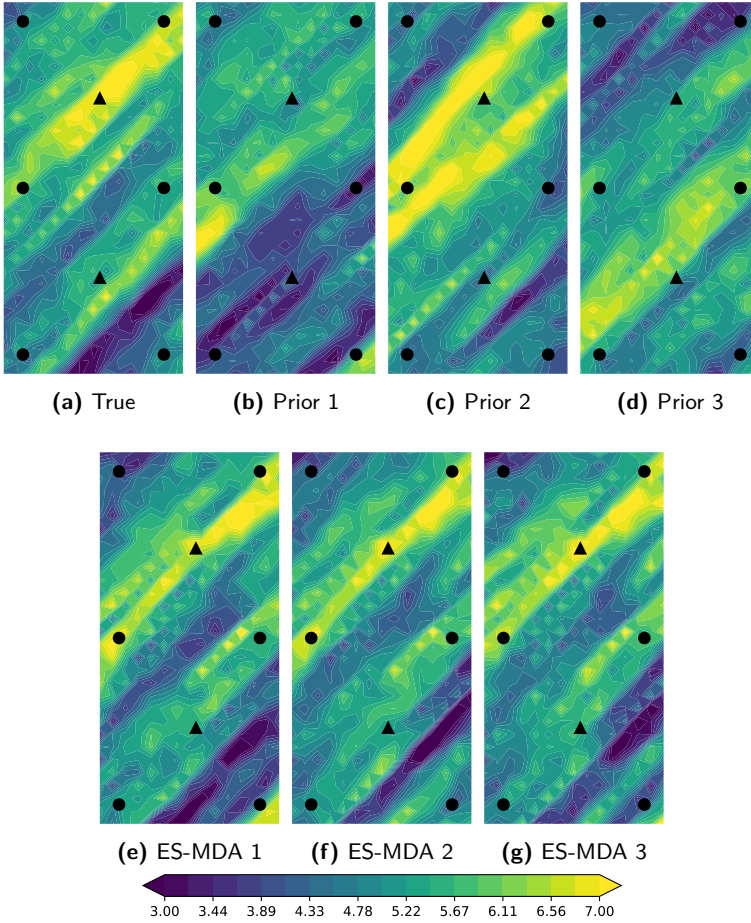


Fig. 5.8: True log-permeability, three prior realizations, and three posterior realizations obtained using ES-MDA for a 2D model under two-phase flow. Black circles indicate the positions of producers, and black triangles indicate the positions of water injection wells. Example 5.3.4.5.

and regularization methods implies that initiating the data assimilation with large values for α_ℓ and gradually decreasing them during the process may enhance the quality of the final models. This approach is rooted in the observation that at the beginning of a data assimilation process, the predicted data are typically distant from the observations. This situation may lead to overcorrections in the values of the model parameters. In this scenario, large values of α_ℓ work as a regularization coefficient, damping the model corrections at early iterations.

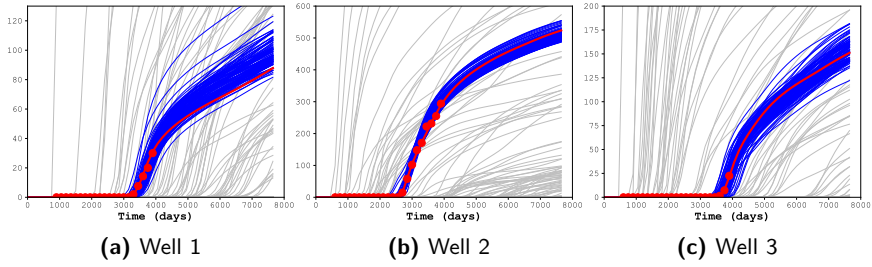


Fig. 5.9: Water production rate in bbl/day for three wells. Red circles are the observations, and the red line is the prediction with the true log-permeability distribution. The gray and blue lines show the predictions from 100 prior and posterior realizations, respectively. The posterior realizations were obtained with ES-MDA. Example 5.3.4.5.

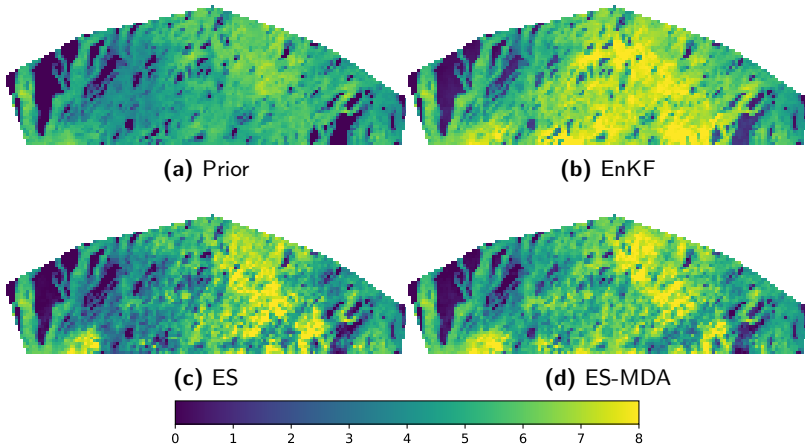


Fig. 5.10: Log-permeability for the first layer of one realization of the Brugge case. Example 5.3.4.6. Reproduced from Emerick and Reynolds [132] with permission from Elsevier.

5.3.5.1 Adaptive Selection of ES-MDA Inflation

Emerick [118] introduced a straightforward adaptive procedure for choosing N_a and α_ℓ based on the evolution of the data-mismatch objective function. This method involves specifying the maximum number of data assimilations, denoted as $N_{a,\max}$, the maximum inflation factor $\alpha_{\max}$, and a coefficient a_0 within the range $(0, 1]$, for example, $a_0 = 0.25$. The steps of this adaptive procedure are outlined in Pseudo-code 5.3.

The adaptive scheme outlined in Pseudo-code 5.3 was originally proposed based on practical observations without a formal mathematical justification. However, a more recent work by Iglesias and Yang [208] has provided a theoretical basis for a similar scheme. The interpretation involves considering

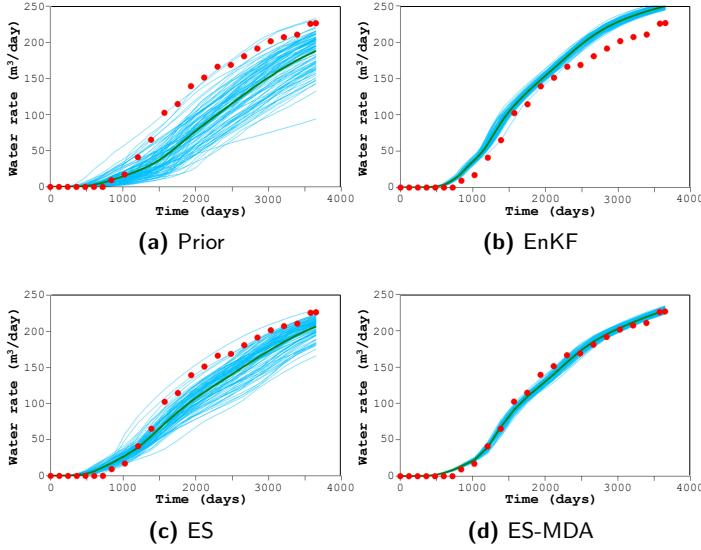


Fig. 5.11: Water production rate for well P5 of the Brugge case. Red circles are the observations, blue lines are the predicted data and the green line is the ensemble mean prediction. Example 5.3.4.6. Reproduced from Emerick and Reynolds [132] with permission from Elsevier.

the process as a Bayesian tempering process [315], where the updates are controlled by imposing a threshold on the Jeffreys' divergence [214] between two consecutive updates. The threshold is determined using the discrepancy principle [306, 184]. The resulting expression for selection α_ℓ is given by

$$\alpha_{\ell+1}^{-1} = \min \left\{ \max \left\{ \frac{N_d}{2\bar{\mathcal{O}}_d^\ell}, \sqrt{\frac{N_d}{2\mathbb{V}[\mathcal{O}_d^\ell]}} \right\}, 1 - \beta_\ell \right\}, \quad (5.55)$$

where $\bar{\mathcal{O}}_d^\ell$ and $\mathbb{V}[\mathcal{O}_d^\ell]$ are the mean and variance of the data mismatch objective function computed over the N_e realizations of the ensemble, and $\beta_\ell = \sum_{i=1}^\ell \alpha_i^{-1}$.

Two additional adaptive MDA schemes are presented in [253]. The first one follows the procedure outlined in Pseudo-code 5.3, incorporating an additional check for the maximum allowed change in model parameters. If the change in any parameter for any ensemble member exceeds twice its prior standard deviation, the update is rejected. Subsequently, a new update is attempted with a doubled value of $\alpha_{\ell+1}$. This process is iterated until no further violations occur. It's important to note that this scheme does not require new reservoir simulations; rather, it involves repeated iterations of the analysis step.

Pseudo-code 5.3: Adaptive ES-MDA

1. **Initialization:** generate the initial ensemble, $\{\mathbf{m}_j^0\}_{j=1}^{N_e}$. Choose the values for $\alpha_{\max}$, $N_{a,\max}$ and a_0 . Set $\ell = 0$ and $\beta_0 = 0$ and

$$a = \min \left\{ \frac{\alpha_{\max}}{\bar{\mathcal{O}}_{N,d}^0}, a_0 \right\}. \quad (5.56)$$

2. **Forecast step:** run the forward model from time zero until the end of the historical period to compute the vector of predicted data

$$\mathbf{d}_j^\ell = \mathbf{g}(\mathbf{m}_j^\ell), \quad \text{for } j = 1, 2, \dots, N_e. \quad (5.57)$$

3. Compute the average normalized data-mismatch objective function

$$\bar{\mathcal{O}}_{N,d}^\ell = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathcal{O}_{N,d}(\mathbf{m}_j^\ell), \quad \text{where} \quad (5.58)$$

$$\mathcal{O}_{N,d}(\mathbf{m}_j^\ell) = \frac{1}{2N_d} (\mathbf{d}_{\text{obs}} - \mathbf{d}_j^\ell)^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{d}_j^\ell). \quad (5.59)$$

4. Compute:

$$\alpha_{\ell+1} = \min \left\{ a \bar{\mathcal{O}}_{N,d}^\ell, \alpha_{\max} \right\}. \quad (5.60)$$

$$\beta_{\ell+1} = \beta_\ell + \frac{1}{\alpha_{\ell+1}}. \quad (5.61)$$

5. If $\left(\beta_{\ell+1} > 1 - \frac{1}{\alpha_{\max}} \right)$ or $(\ell = N_{a,\max} - 1)$ then

- Recompute the inflation factor using

$$\alpha_{\ell+1} = \frac{1}{1 - \beta_\ell}. \quad (5.62)$$

- Set $\beta_{\ell+1} = 1$.

6. Perturb the vector of observation:

$$\mathbf{d}_{\text{obs},j}^\ell = \mathbf{d}_{\text{obs}} + \sqrt{\alpha_{\ell+1}} \mathbf{C}_e^{1/2} \mathbf{z}_j^\ell, \quad (5.63)$$

for $j = 1, 2, \dots, N_e$, where $\mathbf{z}_j^\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

7. **Analysis step:** update the vector of model parameters using

$$\mathbf{m}_j^{\ell+1} = \mathbf{m}_j^\ell + \tilde{\mathbf{C}}_{\text{md}}^\ell \left(\tilde{\mathbf{C}}_{\text{dd}}^\ell + \alpha_{\ell+1} \mathbf{C}_e \right)^{-1} (\mathbf{d}_{\text{obs},j}^\ell - \mathbf{d}_j^\ell), \quad (5.64)$$

for $j = 1, 2, \dots, N_e$.

8. If $\beta_{\ell+1} = 1$ then stop, else set $\ell = \ell + 1$ and return to step 2.

The second adaptive scheme introduced in [253] draws inspiration from the earlier work by Iglesias and Dawson [209], which employed the regularized Levenberg-Marquardt scheme from [184]. In this scheme, the acceptance of an update is contingent on satisfying the following condition for all ensemble members:

$$\rho^2 \left\| \mathbf{C}_e^{-1/2} (\mathbf{d}_{\text{obs},j} - \mathbf{d}_j^\ell) \right\|^2 \leq \alpha_{\ell+1}^2 \left\| \mathbf{C}_e^{1/2} \left(\tilde{\mathbf{C}}_{\text{dd}}^\ell + \alpha_{\ell+1} \mathbf{C}_e \right)^{-1} (\mathbf{d}_{\text{obs},j}^\ell - \mathbf{d}_j^\ell) \right\|^2, \quad (5.65)$$

for $\rho \in (0, 1)$. Large values for ρ require larger $\alpha_{\ell+1}$, and consequently, more MDA iterations are necessary. However, this also introduces more regularization into the process, which should reduce overcorrections in the models, resulting in smoother updates. The authors used $\rho = 0.2$ in their example. If the condition (5.65) is not satisfied, the update is repeated by doubling the value of $\alpha_{\ell+1}$.

5.3.5.2 Geometric Selection of ES-MDA Inflation

Practical experience with the adaptive schemes proposed in [118, 253] has revealed that they frequently result in a sluggish data assimilation process, demanding a substantial number of iterations. Consequently, this makes the method less appealing for large-scale applications due to the associated high computational costs. Motivated by this issue, Rafiee and Reynolds [356] proposed selecting α_ℓ based on a geometrically decreasing sequence such that

$$\alpha_{\ell+1} = \gamma \alpha_\ell = \gamma^\ell \alpha_1 \quad (5.66)$$

for $\gamma \in (0, 1]$. For a given value of α_1 and N_a , we can compute γ by solving $f(\gamma) = 0$, where

$$f(\gamma) = \sum_{\ell=1}^{N_a} \frac{1}{\gamma^{\ell-1} \alpha_1} - 1. \quad (5.67)$$

This equation can be solved using, for example, the bisection method.

To motivate the selection of α_1 , it is convenient to express the first MDA iteration for updating the ensemble mean as follows:

$$\begin{aligned}
 \bar{\mathbf{m}}^1 &= \bar{\mathbf{m}}^0 + \Delta \mathbf{M}^0 (\Delta \mathbf{D}^0)^\top \left(\Delta \mathbf{D}^0 (\Delta \mathbf{D}^0)^\top + \alpha_1 \mathbf{C}_e \right)^{-1} \left(\mathbf{d}_{\text{obs}} - \overline{\mathbf{g}(\mathbf{m}^0)} \right) \\
 &= \bar{\mathbf{m}}^0 + \Delta \mathbf{M}^0 \left(\mathbf{C}_e^{-1/2} \Delta \mathbf{D}^0 \right)^\top \left[\left(\mathbf{C}_e^{-1/2} \Delta \mathbf{D}^0 \right) \left(\mathbf{C}_e^{-1/2} \Delta \mathbf{D}^0 \right)^\top + \alpha_1 \mathbf{I} \right]^{-1} \times \\
 &\quad \mathbf{C}_e^{-1/2} \left(\mathbf{d}_{\text{obs}} - \overline{\mathbf{g}(\mathbf{m}^0)} \right)
 \end{aligned} \tag{5.68}$$

or

$$\begin{aligned}
 (\Delta \mathbf{M}^0)^+ (\bar{\mathbf{m}}^1 - \bar{\mathbf{m}}^0) &= \left(\mathbf{C}_e^{-1/2} \Delta \mathbf{D}^0 \right)^\top \times \\
 &\quad \left(\left(\mathbf{C}_e^{-1/2} \Delta \mathbf{D}^0 \right) \left(\mathbf{C}_e^{-1/2} \Delta \mathbf{D}^0 \right)^\top + \alpha_1 \mathbf{I} \right)^{-1} \times \\
 &\quad \mathbf{C}_e^{-1/2} \left(\mathbf{d}_{\text{obs}} - \overline{\mathbf{g}(\mathbf{m}^0)} \right),
 \end{aligned} \tag{5.69}$$

where $\mathbf{m}^0$ denotes a prior realization and $(\Delta \mathbf{M}^0)^+$ is the pseudo-inverse of $\Delta \mathbf{M}^0$. Now call

$$\hat{\mathbf{m}} = (\Delta \mathbf{M}^0)^+ (\bar{\mathbf{m}}^1 - \bar{\mathbf{m}}^0), \tag{5.70}$$

$$\hat{\mathbf{d}} = \mathbf{C}_e^{-1/2} \left(\mathbf{d}_{\text{obs}} - \overline{\mathbf{g}(\mathbf{m}^0)} \right) \tag{5.71}$$

and

$$\widehat{\Delta \mathbf{D}} = \mathbf{C}_e^{-1/2} \Delta \mathbf{D}^0. \tag{5.72}$$

Hence, Eq. (5.69) can be written as

$$\hat{\mathbf{m}} = \widehat{\Delta \mathbf{D}}^\top \left(\widehat{\Delta \mathbf{D}} \widehat{\Delta \mathbf{D}}^\top + \alpha_1 \mathbf{I} \right)^{-1} \hat{\mathbf{d}}. \tag{5.73}$$

Eq. (5.73) represents the solution of a regularized linear least-squares problem (see Appendix B, Section B.4.4), where α_1 plays the role of the Tikhonov regularization parameter.

Rafiee and Reynolds [356] used the regularization condition of Eq. (5.65) as starting point to derive an expression for α_1 . We can write Eq. (5.65) as

$$\rho^2 \leq \alpha_1^2 \frac{\left\| \mathbf{C}^{-1} \hat{\mathbf{d}} \right\|^2}{\left\| \hat{\mathbf{d}} \right\|^2} \tag{5.74}$$

where $\mathbf{C} = \left(\widehat{\Delta \mathbf{D}} \widehat{\Delta \mathbf{D}}^\top + \alpha_1 \mathbf{I} \right)$. Then, note that

$$\rho^2 \leq \alpha_1^2 \max_i \lambda_i^2 = \alpha_1^2 \max_i \frac{1}{(\sigma_i^2 + \alpha_1)^2} \tag{5.75}$$

with λ_i denoting the i th eigenvalue of $\mathbf{C}$ and σ_i the i th singular value of $\widehat{\Delta \mathbf{D}}^3$. Instead of selecting α_1 that satisfies Eq. (5.75), Rafiee and Reynolds [356] use

$$\rho^2 \leq \alpha_1^2 \frac{1}{(\bar{\sigma}^2 + \alpha_1)^2}. \quad (5.76)$$

The equality occurs for

$$\alpha_1 = \frac{\rho}{1 - \rho} \bar{\sigma}^2, \quad (5.77)$$

where

$$\bar{\sigma} = \frac{1}{N_r} \sum_{i=1}^{N_r} \sigma_i, \quad (5.78)$$

with N_r denoting the number of non-zero singular values of $\mathbf{C}$. Rafiee and Reynolds [356] proposed to adopt $\rho = 0.5$, which leads to $\alpha_1 = \bar{\sigma}^2$. The resulting procedure is summarized in Pseudo-code 5.4.

Pseudo-code 5.4: Geometric MDA 1

1. Specify N_a .
2. Compute

$$\alpha_1 = \max \{ \bar{\sigma}^2, N_a \}, \quad (5.79)$$

where

$$\bar{\sigma} = \frac{1}{N_r} \sum_{i=1}^{N_r} \sigma_i. \quad (5.80)$$

3. Compute $\gamma \in (0, 1]$ by solving

$$f_1(\gamma) = \sum_{k=1}^{N_a} \frac{1}{\gamma^{k-1} \alpha_1} - 1 = 0 \quad (5.81)$$

using the bisection method.

4. Apply ES-MDA with $\alpha_{\ell+1} = \gamma^\ell \alpha_1$.

Emerick [122] proposed a similar procedure based on a geometric sequence of α_ℓ which requires the specification of the inflation factor at the last data assimilation, $\alpha_{N_a} = \alpha_{\min}$. The motivation for this procedure is to ensure that the geometric scheme $\alpha_{\ell+1} = \gamma^\ell \alpha_1$ does not result in a sequence that decays too rapidly. The value of α_1 is computed using

³ See Appendix A for a brief review on linear algebra.

$$\alpha_1 = \gamma^{1-N_a} \alpha_{N_a}, \quad (5.82)$$

where γ is obtained by solving

$$f_2(\gamma) = \sum_{\ell=1}^{N_a} \frac{1}{\gamma^{\ell-N_a} \alpha_{N_a}} - 1, \quad (5.83)$$

using the bisection method. The value α_1 obtained with Eq. (5.82) is checked using the Morozov's discrepancy principle (MDP) [306]. If the current value of α_1 is not sufficiently large to satisfy the MDP, N_a is increased, and a new value for α_1 is computed.

The MDP states that it is not reasonable to expect the solution to yield an error smaller than the noise level in the data. Therefore, we can define a discrepancy function as

$$h(\alpha) = \left\| \widehat{\Delta \mathbf{D}} \widehat{\mathbf{m}} - \widehat{\mathbf{d}} \right\|^2 - \eta^2, \quad (5.84)$$

where η is the noise level. Here, we assume that $\eta^2 = \mathbb{E}[\|\widehat{\mathbf{e}}\|^2] = N_d$, where $\widehat{\mathbf{e}}$ is the normalized data-error vector.

Using Eq. (5.84) we can compute $\alpha^* > 0$ such that $h(\alpha^*) = 0$ by solving a root-finding problem. We start by writing the singular value decomposition of $\widehat{\Delta \mathbf{D}}$ as

$$\widehat{\Delta \mathbf{D}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top, \quad (5.85)$$

where $\mathbf{U} \in \mathbb{R}^{N_d \times N_d}$ and $\mathbf{V} \in \mathbb{R}^{N_m \times N_m}$ are orthogonal matrices and $\mathbf{\Sigma} \in \mathbb{R}^{N_d \times N_m}$ is a matrix with diagonal element given by

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{N_r} > \sigma_{N_r+1} = \dots = \sigma_{\min\{N_d, N_m\}} = 0. \quad (5.86)$$

Replacing Eq. (5.85) in (5.84) and using the properties $\mathbf{U} \mathbf{U}^\top = \mathbf{I}$ and $\mathbf{V} \mathbf{V}^\top = \mathbf{I}$ results in

$$\begin{aligned} h(\alpha) &= \left\| \widehat{\Delta \mathbf{D}} \widehat{\mathbf{m}} - \widehat{\mathbf{d}} \right\|^2 - \eta^2 \\ &= \left\| \mathbf{U}^\top \left(\widehat{\Delta \mathbf{D}} \mathbf{V} \mathbf{V}^\top \widehat{\mathbf{m}} - \widehat{\mathbf{d}} \right) \right\|^2 - \eta^2 \\ &= \left\| \mathbf{\Sigma} \mathbf{V}^\top \widehat{\mathbf{m}} - \mathbf{U}^\top \widehat{\mathbf{d}} \right\|^2 - \eta^2 \\ &= \sum_{i=1}^{N_r} \left(\sigma_i \mathbf{v}_i^\top \widehat{\mathbf{m}} - \mathbf{u}_i^\top \widehat{\mathbf{d}} \right)^2 + \sum_{i=N_r+1}^{N_d} \left(\mathbf{u}_i^\top \widehat{\mathbf{d}} \right)^2 - \eta^2. \end{aligned} \quad (5.87)$$

The vector $\widehat{\mathbf{m}}$ can be written as

$$\begin{aligned}
 \hat{\mathbf{m}} &= \left(\widehat{\Delta \mathbf{D}}^\top \widehat{\Delta \mathbf{D}} + \alpha \mathbf{I} \right)^{-1} \widehat{\Delta \mathbf{D}}^\top \hat{\mathbf{d}} \\
 &= \left(\mathbf{V} \boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} \mathbf{V}^\top + \alpha \mathbf{I} \right)^{-1} \mathbf{V} \boldsymbol{\Sigma}^\top \mathbf{U}^\top \hat{\mathbf{d}} \\
 &= \mathbf{V} \left(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma} + \alpha \mathbf{I} \right) \boldsymbol{\Sigma}^\top \mathbf{U}^\top \hat{\mathbf{d}} \\
 &= \sum_{j=1}^{N_r} \left(\frac{\sigma_j}{\sigma_j^2 + \alpha} \mathbf{u}_j^\top \hat{\mathbf{d}} \right) \mathbf{v}_j.
 \end{aligned} \tag{5.88}$$

Using Eq. (5.88) in (5.87) results in

$$\begin{aligned}
 h(\alpha) &= \sum_{i=1}^{N_r} \left[\sigma_i \mathbf{v}_i^\top \left(\sum_{j=1}^{N_r} \left(\frac{\sigma_j}{\sigma_j^2 + \alpha} \mathbf{u}_j^\top \hat{\mathbf{d}} \right) \mathbf{v}_j \right) - \mathbf{u}_i^\top \hat{\mathbf{d}} \right]^2 + \sum_{i=N_r+1}^{N_d} \left(\mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2 - \eta^2 \\
 &= \sum_{i=1}^{N_r} \left[\frac{\sigma_i^2}{\sigma_i^2 + \alpha} \mathbf{u}_i^\top \hat{\mathbf{d}} - \mathbf{u}_i^\top \hat{\mathbf{d}} \right]^2 + \sum_{i=N_r+1}^{N_d} \left(\mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2 - \eta^2 \\
 &= \sum_{i=1}^{N_r} \left(\frac{\alpha}{\sigma_i^2 + \alpha} \mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2 + \sum_{i=N_r+1}^{N_d} \left(\mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2 - \eta^2.
 \end{aligned} \tag{5.89}$$

We can use the Newton-Raphson method [350] to compute α^* such that $h(\alpha^*) = 0$, in which case it is necessary to compute the derivative

$$\frac{dh(\alpha)}{d\alpha} = \sum_{i=1}^{N_r} \frac{2\alpha\sigma_i^2}{(\sigma_i^2 + \alpha)^3} \left(\mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2. \tag{5.90}$$

Note that $\frac{dh(\alpha)}{d\alpha} > 0$ for any $\alpha > 0$. Therefore, $h(\alpha)$ is a strictly increasing function of α . Moreover, note that

$$\lim_{\alpha \rightarrow 0^+} h(\alpha) = \sum_{i=N_r+1}^{N_d} \left(\mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2 - \eta^2 = \left\| \mathbf{U}_0^\top \hat{\mathbf{d}} \right\|^2 - \eta^2 \tag{5.91}$$

and

$$\lim_{\alpha \rightarrow \infty} h(\alpha) = \sum_{i=1}^{N_d} \left(\mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2 - \eta^2 = \left\| \hat{\mathbf{d}} \right\|^2 - \eta^2. \tag{5.92}$$

Therefore, $h(\alpha) = 0$ has a unique solution as long as the noise level satisfies

$$\left\| \mathbf{U}_0^\top \hat{\mathbf{d}} \right\|^2 \leq \eta^2 \leq \left\| \hat{\mathbf{d}} \right\|^2. \tag{5.93}$$

The matrix $\mathbf{U}_0$ forms a basis for $\Re(\widehat{\Delta \mathbf{D}}^\top)$. The left-hand side of Eq. (5.93) can be interpreted that components of $\hat{\mathbf{d}}$ orthogonal to $\Re(\widehat{\Delta \mathbf{D}}^\top)$ are ex-

pected to be due to noise [226]⁴. However, in ensemble data assimilation, the components of $\widehat{\Delta \mathbf{D}}$ are computed from an ensemble of size N_e . Therefore, $\text{rank}(\widehat{\Delta \mathbf{D}}) \leq \min\{N_d, N_e - 1\}$. If $N_d \gg N_e$, as is often the case for reservoir problems, then $\dim \mathfrak{N}(\widehat{\Delta \mathbf{D}}^\top) \gg N_e$. Consequently, there may be cases where $\|\mathbf{U}_0^\top \widehat{\mathbf{d}}\|^2 > \eta^2$, meaning that the MDP would not be violated for any $\alpha > 0$. To prevent this, a safe simplification was introduced in the proposed method by dropping the term $\sum_{i=N_r+1}^{N_d} (\mathbf{u}_i^\top \widehat{\mathbf{d}})^2$ in Eq. (5.89). This simplification is justified by the fact that, in practice, localization is always used, which expands the space for reconstructing $\widehat{\mathbf{m}}^5$. The right-hand side of (5.93) implies that the noise level is expected not to exceed the signal [226].

Pseudo-code 5.5 summarizes the geometric MDA process.

Pseudo-code 5.5: Geometric MDA 2

1. Specify α_{N_a} , $\alpha_{\max}$ and N_a .
2. Compute $\gamma \in (0, 1]$ by solving

$$f_2(\gamma) = \sum_{\ell=1}^{N_a} \frac{1}{\gamma^{\ell-N_a} \alpha_{N_a}} - 1 = 0 \quad (5.94)$$

using the bisection method.

3. Compute

$$\alpha_1 = \gamma^{1-N_a} \alpha_{N_a}. \quad (5.95)$$

4. Compute $\alpha^* \in [N_a, \alpha_{\max}]$ by solving $h(\alpha) = 0$ using the Newton-Raphson method.
5. If $\alpha_1 < \alpha^*$ set $N_a = N_a + 1$ and return to step 2.
Otherwise apply ES-MDA with $\alpha_{\ell+1} = \gamma^\ell \alpha_1$.

Pseudo-code 5.6 presents the Newton-Raphson to compute α^* by solving $h(\alpha^*) = 0$.

⁴ $\mathfrak{R}(\cdot)$ and $\mathfrak{N}(\cdot)$ denote the range and null space. See Appendix A, Section A.2.

⁵ Localization is discussed in Chapter 7, Section 7.5.

Pseudo-code 5.6: Newton-Rapshon to compute α^*

1. Set $\alpha_{\min} = N_a$ and specify $\alpha_{\max}$ and $N_{\max}$.
2. Compute

$$h(\alpha_{\min}) = \sum_{i=1}^{N_r} \left(\frac{\alpha_{\min}}{\sigma_i^2 + \alpha_{\min}} \mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2 - N_d \quad (5.96)$$

3. If $h(\alpha_{\min}) \geq 0$ set $\alpha^* = \alpha_{\min}$ and exit.
4. Compute

$$\alpha^1 = \frac{\alpha_{\min} + \alpha_{\max}}{2} \quad (5.97)$$

5. For $n = 1$ to $N_{\max}$ do

- Compute:

$$h(\alpha^n) = \sum_{i=1}^{N_r} \left(\frac{\alpha^n}{\sigma_i^2 + \alpha^n} \mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2 - N_d \quad (5.98)$$

$$h'(\alpha^n) = \sum_{i=1}^{N_r} \frac{2\alpha^n \sigma_i^2}{(\sigma_i^2 + \alpha^n)^3} \left(\mathbf{u}_i^\top \hat{\mathbf{d}} \right)^2$$

$$\alpha^{n+1} = \alpha^n - \frac{h(\alpha^n)}{h'(\alpha^n)} \quad (5.99)$$

- If $\alpha^{n+1} \geq \alpha_{\max}$ set $\alpha^* = \alpha_{\max}$ and exit,
 else if $|\alpha^{n+1} - \alpha^n| < 10^{-3}$ set $\alpha^* = \alpha^{n+1}$ and exit.

6. Set $\alpha^* = \alpha^n$.

5.3.5.3 Example: Geometric Selection of ES-MDA Inflation in the UNISIM-I-H Case

Emerick [122] tested the geometric selection of ES-MDA inflation on the UNISIM-I-H benchmark case [20], a synthetic reservoir based on actual data from the Namorado Field (Campos Basin). Fig. 5.12 presents a plot of the average data-mismatch norm versus the average model-change norm of log-permeability for data assimilations using $N_a = 4$ and $N_a = 8$ with both constant and geometric sequences of inflation factors. The dashed curve connects the “best” points, forming an analog to a Pareto front.

The figure shows that cases with constant inflation coefficients (4×-CONST and 8×-CONST) resulted in lower data-mismatch norms but at the cost of larger changes in the model. The points closer to the region of maximum curvature represent the best balance between data match and

model change. For instance, for $N_a = 4$, a suitable choice for α_1 is on the order of 10^2 , while for $N_a = 8$, the optimal α_1 is around 10^3 . The procedure proposed by Rafiee and Reynolds [356] achieved a particularly well-balanced result for the case with eight data assimilations (8×-GEO1), although the same procedure with four assimilations (4×-GEO1) resulted in a larger data-mismatch norm. These results demonstrate that α_1 and N_a must be selected together. The approach proposed by Emerick [122] (8×-GEO2) resulted in $N_a = 8$ and $\alpha_1 = 3274$, yielding results similar to 8×-GEO1 and showing a good compromise between data match and model change.

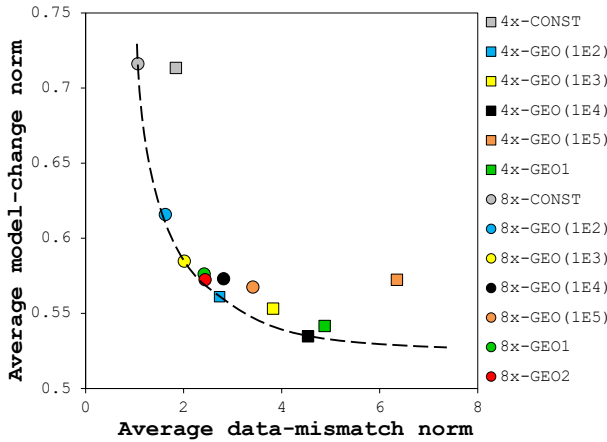


Fig. 5.12: Average data-mismatch and model-change norms. Example 5.3.5.3. Reproduced from Emerick [122] with permission from Elsevier.

5.3.6 Deterministic ES-MDA

ES-MDA relies on the standard perturbed observation scheme, also known as the stochastic scheme. However, there exist deterministic schemes that eliminate the necessity of perturbed observations, such as square-root schemes. One such method is the deterministic ensemble Kalman filter (DEnKF), introduced by Sakov and Oke [374]. While DEnKF is not a square-root filter, it shares the same conceptual objective, aiming to avoid additional sampling errors introduced by perturbing observations. In terms of implementation, DEnKF closely resembles EnKF, involving the computation of the Kalman gain and allowing Schur product localization [197].

Emerick [121] combined ES-MDA with the deterministic scheme proposed by Sakov and Oke [374], giving rise to the deterministic ES-MDA (DES-MDA) method. The motivation behind DES-MDA stems from two key as-

pects: (1) It tends to result in a smaller underestimation of the posterior variance, which is particularly advantageous for problems involving a large number of data points; (2) DES-MDA eliminates the need for factorization to compute $\mathbf{C}_e^{1/2}$, required to obtain perturbed observations. While computing $\mathbf{C}_e^{1/2}$ poses no issue for production data, where it is common to assume uncorrelated measurement errors in time and space (resulting in a diagonal $\mathbf{C}_e$), the scenario is different for 4D seismic data assimilation⁶. In the latter case, a non-diagonal $\mathbf{C}_e$ may be necessary, and the computation of $\mathbf{C}_e^{1/2}$ using Cholesky decomposition becomes computationally prohibitive, especially when N_d is large.

To introduce DES-MDA, it is beneficial to revisit the ES method for the linear case and express it as follows:

$$\mathbf{m}_{c,j} = \mathbf{m}_j + \tilde{\mathbf{K}} (\mathbf{d}_{\text{obs}} + \mathbf{e}_j - \mathbf{G}\mathbf{m}_j), \quad \text{for } j = 1, \dots, N_e, \quad (5.100)$$

where

$$\tilde{\mathbf{K}} = \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{G}^\top (\mathbf{G} \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{G}^\top + \mathbf{C}_e)^{-1}, \quad (5.101)$$

and

$$\mathbf{e}_j = \mathbf{C}_e^{1/2} \mathbf{z}_j \quad (5.102)$$

with $\mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The ES equation to update the mean is⁷

$$\bar{\mathbf{m}}_c = \bar{\mathbf{m}} + \tilde{\mathbf{K}} (\mathbf{d}_{\text{obs}} - \mathbf{G}\bar{\mathbf{m}}). \quad (5.103)$$

Hence,

$$\mathbf{m}_{c,j} - \bar{\mathbf{m}}_c = \mathbf{m}_j - \bar{\mathbf{m}} + \tilde{\mathbf{K}} (\mathbf{e}_j - \mathbf{G}(\mathbf{m}_j - \bar{\mathbf{m}})), \quad \text{for } j = 1, \dots, N_e. \quad (5.104)$$

This can be written in a matrix form as

$$\Delta \mathbf{M}_c = \Delta \mathbf{M} + \tilde{\mathbf{K}} (\mathbf{E} - \mathbf{G} \Delta \mathbf{M}), \quad (5.105)$$

where

$$\mathbf{E} = [\mathbf{e}_1 \cdots \mathbf{e}_{N_e}], \quad (5.106)$$

$$\Delta \mathbf{M} = \frac{1}{\sqrt{N_e - 1}} [\mathbf{m}_1 - \bar{\mathbf{m}}, \dots, \mathbf{m}_{N_e} - \bar{\mathbf{m}}], \quad (5.107)$$

and

$$\Delta \mathbf{M}_c = \frac{1}{\sqrt{N_e - 1}} [\mathbf{m}_{c,1} - \bar{\mathbf{m}}_c, \dots, \mathbf{m}_{c,N_e} - \bar{\mathbf{m}}_c]. \quad (5.108)$$

Computing the covariance matrix of the posterior ensemble gives

⁶ One alternative to compute the data perturbations with non-diagonal $\mathbf{C}_e$ is to use algorithms from geostatistics such as sequential Gaussian simulation.

⁷ Here, it is assumed that $\bar{\mathbf{e}} = \mathbf{0}$. This is easily obtained in practice by correcting the entries of the vector $\mathbf{z}_j$ to have exactly zero mean.

$$\begin{aligned}
 \tilde{\mathbf{C}}_{\mathbf{m}_c} &= \Delta \mathbf{M}_c \Delta \mathbf{M}_c^\top \\
 &= \left[\Delta \mathbf{M} + \tilde{\mathbf{K}} (\mathbf{E} - \mathbf{G} \Delta \mathbf{M}) \right] \left[\Delta \mathbf{M} + \tilde{\mathbf{K}} (\mathbf{E} - \mathbf{G} \Delta \mathbf{M}) \right]^\top \\
 &= \tilde{\mathbf{C}}_{\mathbf{m}} - \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{G}^\top \tilde{\mathbf{K}}^\top - \tilde{\mathbf{K}} \mathbf{G} \tilde{\mathbf{C}}_{\mathbf{m}} + \tilde{\mathbf{K}} \mathbf{G} \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{G}^\top \tilde{\mathbf{K}}^\top + \\
 &\quad \tilde{\mathbf{K}} \mathbf{E} \mathbf{E}^\top \tilde{\mathbf{K}}^\top + \left(\mathbf{I} - \tilde{\mathbf{K}} \mathbf{G} \right) \Delta \mathbf{M} \mathbf{E}^\top \tilde{\mathbf{K}}^\top + \\
 &\quad \tilde{\mathbf{K}} \mathbf{E} \Delta \mathbf{M}^\top \left(\mathbf{I} - \mathbf{G}^\top \tilde{\mathbf{K}}^\top \right). \tag{5.109}
 \end{aligned}$$

This expression converges to the theoretical covariance update formula

$$\tilde{\mathbf{C}}_{\mathbf{m}_c} = \left(\mathbf{I} - \tilde{\mathbf{K}} \mathbf{G} \right) \tilde{\mathbf{C}}_{\mathbf{m}} \tag{5.110}$$

when $N_e \rightarrow \infty$. The fact that the perturbed observation scheme only satisfies the update formula statistically has been regarded as a source of sampling errors⁸.

If we neglect the perturbed observations (setting $\mathbf{E} = \mathbf{0}$ in the previous expression), we obtain

$$\tilde{\mathbf{C}}_{\mathbf{m}_c} = \tilde{\mathbf{C}}_{\mathbf{m}} - 2\tilde{\mathbf{K}} \mathbf{G} \tilde{\mathbf{C}}_{\mathbf{m}} + \tilde{\mathbf{K}} \mathbf{G} \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{G}^\top \tilde{\mathbf{K}}^\top. \tag{5.111}$$

Sakov and Oke [374] observed that by neglecting the quadratic term, $\tilde{\mathbf{K}} \mathbf{G} \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{G}^\top \tilde{\mathbf{K}}^\top$, the theoretical covariance update could be obtained by multiplying the Kalman gain by half. This insight led to the development of DEnKF. In this scheme, the mean is updated using the standard formula, and deviations from the mean are updated using:

$$\Delta \mathbf{M}_c = \Delta \mathbf{M} - \frac{1}{2} \tilde{\mathbf{K}} \mathbf{G} \Delta \mathbf{M}. \tag{5.112}$$

This expression results in the following expression for the posterior covariance

$$\tilde{\mathbf{C}}_{\mathbf{m}_c} = \left(\mathbf{I} - \tilde{\mathbf{K}} \mathbf{G} \right) \tilde{\mathbf{C}}_{\mathbf{m}} + \frac{1}{4} \tilde{\mathbf{K}} \mathbf{G} \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{G}^\top \tilde{\mathbf{K}}^\top. \tag{5.113}$$

Note that compared to the theoretical value, $\tilde{\mathbf{C}}_{\mathbf{m}_c}$ above has an extra positive semidefinite matrix $\frac{1}{4} \tilde{\mathbf{K}} \mathbf{G} \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{G}^\top \tilde{\mathbf{K}}^\top$. Therefore, this expression overestimates the theoretical covariance. Sakov and Oke [374] interpreted this fact as a built-in covariance inflation⁹.

The extension of the deterministic analysis with ES-MDA for nonlinear problems is direct. First, define the matrix with predicted data

$$\Delta \mathbf{D}_\ell = \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{g}(\mathbf{m}_1^\ell) - \overline{\mathbf{g}(\mathbf{m}^\ell)}, \dots, \mathbf{g}(\mathbf{m}_{N_e}^\ell) - \overline{\mathbf{g}(\mathbf{m}^\ell)} \right], \tag{5.114}$$

⁸ That is the motivation for deterministic or square-root schemes, which are designed to exactly satisfy the theoretical covariance update formula.

⁹ Covariance inflation [15] is a common *ad hoc* method used to compensate for excessive variance loss observed in ensemble methods (see Chapter 7, Section 7.4).

and write

$$\mathbf{G}_\ell \Delta \mathbf{M}_\ell = \Delta \mathbf{D}_\ell. \quad (5.115)$$

Then, the following products become

$$\begin{aligned} \tilde{\mathbf{C}}_\mathbf{m}^\ell \mathbf{G}_\ell^\top &= \Delta \mathbf{M}_\ell \Delta \mathbf{M}_\ell^\top \mathbf{G}_\ell^\top \\ &= \Delta \mathbf{M}_\ell \Delta \mathbf{D}_\ell^\top = \tilde{\mathbf{C}}_\mathbf{md}^\ell \end{aligned} \quad (5.116)$$

and

$$\begin{aligned} \mathbf{G}_\ell \tilde{\mathbf{C}}_\mathbf{m}^\ell \mathbf{G}_\ell^\top &= \mathbf{G}_\ell \Delta \mathbf{M}_\ell \Delta \mathbf{M}_\ell^\top \mathbf{G}_\ell^\top \\ &= \Delta \mathbf{D}_\ell \Delta \mathbf{D}_\ell^\top = \tilde{\mathbf{C}}_\mathbf{dd}^\ell. \end{aligned} \quad (5.117)$$

Pseudo-code 5.7 summarizes the resulting DES-MDA method.

Pseudo-code 5.7: DES-MDA

1. **Initialization:** generate the initial ensemble, $\{\mathbf{m}_j^0\}_{j=1}^{N_e}$, choose N_a and $\{\alpha_\ell\}_{\ell=1}^{N_a}$. Set $\ell = 0$.

2. **Forecast step:** compute the vector of predicted data

$$\mathbf{d}_j^\ell = \mathbf{g}(\mathbf{m}_j^\ell), \quad \text{for } j = 1, 2, \dots, N_e \quad (5.118)$$

3. **Analysis step:**

- Compute the Kalman gain

$$\tilde{\mathbf{K}}_\ell = \tilde{\mathbf{C}}_\mathbf{md}^\ell \left(\tilde{\mathbf{C}}_\mathbf{dd}^\ell + \alpha_{\ell+1} \mathbf{C}_\mathbf{e} \right)^{-1} \quad (5.119)$$

- Update the ensemble mean

$$\bar{\mathbf{m}}^{\ell+1} = \bar{\mathbf{m}}^\ell + \tilde{\mathbf{K}}_\ell \left(\mathbf{d}_\text{obs} - \bar{\mathbf{d}}^\ell \right) \quad (5.120)$$

- Update the ensemble deviations

$$\Delta \mathbf{M}_{\ell+1} = \Delta \mathbf{M}_\ell - \frac{1}{2} \tilde{\mathbf{K}}_\ell \Delta \mathbf{D}_\ell \quad (5.121)$$

- Update the ensemble

$$\mathbf{M}_{\ell+1} = \bar{\mathbf{m}}^{\ell+1} \mathbf{1}^\top + \sqrt{N_e - 1} \Delta \mathbf{M}_{\ell+1} \quad (5.122)$$

4. Set $\ell = \ell + 1$.
5. If $\ell = N_a$ then set $\mathbf{m}_{c,j} = \mathbf{m}_j^\ell$ for $j = 1, \dots, N_e$ and stop, else return to step 2.

5.3.6.1 Example: DES-MDA in the UNISIM-I-H Case

Emerick [121] tested ES-MDA and DES-MDA on the UNISIM-I-H case using both production and 4D seismic data. Fig. 5.13 shows the normalized variance of log-permeability after data assimilation with ES-MDA and DES-MDA for an ensemble of $N_e = 200$ realizations. The results indicate that DES-MDA maintained a higher variability in the posterior ensemble without compromising the method's overall ability to match the observed data, as illustrated in Figs. 5.14 and 5.15.

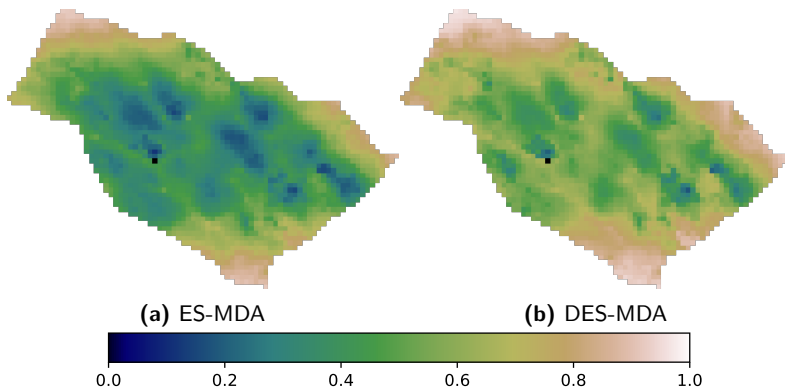


Fig. 5.13: Normalized variance of log-permeability (layer 12) for the UNISIM-I-H case. Example 5.3.6.1. Reproduced from Emerick [121] with permission from Springer Nature.

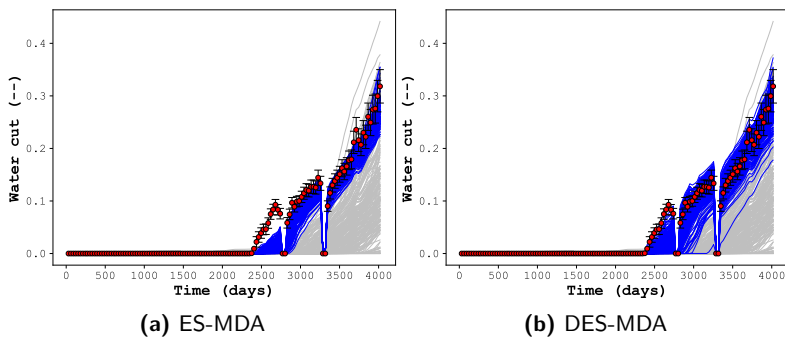


Fig. 5.14: Predicted water cut data for the well NA1A of the UNISIM-I-H case. Red dots are the observed data points, and gray and blue lines correspond to the predicted data from the prior and posterior ensemble, respectively, for the UNISIM-I-H case. Example 5.3.6.1. Reproduced from Emerick [121] with permission from Springer Nature.

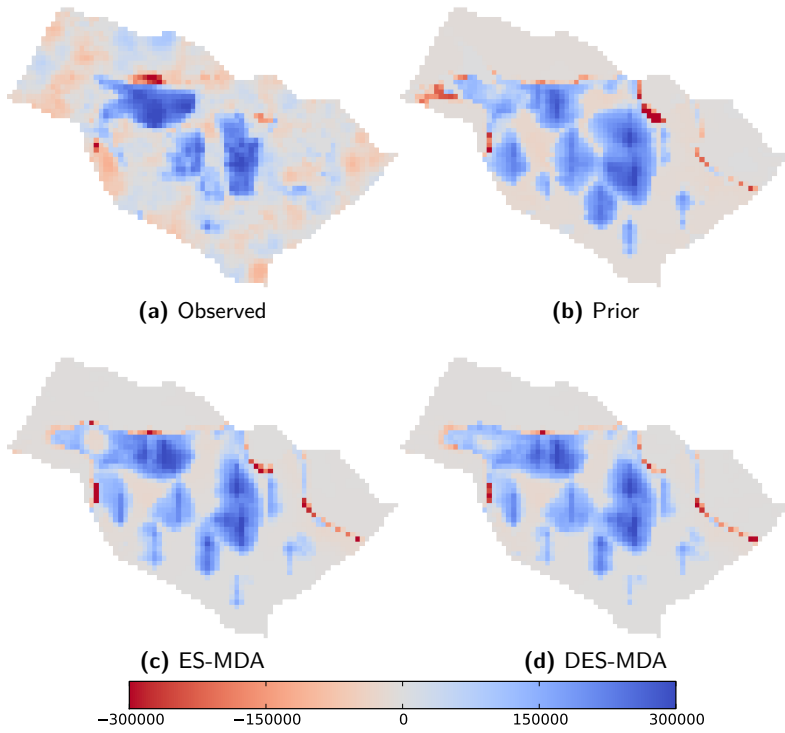


Fig. 5.15: Observed and predicted 4D seismic data (P-impedance changes in $\text{kg/m}^2\text{s}$) for the first realization for the UNISIM-I-H case (layer 12). Example 5.3.6.1. Reproduced from Emerick [121] with permission from Springer Nature.

This page is reserved for your imagination.

Iterative Ensemble Smoothers

Abstract: *This chapter reviews two widely used iterative versions of the ensemble smoother: the ensemble randomized maximum likelihood (EnRML) and the subspace iterative ensemble smoother (SIES).*

6.1 Introduction

This chapter reviews two popular versions of iterative ES: the ensemble RML (EnRML) [78, 79] and the subspace iterative ES (SIES) [354, 142]. Other iterative forms of ES exist in the literature but are not discussed here. Notable examples include the iterative adaptive Gaussian mixture smoother (IAGS) [413], the regularized Levenberg-Marquardt for minimum-average-cost (RLM-MAC) [279], and the iterative local updating ensemble smoother (ILUES) [480].

6.2 Ensemble Randomized Maximum Likelihood

Chen and Oliver [78] introduced one of the earliest iterative forms of ES by adapting the scheme proposed by Gu and Oliver [178] in the context of sequential data assimilation. This method employs a Gauss-Newton (GN) update scheme to minimize the RML objective function:

$$\begin{aligned} \mathcal{O}_r(\mathbf{m}) = & \frac{1}{2} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m})) + \\ & \frac{1}{2} (\mathbf{m} - \mathbf{m}_j)^\top \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_j). \end{aligned} \quad (6.1)$$

In this method, the vector of model parameters is updated using

$$\mathbf{m}_j^{\ell+1} = \mathbf{m}_j^\ell + \beta_\ell \delta \mathbf{m}_j^{\ell+1}, \quad (6.2)$$

where $\beta_\ell \in [0, 1]$ is the step size in the direction $\delta \mathbf{m}_j^{\ell+1}$ given by

$$\begin{aligned} \delta \mathbf{m}_j^{\ell+1} = & - \left(\tilde{\mathbf{C}}_m^{-1} + \tilde{\mathbf{G}}_\ell^\top \mathbf{C}_e^{-1} \tilde{\mathbf{G}}_\ell \right)^{-1} \times \\ & \left[\tilde{\mathbf{C}}_m^{-1} (\mathbf{m}_j^\ell - \mathbf{m}_j) + \tilde{\mathbf{G}}_\ell^\top \mathbf{C}_e^{-1} (\mathbf{g}(\mathbf{m}_j^\ell) - \mathbf{d}_{\text{obs},j}) \right], \end{aligned} \quad (6.3)$$

or equivalently

$$\begin{aligned} \delta \mathbf{m}_j^{\ell+1} = & \mathbf{m}_j - \mathbf{m}_j^\ell - \tilde{\mathbf{C}}_m \tilde{\mathbf{G}}_\ell^\top \left(\tilde{\mathbf{G}}_\ell \tilde{\mathbf{C}}_m \tilde{\mathbf{G}}_\ell^\top + \mathbf{C}_e \right)^{-1} \times \\ & \left[\mathbf{g}(\mathbf{m}_j^\ell) - \mathbf{d}_{\text{obs},j} - \tilde{\mathbf{G}}_\ell (\mathbf{m}_j^\ell - \mathbf{m}_j) \right]. \end{aligned} \quad (6.4)$$

As before, a tilde over a matrix indicates that this matrix is estimated based on the ensemble. Additionally, note that $\mathbf{m}_j$ without the iteration index ℓ refers to a prior realization.

Eq. (6.4) requires the sensitivity matrix at the ℓ th iteration, $\tilde{\mathbf{G}}_\ell$. Gu and Oliver [178] propose to write

$$\tilde{\mathbf{G}}_\ell \Delta \mathbf{M}_\ell = \Delta \mathbf{D}_\ell, \quad (6.5)$$

and compute $\tilde{\mathbf{G}}_\ell$ as

$$\tilde{\mathbf{G}}_\ell = \Delta \mathbf{D}_\ell \Delta \mathbf{M}_\ell^+, \quad (6.6)$$

where $\Delta \mathbf{M}_\ell^+$ is the pseudo-inverse of $\Delta \mathbf{M}_\ell$. As usual, the matrices $\Delta \mathbf{D}_\ell$ and $\Delta \mathbf{M}_\ell$ have the forms

$$\begin{aligned} \Delta \mathbf{D}_\ell = & \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{g}(\mathbf{m}_1^\ell) - \overline{\mathbf{g}(\mathbf{m}^\ell)}, \dots, \mathbf{g}(\mathbf{m}_{N_e}^\ell) - \overline{\mathbf{g}(\mathbf{m}^\ell)} \right] \\ = & \mathbf{D}_\ell \mathbf{A} \end{aligned} \quad (6.7)$$

and

$$\begin{aligned} \Delta \mathbf{M}_\ell = & \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{m}_1^\ell - \overline{\mathbf{m}^\ell}, \dots, \mathbf{m}_{N_e}^\ell - \overline{\mathbf{m}^\ell} \right] \\ = & \mathbf{M}_\ell \mathbf{A} \end{aligned} \quad (6.8)$$

where

$$\mathbf{A} = \frac{1}{\sqrt{N_e - 1}} \left(\mathbf{I} - \frac{1}{N_e} \mathbf{1}\mathbf{1}^\top \right). \quad (6.9)$$

Chen and Oliver [78] demonstrated that this scheme achieved a significant reduction in data mismatch for the Brugge benchmark case [346]. However, this came at the cost of requiring several iterations due to the necessity of small step sizes to ensure acceptable convergence. This observation led Chen and Oliver [79] to replace the GN with a Levenberg-Marquardt (LM) update scheme. The LM scheme was introduced to regularize the search direction and better control the step length, improving the efficiency of the inversion process. The LM analogous of Eq. (6.3) is obtained by modifying the GN Hessian such that

$$\begin{aligned} \delta \mathbf{m}_j^{\ell+1} = & - \left((1 + \lambda_\ell) \tilde{\mathbf{C}}_{\mathbf{m}}^{-1} + \tilde{\mathbf{G}}_\ell^\top \mathbf{C}_{\mathbf{e}}^{-1} \tilde{\mathbf{G}}_\ell \right)^{-1} \times \\ & \left[\tilde{\mathbf{C}}_{\mathbf{m}}^{-1} (\mathbf{m}_j^\ell - \mathbf{m}_j) + \tilde{\mathbf{G}}_\ell^\top \mathbf{C}_{\mathbf{e}}^{-1} (\mathbf{g}(\mathbf{m}_j^\ell) - \mathbf{d}_{\text{obs},j}) \right], \end{aligned} \quad (6.10)$$

or equivalently

$$\begin{aligned} \delta \mathbf{m}_j^{\ell+1} = & \frac{\mathbf{m}_j - \mathbf{m}_j^\ell}{1 + \lambda_\ell} - \tilde{\mathbf{C}}_{\mathbf{m}} \tilde{\mathbf{G}}_\ell^\top \left(\tilde{\mathbf{G}}_\ell \tilde{\mathbf{C}}_{\mathbf{m}} \tilde{\mathbf{G}}_\ell^\top + (1 + \lambda_\ell) \mathbf{C}_{\mathbf{e}} \right)^{-1} \times \\ & \left[\mathbf{g}(\mathbf{m}_j^\ell) - \mathbf{d}_{\text{obs},j} - \frac{\tilde{\mathbf{G}}_\ell (\mathbf{m}_j^\ell - \mathbf{m}_j)}{1 + \lambda_\ell} \right]. \end{aligned} \quad (6.11)$$

One challenge with these schemes is that estimating $\tilde{\mathbf{G}}_\ell$ using the pseudo-inverse of $\Delta \mathbf{M}_\ell$ is noisy and unstable if $N_m \gg N_e$. Chen and Oliver [78] argue that this instability might not be critical because $\tilde{\mathbf{G}}_\ell$ is mostly pre-multiplied by $\tilde{\mathbf{C}}_{\mathbf{m}}$, which tends to smooth the estimates. This effect is illustrated in Fig. 6.1, which presents the values of the matrix $\mathbf{G}_\ell$ ($\ell = 0$) and the product $\mathbf{C}_{\mathbf{m}} \mathbf{G}_\ell^\top$ at two different times for the linear two-phase flow discussed in Section 2.4.5.3. Each panel shows the sensitivity estimated with the ensemble using Eq. (6.6) and calculated with the adjoint method. While the ensemble-based sensitivity values capture the overall structure of the adjoint, they are quite noisy. However, the product with the prior covariance tends to smooth out some of this noise. It is important to note that $\tilde{\mathbf{C}}_{\mathbf{m}}$ corresponds to the prior covariance, which is not updated during the process. However, $\tilde{\mathbf{G}}_\ell$ still appears without the pre-multiplication in the GN and LM update equations, which could lead to instability issues.

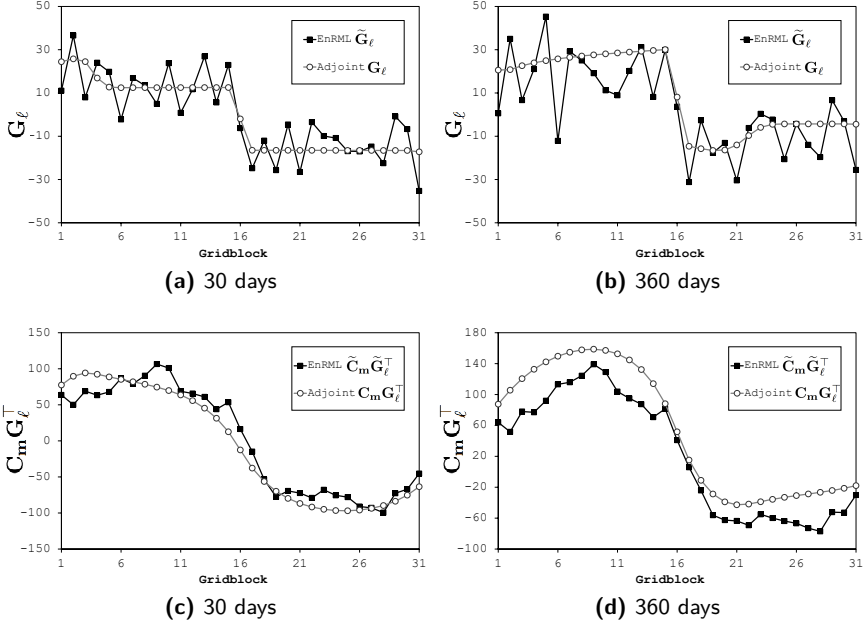


Fig. 6.1: Comparison between adjoint and ensemble-based sensitivity calculations. (a) and (b) display sensitivity values computed at 30 and 360 days, respectively. (c) and (d) show the corresponding products of the prior covariance with the transpose of the sensitivity matrix. Reproduced from Emerick and Reynolds [131] with permission from Springer Nature.

6.2.1 LM-EnRML

The challenges associated with the calculation of $\tilde{\mathbf{G}}_\ell$ led Chen and Oliver [79] to propose an implementation that avoids explicitly calculating $\tilde{\mathbf{G}}_\ell$. To derive this implementation, they modified the Hessian in Eq. (6.10) by replacing $\tilde{\mathbf{C}}_\mathbf{m}$ with $\tilde{\mathbf{C}}_\mathbf{m}^\ell$, which means substituting the prior covariance estimate with the covariance estimate at the current iteration. With this modification, Eq. (6.10) becomes

$$\delta \mathbf{m}_j^{\ell+1} = - \left[(1 + \lambda_\ell) \left(\tilde{\mathbf{C}}_\mathbf{m}^\ell \right)^{-1} + \tilde{\mathbf{G}}_\ell^\top \mathbf{C}_\mathbf{e}^{-1} \tilde{\mathbf{G}}_\ell \right]^{-1} \times \left[\tilde{\mathbf{C}}_\mathbf{m}^{-1} (\mathbf{m}_j^\ell - \mathbf{m}_j) + \tilde{\mathbf{G}}_\ell^\top \mathbf{C}_\mathbf{e}^{-1} (\mathbf{g}(\mathbf{m}_j^\ell) - \mathbf{d}_{\text{obs},j}) \right], \quad (6.12)$$

or equivalently

$$\begin{aligned} \delta \mathbf{m}_j^{\ell+1} = & - \left[(1 + \lambda_\ell) \left(\tilde{\mathbf{C}}_\mathbf{m}^\ell \right)^{-1} + \tilde{\mathbf{G}}_\ell^\top \mathbf{C}_\mathbf{e}^{-1} \tilde{\mathbf{G}}_\ell \right]^{-1} \tilde{\mathbf{C}}_\mathbf{m}^{-1} (\mathbf{m}_j^\ell - \mathbf{m}_j) - \\ & \tilde{\mathbf{C}}_\mathbf{m}^\ell \tilde{\mathbf{G}}_\ell^\top \left[\tilde{\mathbf{G}}_\ell \tilde{\mathbf{C}}_\mathbf{m}^\ell \tilde{\mathbf{G}}_\ell^\top + (1 + \lambda_\ell) \mathbf{C}_\mathbf{e} \right]^{-1} (\mathbf{g}(\mathbf{m}_j^\ell) - \mathbf{d}_{\text{obs},j}). \end{aligned} \quad (6.13)$$

For the computational implementation of Eq. (6.13), Chen and Oliver [79] introduced the following scaled matrices:

$$\widehat{\Delta \mathbf{M}}_\ell = \mathbf{C}_s^{-1/2} \Delta \mathbf{M}_\ell, \quad (6.14)$$

and

$$\widehat{\Delta \mathbf{D}}_\ell = \mathbf{C}_\mathbf{e}^{-1/2} \Delta \mathbf{D}_\ell, \quad (6.15)$$

where $\mathbf{C}_s$ is a scaling diagonal matrix containing the prior model variances, i.e., $\mathbf{C}_s = \text{diag}(\mathbf{C}_\mathbf{m})$.

Using (6.14)–(6.15) in (6.13) leads to

$$\begin{aligned} \delta \mathbf{m}_j^{\ell+1} = & -\mathbf{C}_s^{1/2} \widehat{\Delta \mathbf{M}}_\ell \left[(1 + \lambda_\ell) \mathbf{I} + \widehat{\Delta \mathbf{D}}_\ell^\top \widehat{\Delta \mathbf{D}}_\ell \right]^{-1} \times \\ & \widehat{\Delta \mathbf{M}}_\ell^\top \left(\widehat{\Delta \mathbf{M}}_\ell^{-\top} \widehat{\Delta \mathbf{M}}_\ell^{-1} \right) \mathbf{C}_s^{-1/2} (\mathbf{m}_j^\ell - \mathbf{m}_j) - \\ & \mathbf{C}_s^{1/2} \widehat{\Delta \mathbf{M}}_\ell \widehat{\Delta \mathbf{D}}_\ell^\top \left[(1 + \lambda_\ell) \mathbf{I} + \widehat{\Delta \mathbf{D}}_\ell \widehat{\Delta \mathbf{D}}_\ell^\top \right]^{-1} \times \\ & \mathbf{C}_\mathbf{e}^{-1/2} (\mathbf{g}(\mathbf{m}_j^\ell) - \mathbf{d}_{\text{obs},j}). \end{aligned} \quad (6.16)$$

Note that $\widehat{\Delta \mathbf{M}}$ without the iteration index ℓ refers to the prior. Using truncated singular value decomposition

$$\widehat{\Delta \mathbf{M}} = \mathbf{U}_m \Sigma_m \mathbf{V}_m^\top \quad (6.17)$$

and

$$\widehat{\Delta \mathbf{D}}_\ell = \mathbf{U}_d \Sigma_d \mathbf{V}_d^\top, \quad (6.18)$$

in (6.16) leads to the final LM-EnRML update direction:

$$\begin{aligned} \delta \mathbf{m}_j^{\ell+1} = & -\mathbf{C}_s^{1/2} \widehat{\Delta \mathbf{M}}_\ell \mathbf{V}_d [(1 + \lambda_\ell) \mathbf{I} + \Sigma_d^2]^{-1} \times \\ & \mathbf{V}_d^\top \widehat{\Delta \mathbf{M}}_\ell^\top \mathbf{U}_m \Sigma_m^{-2} \mathbf{U}_m^\top \mathbf{C}_s^{-1/2} (\mathbf{m}_j^\ell - \mathbf{m}_j) - \\ & \mathbf{C}_s^{1/2} \widehat{\Delta \mathbf{M}}_\ell \mathbf{V}_d \Sigma_d [(1 + \lambda_\ell) \mathbf{I} + \Sigma_d^2]^{-1} \times \\ & \mathbf{U}_d^\top \mathbf{C}_\mathbf{e}^{-1/2} (\mathbf{g}(\mathbf{m}_j^\ell) - \mathbf{d}_{\text{obs},j}). \end{aligned} \quad (6.19)$$

Chen and Oliver [79] also proposed an approximation to this expression by neglecting the term corresponding to the model mismatch in Eq. (6.16), leading to an update direction given by

$$\delta \mathbf{m}_j^{\ell+1} = -\mathbf{C}_s^{1/2} \widehat{\Delta \mathbf{M}}_\ell \mathbf{V}_d \boldsymbol{\Sigma}_d \left[(1 + \lambda_\ell) \mathbf{I} + \boldsymbol{\Sigma}_d^2 \right]^{-1} \times \mathbf{U}_d^\top \mathbf{C}_e^{-1/2} \left(\mathbf{g}(\mathbf{m}_j^\ell) - \mathbf{d}_{\text{obs},j} \right). \quad (6.20)$$

It is interesting to note that the update equation of the approximated version of LM-EnRML is very similar to the ES-MDA update equation, with $(1 + \lambda_\ell)$ taking the role of the MDA data-error covariance inflation coefficient, α_ℓ . However, there are two noticeable differences: First, ES-MDA terminates based on the condition $\sum_{\ell=1}^{N_a} \alpha_\ell^{-1} = 1$, while LM-EnRML stops based on convergence criteria or the maximum number of iterations. Second, ES-MDA computes the data perturbations using the inflated data-error covariance, $\alpha_\ell \mathbf{C}_e$, and resamples each iteration, while LM-EnRML computes the perturbations based on $\mathbf{C}_e$ and keeps the values constant throughout the entire iterative process.

Pseudo-code 6.1 summarizes the LM-EnRML method. This process requires the specification of the initial LM parameter, $\lambda_0 > 0$, the maximum LM parameter, $\lambda_{\text{max}} > 0$, and the increase factor, $\gamma > 1$. Chen and Oliver [79] suggest selecting λ_0 on the order of $\overline{\mathcal{O}}_{N,d}^0$ and $\gamma = 10$. The maximum λ_{max} is introduced in Pseudo-code 6.1 to halt the process when the data mismatch ceases to decrease. Additional termination criteria include the maximum number of iterations, a small reduction in the objective function, and minimal updates in the models over an iteration.

Pseudo-code 6.1: LM-EnRML

1. **Initialization:** generate the initial ensemble, $\{\mathbf{m}_j^0\}_{j=1}^{N_e}$, by sampling the prior distribution of parameters, select the values for λ_0 , $\lambda_{\max}$, and γ , set $\ell = 0$, and compute:

$$\widehat{\Delta \mathbf{M}} = \mathbf{C}_s^{-1/2} \Delta \mathbf{M} = \mathbf{U}_m \Sigma_m \mathbf{V}_m^\top \quad (6.21)$$

$$\mathbf{d}_{\text{obs},j} = \mathbf{d}_{\text{obs}} + \mathbf{C}_e^{1/2} \mathbf{z}_j, \text{ where } \mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (6.22)$$

$$\mathbf{d}_j^0 = \mathbf{g}(\mathbf{m}_j^0), \text{ for } j = 1, 2, \dots, N_e \quad (6.23)$$

$$\overline{\mathcal{O}}_{N,d}^0 = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathcal{O}_{N,d}(\mathbf{m}_j^0) \quad (6.24)$$

2. **Model update:**

$$\mathbf{m}_j^{\ell+1} = \mathbf{m}_j^\ell + \delta \mathbf{m}_j^\ell, \text{ for } j = 1, 2, \dots, N_e, \quad (6.25)$$

where

$$\begin{aligned} \delta \mathbf{m}_j^{\ell+1} = & -\mathbf{C}_s^{1/2} \widehat{\Delta \mathbf{M}}_\ell \mathbf{V}_d [(1 + \lambda_\ell) \mathbf{I} + \Sigma_d^2]^{-1} \times \\ & \mathbf{V}_d^\top \widehat{\Delta \mathbf{M}}_\ell^\top \mathbf{U}_m \Sigma_m^{-2} \mathbf{U}_m^\top \mathbf{C}_s^{-1/2} (\mathbf{m}_j^\ell - \mathbf{m}_j) - \\ & \mathbf{C}_s^{1/2} \widehat{\Delta \mathbf{M}}_\ell \mathbf{V}_d \Sigma_d [(1 + \lambda_\ell) \mathbf{I} + \Sigma_d^2]^{-1} \times \\ & \mathbf{U}_d^\top \mathbf{C}_e^{-1/2} (\mathbf{d}_j^\ell - \mathbf{d}_{\text{obs},j}), \end{aligned} \quad (6.26)$$

$$\widehat{\Delta \mathbf{M}}_\ell = \mathbf{C}_s^{-1/2} \Delta \mathbf{M}_\ell, \quad (6.27)$$

and

$$\widehat{\Delta \mathbf{D}}_\ell = \mathbf{C}_e^{-1/2} \Delta \mathbf{D}_\ell = \mathbf{U}_d \Sigma_d \mathbf{V}_d^\top. \quad (6.28)$$

3. **Model prediction:**

$$\mathbf{d}_j^{\ell+1} = \mathbf{g}(\mathbf{m}_j^{\ell+1}), \text{ for } j = 1, 2, \dots, N_e \quad (6.29)$$

$$\overline{\mathcal{O}}_{N,d}^{\ell+1} = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathcal{O}_{N,d}(\mathbf{m}_j^{\ell+1}) \quad (6.30)$$

4. If $\overline{\mathcal{O}}_{N,d}^{\ell+1} > \overline{\mathcal{O}}_{N,d}^\ell$ then increase $\lambda_\ell = \gamma \lambda_\ell$. If $\lambda_\ell > \lambda_{\max}$ stop, else return to step 3.
5. If $\overline{\mathcal{O}}_{N,d}^{\ell+1} \leq \overline{\mathcal{O}}_{N,d}^\ell$ check the termination criteria.
6. If not terminated then reduce $\lambda_{\ell+1} = \lambda_\ell / \gamma$, set $\ell = \ell + 1$ and return to step 2.

6.3 Subspace Iterative Ensemble Smoother

Raanes et al. [354] revisited the EnRML method and proposed a new derivation based on the observation that the solution provided by EnRML can be written as a linear combination of the prior realizations. The resulting method, called the subspace iterative ensemble smoother (SIES), is argued to be simpler both in terms of derivation and implementation than EnRML. Since both methods start from the same premise, they converge to the same solution, except for roundoff errors and specific implementation details such as step size and singular value truncation. Evensen et al. [145] provided a detailed discussion on the implementation of SIES.

The starting point for SIES is to write the vector of model parameters at the ℓ th iteration as a linear combination of the prior realizations as

$$\mathbf{m}_j^\ell = \mathbf{m}_j + \Delta \mathbf{M} \mathbf{w}_j^\ell, \quad (6.31)$$

where $\mathbf{w}_j^\ell \in \mathbb{R}^{N_e}$ is the vector containing the coefficients of the linear combination. Note that $\mathbf{m}_j$ is the j th prior realization and $\Delta \mathbf{M}$ contains the prior ensemble variations around the prior mean. The RML objective function can be written in terms of $\mathbf{w}$ as

$$\begin{aligned} \mathcal{O}_r(\mathbf{w}) = & \frac{1}{2} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}_j + \Delta \mathbf{M} \mathbf{w}))^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}_j + \Delta \mathbf{M} \mathbf{w})) + \\ & \frac{1}{2} \mathbf{w}^\top \mathbf{w}, \end{aligned} \quad (6.32)$$

with $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Raanes et al. [354] propose to minimize (6.32) using a GN scheme such that

$$\mathbf{w}_j^{\ell+1} = \mathbf{w}_j^\ell + \beta_\ell \delta \mathbf{w}_j^{\ell+1} \quad (6.33)$$

where

$$\begin{aligned} \mathbf{w}_j^{\ell+1} = & -\mathbf{w}_j^\ell - \left(\tilde{\mathbf{G}}_\ell \Delta \mathbf{M} \right)^\top \left[\left(\tilde{\mathbf{G}}_\ell \Delta \mathbf{M} \right) \left(\tilde{\mathbf{G}}_\ell \Delta \mathbf{M} \right)^\top + \mathbf{C}_e \right]^{-1} \times \\ & \left(\mathbf{g}(\mathbf{m}_j + \Delta \mathbf{M} \mathbf{w}_j^\ell) - \mathbf{d}_{\text{obs},j} - \tilde{\mathbf{G}}_\ell \Delta \mathbf{M} \mathbf{w}_j^\ell \right). \end{aligned} \quad (6.34)$$

Eq. (6.34) requires to compute the sensitivity matrix $\tilde{\mathbf{G}}_\ell$, which can be estimated using the pseudo-inverse of $\Delta \mathbf{M}_\ell$. Note that $\tilde{\mathbf{G}}_\ell$ always appears multiplied by $\Delta \mathbf{M}$ in Eq. (6.34), hence

$$\tilde{\mathbf{G}}_\ell \Delta \mathbf{M} = \Delta \mathbf{D}_\ell \Delta \mathbf{M}_\ell^+ \Delta \mathbf{M}. \quad (6.35)$$

Instead of computing $\Delta \mathbf{M}_\ell^+$, Raanes et al. [354] noted that

$$\begin{aligned}
 \Delta \mathbf{M}_\ell &= \mathbf{M}_\ell \mathbf{A} \\
 &= (\mathbf{M} + \Delta \mathbf{M} \mathbf{W}_\ell) \mathbf{A} \\
 &= \Delta \mathbf{M} (\mathbf{I} + \mathbf{W}_\ell \mathbf{A}),
 \end{aligned} \tag{6.36}$$

where $\mathbf{M}$ and $\mathbf{W}$ are the matrices containing the realizations of $\mathbf{m}$ and $\mathbf{w}$ in their columns. The matrix $(\mathbf{I} + \mathbf{W}_\ell \mathbf{A})$ is positive definite and we can write

$$\Delta \mathbf{M} = \Delta \mathbf{M}_\ell (\mathbf{I} + \mathbf{W}_\ell \mathbf{A})^{-1} \tag{6.37}$$

and

$$\tilde{\mathbf{G}}_\ell \Delta \mathbf{M} = \Delta \mathbf{D}_\ell \Delta \mathbf{M}_\ell^+ \Delta \mathbf{M}_\ell (\mathbf{I} + \mathbf{W}_\ell \mathbf{A})^{-1}. \tag{6.38}$$

Raanes et al. [354] noted that if the forward model is linear, $\mathbf{d} = \mathbf{G}\mathbf{m}$, Eq. (6.38) simplifies to

$$\begin{aligned}
 \tilde{\mathbf{G}}_\ell \Delta \mathbf{M} &= \Delta \mathbf{D}_\ell \Delta \mathbf{M}_\ell^+ \Delta \mathbf{M}_\ell (\mathbf{I} + \mathbf{W}_\ell \mathbf{A})^{-1} \\
 &= \mathbf{G} \Delta \mathbf{M}_\ell \Delta \mathbf{M}_\ell^+ \Delta \mathbf{M}_\ell (\mathbf{I} + \mathbf{W}_\ell \mathbf{A})^{-1} \\
 &= \mathbf{G} \Delta \mathbf{M}_\ell (\mathbf{I} + \mathbf{W}_\ell \mathbf{A})^{-1} \\
 &= \Delta \mathbf{D}_\ell (\mathbf{I} + \mathbf{W}_\ell \mathbf{A})^{-1}.
 \end{aligned} \tag{6.39}$$

If the forward model is nonlinear, Raanes et al. [354] discuss two situations: First, if $N_m \geq N_e - 1$, then $\text{rank}(\Delta \mathbf{M}_\ell) = N_e - 1$ and

$$\Delta \mathbf{M}_\ell^+ \Delta \mathbf{M}_\ell = \mathbf{I} - \frac{1}{N_e} \mathbf{1} \mathbf{1}^\top. \tag{6.40}$$

In this case, Eq. (6.38) reduces to

$$\tilde{\mathbf{G}}_\ell \Delta \mathbf{M} = \Delta \mathbf{D}_\ell (\mathbf{I} + \mathbf{W}_\ell \mathbf{A})^{-1}, \tag{6.41}$$

which is the same result for the linear case. However, if the forward model is nonlinear and $N_m < N_e - 1$, we do not have the simplification obtained in Eq. (6.40). In this case, Raanes et al. [354] propose to use the SVD of $\Delta \mathbf{M}_\ell$ and write

$$\Delta \mathbf{M}_\ell^+ \Delta \mathbf{M}_\ell = \mathbf{V}_m \Sigma_m^+ \mathbf{U}_m^\top \mathbf{U}_m \Sigma_m \mathbf{V}_m^\top = \mathbf{V}_m \mathbf{V}_m^\top. \tag{6.42}$$

Note that we only need the first N_m right singular vectors of $\Delta \mathbf{M}_\ell$. In terms of computational implementation, Evensen et al. [145] note that instead of computing $(\mathbf{I} + \mathbf{W}_\ell \mathbf{A})^{-1}$, it is more efficient to solve the linear problem

$$(\mathbf{I} + \mathbf{W}_\ell \mathbf{A}) \tilde{\mathbf{G}}_\ell \Delta \mathbf{M} = \Delta \mathbf{D}_\ell, \tag{6.43}$$

for $\tilde{\mathbf{G}}_\ell \Delta \mathbf{M}$.

Pseudo-code 6.2 summarizes the SIES method. This pseudo-code requires the specification of an initial step size, $\beta_0 \leq 1$, a minimum step size $\beta_{\min} > 0$, and a reduction factor $\gamma < 1$. For linear problems, the method converges in one iteration with $\beta_0 = 1$. For nonlinear problems, however, it may be beneficial to use smaller step sizes. Evensen [142] suggest $\beta_0 = 0.6$, but this may vary depending on the problem. Pseudo-code 6.2 includes a heuristic to decrease the step size whenever an iteration fails to reduce the average data-mismatch objective function. A reasonable choice for the reduction factor is $\gamma = 0.5$. A minimal value for the step size, such as $\beta_{\min} = 0.01$, is introduced to avoid unnecessary executions of the forward model in case of a non-decreasing search direction. The iterative process stops when $\beta_\ell < \beta_{\min}$, the maximum number of iterations is reached, or when a small reduction in the objective function or a small update in the models over an iteration is observed.

The SIES method operates in the ensemble subspace, meaning that the data assimilation parameters are the coefficients in the vectors $\mathbf{w}$. This approach enables an efficient implementation that scales linearly with the number of data points, N_d , and model parameters, N_m . However, it complicates the application of standard Schur-product localization (see Chapter 7, Section 7.5). As a solution, a domain localization scheme can be employed (Chapter 7, Section 7.5.5).

Pseudo-code 6.2: SIES

1. **Initialization:** generate the initial ensemble, $\{\mathbf{m}_j^0\}_{j=1}^{N_e}$, by sampling the prior distribution of parameters, select the values for β_0 , $\beta_{\min}$, and γ , set $\ell = 0$ and $\mathbf{W}_0 = \mathbf{O}$, and compute:

$$\mathbf{d}_{\text{obs},j} = \mathbf{d}_{\text{obs}} + \mathbf{C}_e^{1/2} \mathbf{z}_j, \text{ where } \mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (6.44)$$

$$\mathbf{d}_j^0 = \mathbf{g}(\mathbf{m}_j + \Delta \mathbf{M} \mathbf{w}_j^0), \text{ for } j = 1, 2, \dots, N_e \quad (6.45)$$

$$\bar{\mathcal{O}}_{N,d}^0 = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathcal{O}_{N,d}(\mathbf{w}_j^0) \quad (6.46)$$

2. Compute $\tilde{\mathbf{G}}_\ell \Delta \mathbf{M}$ by solving one of the following cases:

- If $N_m < N_e - 1$:

$$(\mathbf{I} + \mathbf{W}_\ell \mathbf{A}) \tilde{\mathbf{G}}_\ell \Delta \mathbf{M} = \Delta \mathbf{D}_\ell \mathbf{V}_m \mathbf{V}_m^\top \quad (6.47)$$

where

$$\Delta \mathbf{M}_\ell = \mathbf{U}_m \Sigma_m \mathbf{V}_m^\top \quad (6.48)$$

- If $N_m \geq N_e - 1$:

$$(\mathbf{I} + \mathbf{W}_\ell \mathbf{A}) \tilde{\mathbf{G}}_\ell \Delta \mathbf{M} = \Delta \mathbf{D}_\ell \quad (6.49)$$

3. **Model update:**

$$\mathbf{w}_j^{\ell+1} = \mathbf{w}_j^\ell + \beta_\ell \delta \mathbf{w}_j^{\ell+1} \quad (6.50)$$

where

$$\begin{aligned} \mathbf{w}_j^{\ell+1} = & -\mathbf{w}_j^\ell - \left(\tilde{\mathbf{G}}_\ell \Delta \mathbf{M} \right)^\top \left[\left(\tilde{\mathbf{G}}_\ell \Delta \mathbf{M} \right) \left(\tilde{\mathbf{G}}_\ell \Delta \mathbf{M} \right)^\top + \mathbf{C}_e \right]^{-1} \times \\ & \left(\mathbf{g}(\mathbf{m}_j + \Delta \mathbf{M} \mathbf{w}_j^\ell) - \mathbf{d}_{\text{obs},j} - \tilde{\mathbf{G}}_\ell \Delta \mathbf{M} \mathbf{w}_j^\ell \right). \end{aligned} \quad (6.51)$$

4. **Model prediction:**

$$\mathbf{d}_j^{\ell+1} = \mathbf{g}(\mathbf{m}_j + \Delta \mathbf{M} \mathbf{w}_j^{\ell+1}), \text{ for } j = 1, 2, \dots, N_e \quad (6.52)$$

$$\bar{\mathcal{O}}_{N,d}^{\ell+1} = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathcal{O}_{N,d}(\mathbf{w}_j^{\ell+1}) \quad (6.53)$$

5. If $\bar{\mathcal{O}}_{N,d}^{\ell+1} > \bar{\mathcal{O}}_{N,d}^\ell$ then reduce $\beta_\ell = \gamma \beta_\ell$. If $\beta_\ell < \beta_{\min}$ stop, else return to step 3.
6. If $\bar{\mathcal{O}}_{N,d}^{\ell+1} \leq \bar{\mathcal{O}}_{N,d}^\ell$ check the termination criteria.
7. If not terminated, increase $\beta_{\ell+1} = \min \{1, \beta_\ell / \gamma\}$, set $\ell = \ell + 1$, and return to step 2.

This page is proof that less is more.

Sampling Errors and Rank Deficiency

Abstract: *This chapter addresses the challenges posed by limited-sized ensembles, including sampling errors and restricted degrees of freedom. It also examines strategies to mitigate their adverse effects, with a particular emphasis on the method known as localization.*

7.1 Introduction

The use of ensembles to estimate covariances is one of the primary reasons for the widespread adoption of ensemble methods across various fields. Ensembles allow handling nonlinear problems while avoiding complex derivative calculations, enabling seamless integration with sophisticated forward models at a reasonable computational cost. Furthermore, since these methods can be formulated within a Bayesian framework, posterior ensembles can be used to estimate the uncertainty in future performance predictions.

Unfortunately, the use of ensembles is also the “Achilles heel” of these methods. The computational demands of large-scale problems require limiting the ensemble size to a few hundred components, which degrades data assimilation performance, leading to two primary issues:

- **Rank deficiency:** This limits the number of degrees of freedom available to assimilate data.
- **Spurious correlations:** These lead to poor estimates of covariances, resulting in erroneous model updates.

The repercussions of these problems are often observed as incorrect updates in the models and excessive reduction in the ensemble variance [3]. In extreme cases, they can lead to the collapse of the ensemble, rendering the methods incapable of assimilating further data.

This chapter discusses the challenges posed by using limited-sized ensembles and explores strategies to mitigate their negative effects. In particular, it focuses on the method known as localization.

Remark 7.1: Uncertainty Underestimation

Ensemble methods typically perform well in generating conditional realizations with satisfactory data matches. However, numerous studies have consistently reported excessive reductions in ensemble variability and overly narrow prediction ranges, limiting their effectiveness in uncertainty quantification [77, 130, 118, 119, 245, 277].

Uncertainty underestimation in ensemble methods is often attributed to sampling errors resulting from the limited ensemble size. However, other factors may contribute, including poorly characterized prior uncertainties, insufficient parametrization, underestimation of data uncertainties, and overconfidence in the model's ability to accurately simulate the system behavior.

7.2 Rank Deficiency

One of the most severe consequences of the ensemble sizes used in practice is the limitation of degrees of freedom available for assimilating data. To understand this, it is convenient to review the ES analysis equation:

$$\mathbf{m}_{c,j} = \mathbf{m}_j + \tilde{\mathbf{C}}_{\text{md}} \left(\tilde{\mathbf{C}}_{\text{dd}} + \mathbf{C}_e \right)^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{d}_j), \quad (7.1)$$

for $j = 1, 2, \dots, N_e$, where $\mathbf{d}_j = \mathbf{g}(\mathbf{m}_j)$. Note that after computing the $N_d \times N_d$ matrix $(\tilde{\mathbf{C}}_{\text{dd}} + \mathbf{C}_e)$, we can solve the linear problem

$$\equiv \left(\tilde{\mathbf{C}}_{\text{dd}} + \mathbf{C}_e \right) \mathbf{x}_j = (\mathbf{d}_{\text{obs},j} - \mathbf{d}_j) \quad (7.2)$$

for the vector $\mathbf{x}_j$. Then, the analysis equation can be written as

$$\begin{aligned} \mathbf{m}_{c,j} &= \mathbf{m}_j + \tilde{\mathbf{C}}_{\text{md}} \mathbf{x}_j \\ &= \mathbf{m}_j + \frac{1}{N_e - 1} \sum_{k=1}^{N_e} (\mathbf{m}_k - \bar{\mathbf{m}}) \overbrace{(\mathbf{d}_k - \bar{\mathbf{d}})^\top}^{\theta_{jk}} \mathbf{x}_j \\ &= \mathbf{m}_j + \frac{1}{N_e - 1} \sum_{k=1}^{N_e} \theta_{jk} (\mathbf{m}_k - \bar{\mathbf{m}}). \end{aligned} \quad (7.3)$$

The vector $\bar{\mathbf{m}}$ is itself represented by a linear combination of the set of prior state vectors, $\mathbf{m}_j$. Thus, it follows that we can write $\mathbf{m}_{c,j}$ as

$$\mathbf{m}_{c,j} = \sum_{k=1}^{N_e} \gamma_{jk} \mathbf{m}_k \quad \text{for } j = 1, 2, \dots, N_e, \quad (7.4)$$

where the γ_{jk} 's are real numbers. This result shows that the j th analyzed state, $\mathbf{m}_{c,j}$, in the ES is a linear combination of the set of N_e prior vectors, $\mathbf{m}_k$. This means that we effectively have at most N_e coefficients, $\{\gamma_j\}_{k=1}^{N_e}$, that can be adjusted to match the observed data. In other words, $\{\mathbf{m}_k\}_{k=1}^{N_e}$ defines a subspace of $\mathbb{R}^{N_m}$, with dimension $\mathbb{R}^{N_e}$, in which we can generate estimates. The same conclusion holds for EnKF, ES-MDA, and other ensemble methods.

Another way to look at this issue is to investigate the rank of the ensemble-based covariance matrix.

$$\begin{aligned} \mathbf{C}_{\mathbf{md}} &\approx \tilde{\mathbf{C}}_{\mathbf{md}} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{m}_j - \bar{\mathbf{m}})(\mathbf{d}_j - \bar{\mathbf{d}})^\top \\ &= \Delta \mathbf{M} \Delta \mathbf{D}^\top. \end{aligned} \quad (7.5)$$

Note that $\Delta \mathbf{M}$ is a $N_m \times N_e$ matrix computed by subtracting the ensemble mean from each of its columns. Thus, $\text{rank}(\Delta \mathbf{M}) \leq \min\{N_m, N_e - 1\}$. Typically, $N_e \ll N_m$, meaning that $\text{rank}(\Delta \mathbf{M}) \leq N_e - 1$. A similar conclusion holds for $\Delta \mathbf{D}$. Hence, $\text{rank}(\tilde{\mathbf{C}}_{\mathbf{md}}) \leq N_e - 1$.

Moreover, Lorenc [270] showed that the assimilation of a perfect datum results in a loss of rank, effectively removing one degree of freedom. While we never assimilate perfect data, it is reasonable to expect that, after assimilating several data points, there will be a reduction in the available degrees of freedom.

7.2.1 Example: Porosity and Permeability of a Rock Sample

The loss of rank is straightforward to illustrate. For instance, consider the problem of estimating porosity and permeability in a core sample introduced in Section 2.3.1.3. We showed that the posterior covariance after assimilation of one observation of porosity is given by

$$\begin{aligned} \mathbf{C}_{\mathbf{m}_c} &= \mathbf{C}_{\mathbf{m}} - \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top (\mathbf{C}_e + \mathbf{G} \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top)^{-1} \mathbf{G} \mathbf{C}_{\mathbf{m}} \\ &= \begin{bmatrix} \sigma_\phi^2 \left(1 - \frac{\sigma_\phi^2}{\sigma_e^2 + \sigma_\phi^2}\right) & \rho \sigma_\phi \sigma_{\ln \kappa} \left(1 - \frac{\sigma_\phi^2}{\sigma_e^2 + \sigma_\phi^2}\right) \\ \rho \sigma_\phi \sigma_{\ln \kappa} \left(1 - \frac{\sigma_\phi^2}{\sigma_e^2 + \sigma_\phi^2}\right) & \sigma_{\ln \kappa}^2 \left(1 - \frac{\rho^2 \sigma_\phi^2}{\sigma_e^2 + \sigma_\phi^2}\right) \end{bmatrix}. \end{aligned} \quad (7.6)$$

Now, if we have a perfect measurement of porosity, i.e., $\sigma_e = 0$, then $\mathbf{C}_{\mathbf{m}_c}$ becomes

$$\mathbf{C}_{\mathbf{m}_c} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{\ln \kappa}^2 (1 - \rho^2) \end{bmatrix}, \quad (7.7)$$

which is singular. Therefore, assimilation of a perfect datum reduced the rank of the posterior covariance, now $\text{rank}(\mathbf{C}_{\mathbf{m}_c}) = 1$.

7.3 Spurious Correlations

The second adverse effect of the limited ensemble size is its potential to yield inaccurate estimates of the covariances used in the computation of the Kalman gain. In this scenario, the estimated covariance matrices are prone to spurious correlations¹, leading to erroneous updates in the ensemble members. Consequently, we observe changes in model parameters which have no influence on the assimilated data, exacerbating problems of excessive loss of ensemble variance².

7.3.1 Example: Porosity and Permeability of a Rock Sample

To illustrate the impact of sampling errors, let's revisit the problem of estimating porosity and permeability in a core sample. However, let's assume that there are no prior correlations between porosity and permeability. In this case, the prior covariance takes the form:

¹ There is well-known result in random matrix theory [25] that states if $\mathbf{x}_j$'s are independent and identically distributed samples from $\mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{x}})$, the error in covariance estimation with an ensemble of N_e samples is

$$\|\mathbf{C}_{\mathbf{x}} - \tilde{\mathbf{C}}_{\mathbf{x}}\| = O\left(\sqrt{\frac{N_x}{N_e}}\right).$$

This means that the estimated covariance $\tilde{\mathbf{C}}_{\mathbf{x}}$ is very inaccurate in high dimensions for small ensembles, i.e., $N_e \ll N_x$.

² Model nonlinearity also exacerbates the effect of sampling errors, as illustrated by Raanes et al. [353] in a compelling example involving a univariate problem with a Gaussian prior and two models—one linear and the other nonlinear. Despite the strong nonlinearity, the nonlinear model was designed to produce the exact same posterior Gaussian PDF as the linear case. Data were assimilated using a square root filter with an ensemble of 40 members. For the linear case, sampling errors were quickly attenuated, and the ensemble statistics converged to the exact values. However, in the nonlinear case, the sampling errors persisted, demonstrating a chronic issue.

$$\mathbf{C}_m = \begin{bmatrix} \sigma_\phi^2 & 0 \\ 0 & \sigma_{\ln \kappa}^2 \end{bmatrix}. \quad (7.8)$$

If we assimilate a single measurement of porosity, $\mathbf{d}_{\text{obs}} = \phi_{\text{obs}}$ and $\mathbf{C}_e = \sigma_e^2$, it results in the following expression for the posterior covariance

$$\begin{aligned} \mathbf{C}_{m_c} &= \mathbf{C}_m - \mathbf{C}_m \mathbf{G}^\top (\mathbf{C}_e + \mathbf{G} \mathbf{C}_m \mathbf{G}^\top)^{-1} \mathbf{G} \mathbf{C}_m \\ &= \begin{bmatrix} \sigma_\phi^2 - \frac{\sigma_\phi^4}{\sigma_e^2 + \sigma_\phi^2} & 0 \\ 0 & \sigma_{\ln \kappa}^2 \end{bmatrix}. \end{aligned} \quad (7.9)$$

Because there is no prior correlation between ϕ and $\ln \kappa$, one observation of ϕ does not reduce the variance of $\ln \kappa$, which means no uncertainty reduction in our estimate of $\ln \kappa$.

Now, let's consider that the prior covariance is estimated by an ensemble. Due to sampling errors, a spurious correlation may be introduced between ϕ and $\ln \kappa$. For instance, if we sample ϕ and $\ln \kappa$ independently and compute the correlation coefficient, it can result in spurious correlations, as illustrated in Fig. 7.1.

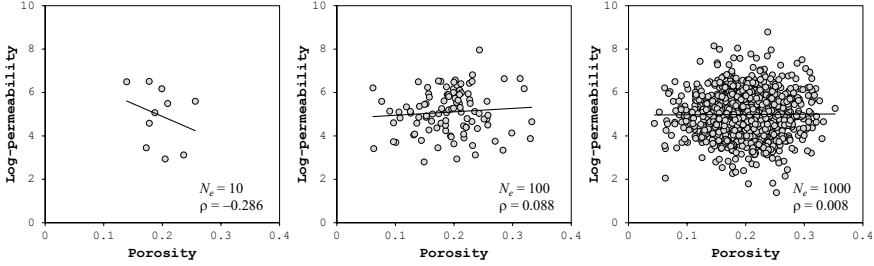


Fig. 7.1: Example showing spurious correlations between ϕ and $\ln \kappa$. Note that the correlation coefficient vanishes as we increase the size of the ensemble.

For concreteness, let us assume that the prior covariance estimated from the ensemble includes a small spurious cross-covariance c_ε . Also, assume that $|c_\varepsilon| < \sigma_\phi \sigma_{\ln \kappa}$.

$$\tilde{\mathbf{C}}_m = \begin{bmatrix} \sigma_\phi^2 & c_\varepsilon \\ c_\varepsilon & \sigma_{\ln \kappa}^2 \end{bmatrix}. \quad (7.10)$$

The posterior covariance becomes

$$\tilde{\mathbf{C}}_{m_c} = \begin{bmatrix} \sigma_\phi^2 - \frac{\sigma_\phi^4}{\sigma_e^2 + \sigma_\phi^2} & c_\varepsilon - \frac{c_\varepsilon \sigma_\phi^2}{\sigma_e^2 + \sigma_\phi^2} \\ c_\varepsilon - \frac{c_\varepsilon \sigma_\phi^2}{\sigma_e^2 + \sigma_\phi^2} & \sigma_{\ln \kappa}^2 - \frac{c_\varepsilon^2}{\sigma_e^2 + \sigma_\phi^2} \end{bmatrix}. \quad (7.11)$$

Therefore, the variance of $\ln \kappa$ is reduced regardless of the sign of c_ε , indicating that the uncertainty in $\ln \kappa$ is underestimated. Additionally, the mean is also altered. This effect diminishes as the ensemble size increases, since $c_\varepsilon \rightarrow 0$ as $N_e \rightarrow \infty$.

We can extend this example³ and assume that there is an actual correlation between ϕ and $\ln \kappa$ so that

$$\mathbf{C}_\mathbf{m} = \begin{bmatrix} \sigma_\phi^2 & \rho\sigma_\phi\sigma_{\ln \kappa} \\ \rho\sigma_\phi\sigma_{\ln \kappa} & \sigma_{\ln \kappa}^2 \end{bmatrix}. \quad (7.12)$$

But suppose, once again, that instead of $\mathbf{C}_\mathbf{m}$, we have an inaccurate version, $\tilde{\mathbf{C}}_\mathbf{m}$, where the variances are correct but the cross-covariances are corrupted with additive sampling errors

$$\tilde{\mathbf{C}}_\mathbf{m} = \begin{bmatrix} \sigma_\phi^2 & \rho\sigma_\phi\sigma_{\ln \kappa} + c_\varepsilon \\ \rho\sigma_\phi\sigma_{\ln \kappa} + c_\varepsilon & \sigma_{\ln \kappa}^2 \end{bmatrix}. \quad (7.13)$$

We already showed (Eq. (2.59)) that the posterior mean of $\ln \kappa$ for the case with correct $\mathbf{C}_\mathbf{m}$ is

$$(\ln \kappa)_{\text{map}} = (\ln \kappa)_{\text{pr}} + \frac{\rho\sigma_\phi\sigma_{\ln \kappa}}{\sigma_\phi^2 + \sigma_e^2} (\phi_{\text{obs}} - \phi_{\text{pr}}). \quad (7.14)$$

The corresponding value for the corrupted covariance is

$$\widetilde{(\ln \kappa)}_{\text{map}} = (\ln \kappa)_{\text{pr}} + \left(\frac{\rho\sigma_\phi\sigma_{\ln \kappa} + c_\varepsilon}{\sigma_\phi^2 + \sigma_e^2} \right) (\phi_{\text{obs}} - \phi_{\text{pr}}). \quad (7.15)$$

We want to analyze the error in the estimate of $\ln \kappa$.

$$\ln \kappa - \widetilde{(\ln \kappa)}_{\text{map}} = \ln \kappa - (\ln \kappa)_{\text{pr}} - \left(\frac{\rho\sigma_\phi\sigma_{\ln \kappa} + c_\varepsilon}{\sigma_\phi^2 + \sigma_e^2} \right) (\phi + e_\phi - \phi_{\text{pr}}), \quad (7.16)$$

where we wrote $\phi_{\text{obs}} = \phi + e_\phi$ with $e_\phi \sim \mathcal{N}(0, \sigma_e^2)$. The error variance becomes

³ This extension is inspired in the discussion presented in [183] and [245].

$$\begin{aligned}
 \mathbb{E} \left[\left(\ln \kappa - \widetilde{(\ln \kappa)_{\text{map}}} \right)^2 \right] &= \\
 &= \mathbb{E} \left[\left((\ln \kappa - (\ln \kappa)_{\text{pr}}) - \left(\frac{\rho \sigma_\phi \sigma_{\ln \kappa} + c_\varepsilon}{\sigma_\phi^2 + \sigma_e^2} \right) ((\phi - \phi_{\text{pr}}) + e_\phi) \right)^2 \right] \\
 &= \underbrace{\mathbb{E} \left[(\ln \kappa - (\ln \kappa)_{\text{pr}})^2 \right]}_{=\sigma_{\ln \kappa}^2} - \frac{2(\rho \sigma_\phi \sigma_{\ln \kappa} + c_\varepsilon)}{\sigma_\phi^2 + \sigma_e^2} \underbrace{\mathbb{E} [(\ln \kappa - (\ln \kappa)_{\text{pr}}) (\phi - \phi_{\text{pr}})]}_{=c_{\ln \kappa, \phi} = \rho \sigma_\phi \sigma_{\ln \kappa}} \\
 &\quad - \frac{2(\rho \sigma_\phi \sigma_{\ln \kappa} + c_\varepsilon)}{\sigma_\phi^2 + \sigma_e^2} \underbrace{\mathbb{E} [(\ln \kappa - (\ln \kappa)_{\text{pr}}) e_\phi]}_{=0} + \left(\frac{\rho \sigma_\phi \sigma_{\ln \kappa} + c_\varepsilon}{\sigma_\phi^2 + \sigma_e^2} \right)^2 \underbrace{\mathbb{E} [(\phi - \phi_{\text{pr}})^2]}_{=\sigma_\phi^2} \\
 &\quad + 2 \left(\frac{\rho \sigma_\phi \sigma_{\ln \kappa} + c_\varepsilon}{\sigma_\phi^2 + \sigma_e^2} \right)^2 \underbrace{\mathbb{E} [(\phi - \phi_{\text{pr}}) e_\phi]}_{=0} + \left(\frac{\rho \sigma_\phi \sigma_{\ln \kappa} + c_\varepsilon}{\sigma_\phi^2 + \sigma_e^2} \right)^2 \underbrace{\mathbb{E} [e_\phi^2]}_{=\sigma_e^2} \\
 &= \sigma_{\ln \kappa}^2 - \frac{2(\rho \sigma_\phi \sigma_{\ln \kappa} + c_\varepsilon)}{\sigma_\phi^2 + \sigma_e^2} \rho \sigma_\phi \sigma_{\ln \kappa} + \left(\frac{\rho \sigma_\phi \sigma_{\ln \kappa} + c_\varepsilon}{\sigma_\phi^2 + \sigma_e^2} \right)^2 (\sigma_\phi^2 + \sigma_e^2) \\
 &= \sigma_{\ln \kappa}^2 - \frac{\rho^2 \sigma_\phi^2 \sigma_{\ln \kappa}^2}{\sigma_\phi^2 + \sigma_e^2} + \frac{c_\varepsilon^2}{\sigma_\phi^2 + \sigma_e^2}. \quad (7.17)
 \end{aligned}$$

The last equality can be written as

$$\mathbb{E} \left[\left(\ln \kappa - \widetilde{(\ln \kappa)_{\text{map}}} \right)^2 \right] = \sigma_{\ln \kappa}^2 - \frac{\rho^2 \sigma_\phi^2 \sigma_{\ln \kappa}^2}{\sigma_\phi^2 + \sigma_e^2} \left(1 - \left(\frac{c_\varepsilon}{\rho \sigma_\phi \sigma_{\ln \kappa}} \right)^2 \right). \quad (7.18)$$

Hamill et al. [183] interpreted the term $\frac{c_\varepsilon}{\rho \sigma_\phi \sigma_{\ln \kappa}}$ as a “relative error” in the covariance. From Eq. (7.18), we can highlight two points: (i) if $\left\| \frac{c_\varepsilon}{\rho \sigma_\phi \sigma_{\ln \kappa}} \right\| > 1$, the variance of $\ln \kappa$ increases by assimilating data, which is clearly inconsistent. There is a degradation in the estimate of $\ln \kappa$ due to data assimilation. (ii) If the actual correlation coefficient between ϕ and $\ln \kappa$ is small, there is a higher likelihood of degradation in the estimates due to sampling error.

The second point is particularly significant because weak correlations are more challenging to estimate with limited samples. Lacerda et al. [245] illustrated this aspect by conducting a small numerical experiment to estimate the correlation coefficient between two parameters, ϕ and $\ln \kappa$, based on an ensemble of $N_e = 50$ members. They considered two cases where the true correlations are $\rho = 0$ and $\rho = 0.8$, and repeated the sampling 100 times. Fig. 7.2 shows the histograms of the estimates of ρ , indicating a large standard deviation for the case with $\rho = 0$.

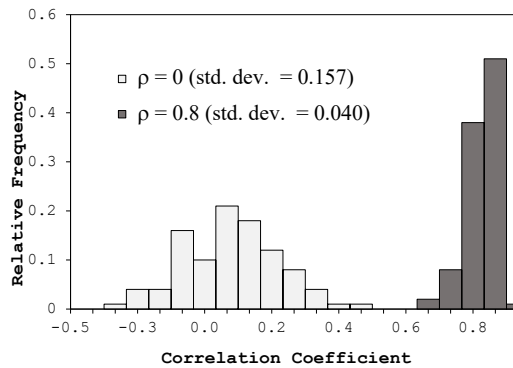


Fig. 7.2: Histograms of the estimated correlation coefficient for two cases: uncorrelated variables ($\rho = 0$) and strongly correlated variables ($\rho = 0.8$). Reproduced from Lacerda et al. [245] with permission from Elsevier.

7.3.2 Example: Porosity Distribution in a Core Sample

Recall the problem introduced in Section 2.3.1.4. In Fig. 4.5, the EnKF results for 100 members were presented. Now, we investigate the same problem using EnKF with $N_e = 10$. Fig. 7.3 shows the data assimilation results for each of the 10 data points. It can be observed that the ensemble was able to match most data points, albeit with a significant reduction in ensemble variability. However, at the last datum, it is possible to note an overcorrection in the values of porosity. This essentially happens because the ensemble variance becomes so small that the inversion becomes unstable.

7.4 Covariance Inflation

One *ad hoc* strategy to compensate for the excessive reduction in ensemble variance is known as covariance inflation [15]. In covariance inflation, we replace the forecast ensemble by

$$\mathbf{y}_{\text{inf},j}^{n,f} = \gamma(\mathbf{y}_j^{n,f} - \bar{\mathbf{y}}^{n,f}) + \bar{\mathbf{y}}^{n,f} \quad \text{for } j = 1, 2, \dots, N_e, \quad (7.19)$$

where γ is the inflation factor, a number slightly greater than one, say $\gamma = 1.05$.

Note that Eq. (7.19) increases the variance but does not change the mean of $\mathbf{y}^{n,f}$. The optimal inflation factor is problem-dependent. There are several methods in the literature to compute the inflation coefficient; see, e.g., Raanes et al. [353] and references therein. Covariance inflation finds common use in oceanography and numerical weather prediction applications, particularly in the context of sequential data assimilation. However, it has

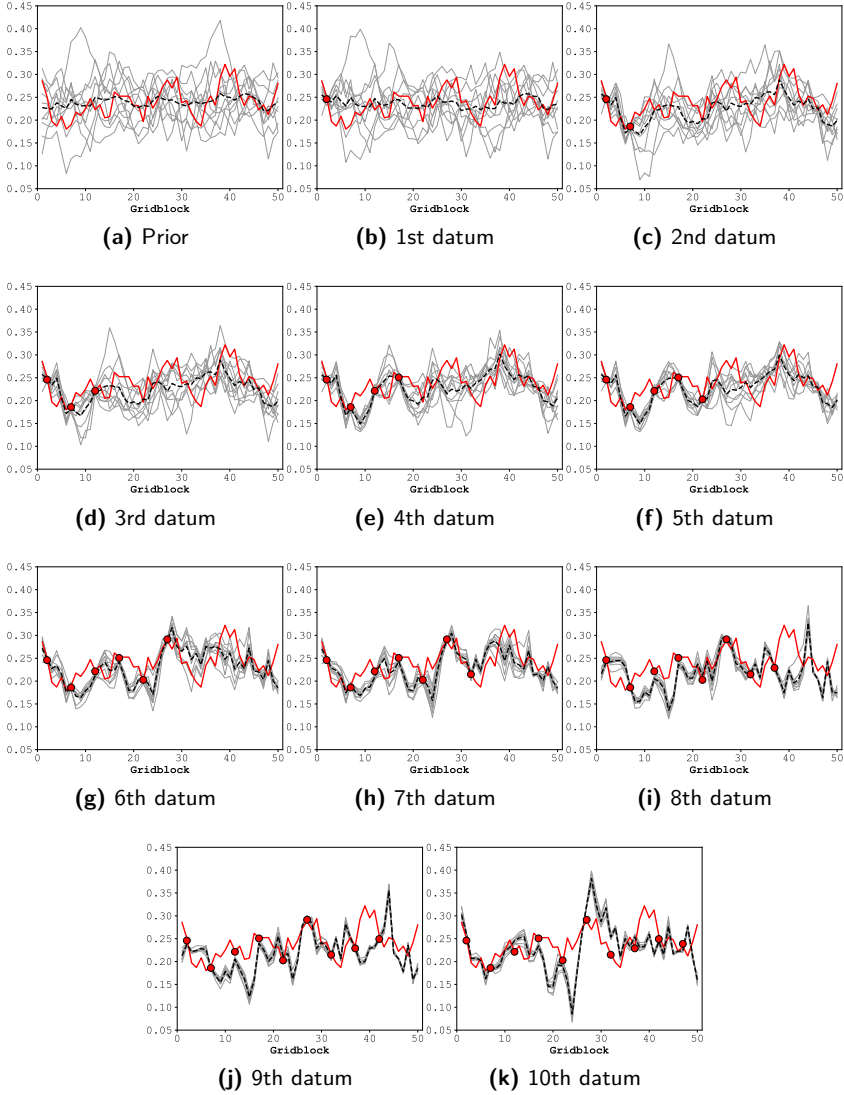


Fig. 7.3: EnKF sampling with $N_e = 10$ for porosity in a core sample. The solid red line represents the ground truth. The red circles represent the observed data points. The gray lines represent samples obtained by EnKF for different numbers of observations. The dashed lines represent the ensemble mean. Example 7.3.2.

rarely been applied with iterative ensemble smoothers for reservoir data assimilation.

7.5 Localization

Localization [197] is another *ad hoc* procedure employed to mitigate the negative effects of small ensembles. Typically, localization assumes that the covariances between model parameters and data are a function of distance. In its simplest form, localization aims to restrict the influence of a component of the innovation vector (data mismatch term) to a region around the data location. Localization is highly effective, simple to implement, and computationally inexpensive. However, determining the appropriate localization region depends on the specific problem.

The ES-MDA analysis with covariance localization can be expressed as:

$$\mathbf{m}_j^{\ell+1} = \mathbf{m}_j^\ell + \left(\mathbf{R}_{\text{md}} \circ \tilde{\mathbf{C}}_{\text{md}}^\ell \right) \left(\mathbf{R}_{\text{dd}} \circ \tilde{\mathbf{C}}_{\text{dd}}^\ell + \alpha_{\ell+1} \mathbf{C}_e \right)^{-1} \times \left(\mathbf{d}_{\text{obs},j}^\ell - \mathbf{g}(\mathbf{m}_j^\ell) \right). \quad (7.20)$$

Here, $\mathbf{R}_{\text{md}}$ and $\mathbf{R}_{\text{dd}}$ are the localization matrices with dimensions $N_m \times N_d$ and $N_d \times N_d$, respectively. The symbol “ $\circ$ ” denotes the Schur or Hadamard product, which represents the element-wise product of the components of the matrices (see Remark 7.2).

Often, instead of “localizing” the matrices $\tilde{\mathbf{C}}_{\text{md}}^\ell$ and $\tilde{\mathbf{C}}_{\text{dd}}^\ell$, we apply the Schur product directly to the Kalman gain, in which case the ES-MDA analysis can be expressed as:

$$\mathbf{m}_j^{\ell+1} = \mathbf{m}_j^\ell + \left(\mathbf{R}_{\text{md}} \circ \tilde{\mathbf{K}}^\ell \right) \left(\mathbf{d}_{\text{obs},j}^\ell - \mathbf{g}(\mathbf{m}_j^\ell) \right) \quad (7.21)$$

where

$$\tilde{\mathbf{K}}^\ell = \tilde{\mathbf{C}}_{\text{md}}^\ell \left(\tilde{\mathbf{C}}_{\text{dd}}^\ell + \alpha_{\ell+1} \mathbf{C}_e \right)^{-1}. \quad (7.22)$$

Covariance and Kalman gain localization appear to yield similar results [126]. However, Kalman gain localization offers some advantages in terms of computational implementation; in particular, Kalman gain localization allows the use of subspace inversion, as discussed in Chapter 8.

Remark 7.2: Schur (Hadamard) Product

Let $\mathbf{A}$ and $\mathbf{B}$ be $N_1 \times N_2$ matrices. The Schur product of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A} \circ \mathbf{B}$, is defined by the element-by-element product

$$[A \circ B]_{ij} = A_{ij}B_{ij} \quad (7.23)$$

where A_{ij} and B_{ij} are the (i, j) entry of the matrices $\mathbf{A}$ and $\mathbf{B}$, respectively. Note that $\mathbf{A}$ and $\mathbf{B}$ must be of same size but not necessarily square.

Schur Product Theorem: *if the matrix $\mathbf{A}$ is positive definite and the matrix $\mathbf{B}$ is positive semidefinite with all of its main diagonal entries positive, then the product $\mathbf{A} \circ \mathbf{B}$ is positive definite [196].*

7.5.1 Spurious Correlation

The original motivation behind localization is to eliminate or minimize the presence of spurious values resulting from sampling errors in the estimates of covariances and/or Kalman gain. Initially, localization was conceptualized as a cut-off radius, where only data within the radius were considered for computing the Kalman gain [197]. However, this method could lead to discontinuities in the estimates. To address this issue, Houtekamer and Mitchell [198] proposed using the Schur product with a localization matrix computed with a smooth correlation function. This alternative has proven to be more effective than a simple cut-off radius because it acts as a regularization strategy for the covariance/Kalman gain estimates.

Fig. 7.4 illustrates the effect of localization on covariance estimates. Panel (a) displays covariance values estimated with ensembles of 50, 100, and 1,000 samples. For comparison, the actual covariance computed with a spherical covariance function is included in each plot. Notice that while the covariance estimates tend to capture regions with higher covariance, they fail to vanish at positions with zero covariance. Increasing the size of the ensemble improves the estimates, but a considerable amount of spurious covariance values persists even for $N_e = 1,000$. Panel (b) shows the correlation function used to taper (localize) the covariance estimates. Panel (c) displays the covariance estimates after applying the Schur product with the correlation function, demonstrating significant improvement in the estimated covariance values.

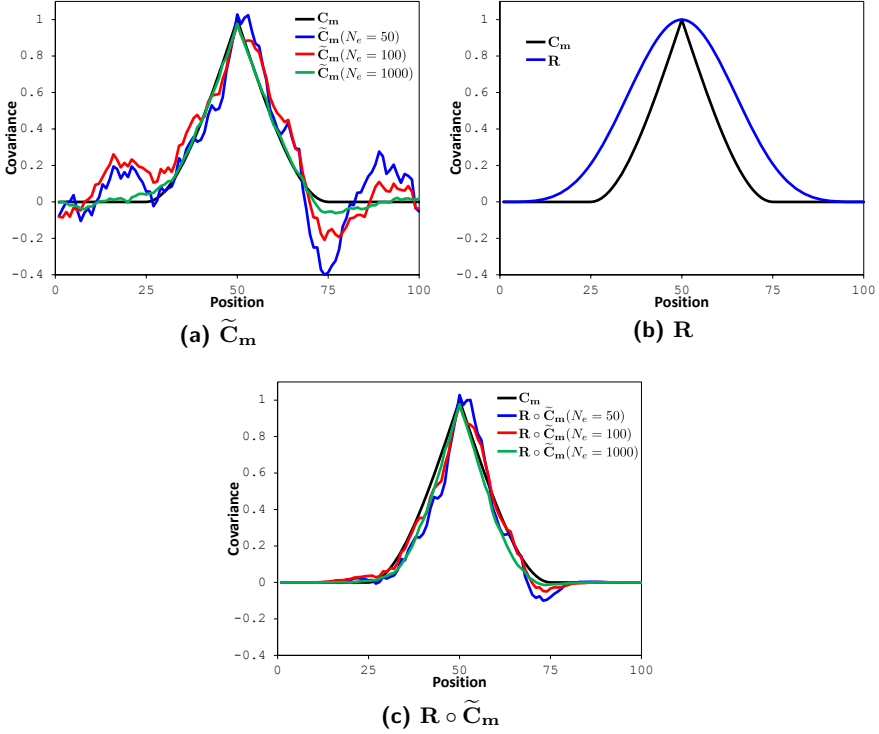


Fig. 7.4: Illustration of the effect of localization in removing spurious correlations. (a) Covariance estimates with ensembles of 50, 100, and 1,000 samples. (b) Gaspari-Cohn correlation function used to taper the covariance estimates. (c) Covariance estimates after localization.

7.5.2 Degrees of Freedom

As discussed in Section 7.2, $\text{rank}(\tilde{\mathbf{C}}_{\mathbf{m}}) \leq \{N_m, N_e - 1\}$. However, this matrix is at least positive semidefinite⁴ with positive values on the diagonal. Consequently, when we select a positive definite localization matrix $\mathbf{R}$, the localized version, $\mathbf{R} \circ \tilde{\mathbf{C}}_{\mathbf{m}}$, becomes full-rank (due to the Schur product theorem). This implies that localization enables the model update to be derived from a significantly larger space compared to the one spanned by the ensemble.

⁴ To show that $\tilde{\mathbf{C}}_{\mathbf{m}}$ is at least positive semidefinite, suffices to show that $\mathbf{x}^\top \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{x} \geq 0$ for any vector $\mathbf{x}$ (see Appendix A, Section A.2). Note that $\tilde{\mathbf{C}}_{\mathbf{m}} = \Delta \mathbf{M} \Delta \mathbf{M}^\top$. Hence

$$\mathbf{x}^\top \tilde{\mathbf{C}}_{\mathbf{m}} \mathbf{x} = (\Delta \mathbf{M} \mathbf{x})^\top (\Delta \mathbf{M} \mathbf{x}) = \|\Delta \mathbf{M} \mathbf{x}\|_2^2 \geq 0. \quad (7.24)$$

Fig. 7.5 demonstrates the impact of the Schur product on increasing the rank of a $1,000 \times 1,000$ covariance matrix. In Fig. 7.5a, the number of non-zero singular values is constrained by the ensemble size. However, after applying the Schur product with a positive definite localization matrix, the number of non-zero singular values becomes 1,000, regardless of the ensemble size.

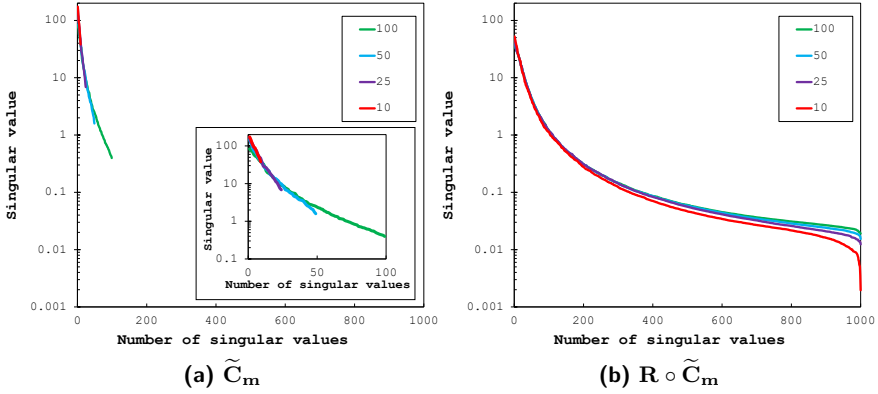


Fig. 7.5: Singular values of a covariance matrix estimated by ensembles, shown before (a) and after (b) applying the Schur product with a positive definite localization matrix. The legend indicates the ensemble sizes.

Another perspective on the impact of localization is to examine the effect of the Schur product in updating each component of $\mathbf{m}$. Previously, we demonstrated that $\mathbf{m}_{c,j}$ for ES without localization is a linear combination of the N_e prior vectors, $\mathbf{m}_j$. Following the same procedure for the case with Kalman gain localization, we obtain:

$$\begin{aligned}
 \mathbf{m}_{c,j} &= \mathbf{m}_j + \left(\mathbf{R}_{\text{md}} \circ \tilde{\mathbf{K}} \right) (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}_j)) \\
 &= \mathbf{m}_j + \left(\mathbf{R}_{\text{md}} \circ \tilde{\mathbf{K}} \right) \delta \mathbf{d}_j \\
 &= \mathbf{m}_j + \sum_{k=1}^{N_d} \left(\mathbf{R}_{\text{md}_k} \circ \tilde{\mathbf{K}}_k \right) \delta d_{jk},
 \end{aligned} \tag{7.25}$$

where $\mathbf{R}_{\text{md}_k}$ and $\tilde{\mathbf{K}}_k$ denote the k th columns of the localization matrix and Kalman gain, respectively. Now let's examine the update equation for the i th entry of $\mathbf{m}_j$, denoted by m_{ij} :

$$\begin{aligned}
 m_{c,ij} &= m_{ij} + \sum_{k=1}^{N_d} \tilde{K}_{ik} \overbrace{r_{ik} \delta d_{jk}}^{\theta_{ijk}} \\
 &= m_{ij} + \sum_{k=1}^{N_d} \theta_{ijk} \tilde{K}_{ik},
 \end{aligned} \tag{7.26}$$

where $i = 1, \dots, N_m$. r_{ik} is the (i, k) entry of the localization matrix $\mathbf{R}_{\text{md}}$, corresponding to the localization coefficient between the i th model parameter and the k th observation. This result shows that for each i , we have a different set of coefficients θ_{ijk} 's. Therefore, each component of $\mathbf{m}_j$ is updated with a different linear combination of the columns of $\tilde{\mathbf{K}}$. Thus, localization expands the degrees of freedom to assimilate data.

7.5.3 Distance-Based Localization

In distance-based localization, the localization matrix $\mathbf{R}_{\text{md}}$ used in the Schur product is computed based on the distance between model parameters and observation locations. Essentially, the localization coefficients assign weights to covariance values based on their spatial separation: closer points receive higher weights, while distant points receive lower weights. A common choice is to use the fifth-order compact correlation function proposed by Gaspari and Cohn [162] (Fig. 7.6) to compute the entries of $\mathbf{R}_{\text{md}}$. The Gaspari-Cohn correlation function can be expressed as:

$$r(h) = \begin{cases} -\frac{1}{4} \left(\frac{h}{L}\right)^5 + \frac{1}{2} \left(\frac{h}{L}\right)^4 + \frac{5}{8} \left(\frac{h}{L}\right)^3 - \frac{5}{3} \left(\frac{h}{L}\right)^2 + 1 & 0 \leq h \leq L \\ \frac{1}{12} \left(\frac{h}{L}\right)^5 - \frac{1}{2} \left(\frac{h}{L}\right)^4 + \frac{5}{8} \left(\frac{h}{L}\right)^3 + \frac{5}{3} \left(\frac{h}{L}\right)^2 - 5 \left(\frac{h}{L}\right) + 4 - \frac{2}{3} \left(\frac{h}{L}\right)^{-1} & L \leq h \leq 2L \\ 0 & h > 2L, \end{cases} \tag{7.27}$$

where h is the spatial distance and L is known as the critical length.

Another function often adopted for computing localization coefficients is the Furrer-Bengtsson pseudo-optimal taper [156]:

$$r(h) = \frac{1}{1 + (1 + \mathcal{C}(0)^2 / \mathcal{C}(h)^2) / N_e}, \tag{7.28}$$

where $\mathcal{C}(h)$ is a covariance function with compact support. Eq. (7.28) indicates that the strength of localization depends on the ensemble size, with the localization coefficients approaching one as the ensemble size tends to infinity.

In some situations, it may be desirable to use different critical lengths in different directions. This can be achieved by defining an ellipsoid with

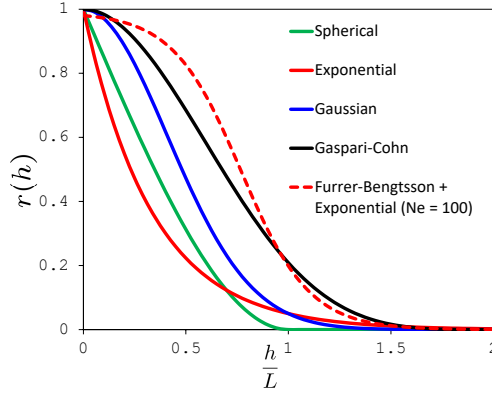


Fig. 7.6: Correlation functions frequently utilized in reservoir data assimilation. The dashed red curve represents the Furrer-Bengtsson taper with $\mathcal{C}(h)$ computed with an exponential correlation function and an ensemble with $N_e = 100$.

critical lengths in three orthogonal directions. This ellipsoid can be rotated to align, for instance, with the main directions specified in the geostatistical model used to generate the prior realizations. An ellipsoid rotated in the $x - y$ plane⁵ (Fig. 7.7) can be obtained using

$$\frac{h}{L} = \sqrt{\left(\frac{\Delta x'}{L_x}\right)^2 + \left(\frac{\Delta y'}{L_y}\right)^2 + \left(\frac{\Delta z'}{L_z}\right)^2}, \quad (7.29)$$

where

$$\begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}. \quad (7.30)$$

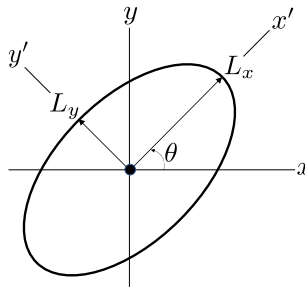


Fig. 7.7: Localization ellipsoid in the $x - y$ plane.

⁵ Appendix C, Section C.4.3 shows the equations for an ellipsoid rotated in the 3D space.

7.5.3.1 Example: Distance-Based Localization of the Kalman Gain

To illustrate the impact of distance-based localization on the Kalman gain, we examine a 2D reservoir model consisting of a grid of 60×60 blocks, operated with 16 oil-producing wells and nine water injection wells arranged in nine five-spot patterns. Figs. 7.8c and 7.8f show the Kalman gain values estimated from an ensemble of 100 permeability realizations for a datum of water cut and a datum of water injection rate for two wells, whose positions are highlighted in white. In both cases, large Kalman gain values are observed around the respective wells. However, especially in the water cut case, significant Kalman gain values are also observed in other regions of the model, far from the observation locations. For both cases, localization coefficients were computed using the Gaspari-Cohn function, with the main anisotropic direction aligned with the primary continuity direction in the permeability field (Figs. 7.8a and 7.8b). Figs. 7.8d and 7.8g show the corresponding Kalman gain estimates after applying localization. For comparison, we also computed the Kalman gain estimates with an ensemble of 5,000 realizations (Figs. 7.8e and 7.8h). Clearly, localization improved the Kalman gain estimates by removing spurious values away from the observation locations. However, comparing Fig. 7.8d with Fig. 7.8e, we observe that localization did not improve the Kalman gain estimate in the region connecting the injector and producer wells.

7.5.3.2 Example: Porosity Distribution in a Core Sample

Fig. 7.9 compares the data assimilation results for the example of porosity in a core sample (Example 7.3.2) with three different configurations: $N_e = 100$, $N_e = 10$, and $N_e = 10$ with distance-based localization. Clearly, localization led to significant improvements in the final estimates. It is notable that the ensemble mean is reasonably close to the mean estimate with a larger ensemble. Moreover, we still observe relevant variability in the posterior realizations. Additionally, we no longer observe overcorrection in the porosity estimates.

7.5.3.3 Example: Field Data Assimilation without Localization

Fig. 7.10 shows the predicted data for two wells in an oilfield obtained by assimilating data with ES-MDA, with and without localization [118]. These results illustrate that ES-MDA without localization was still able to reasonably match the data, although it resulted in considerably lower variability in the predictions compared to the case using localization. Fig. 7.11 presents the final permeability fields. Here, it is evident that in the absence of localization, the values of permeability were driven to the minimum and maximum limits (truncated at 1 mD and 5,000 mD, respectively). Another interesting

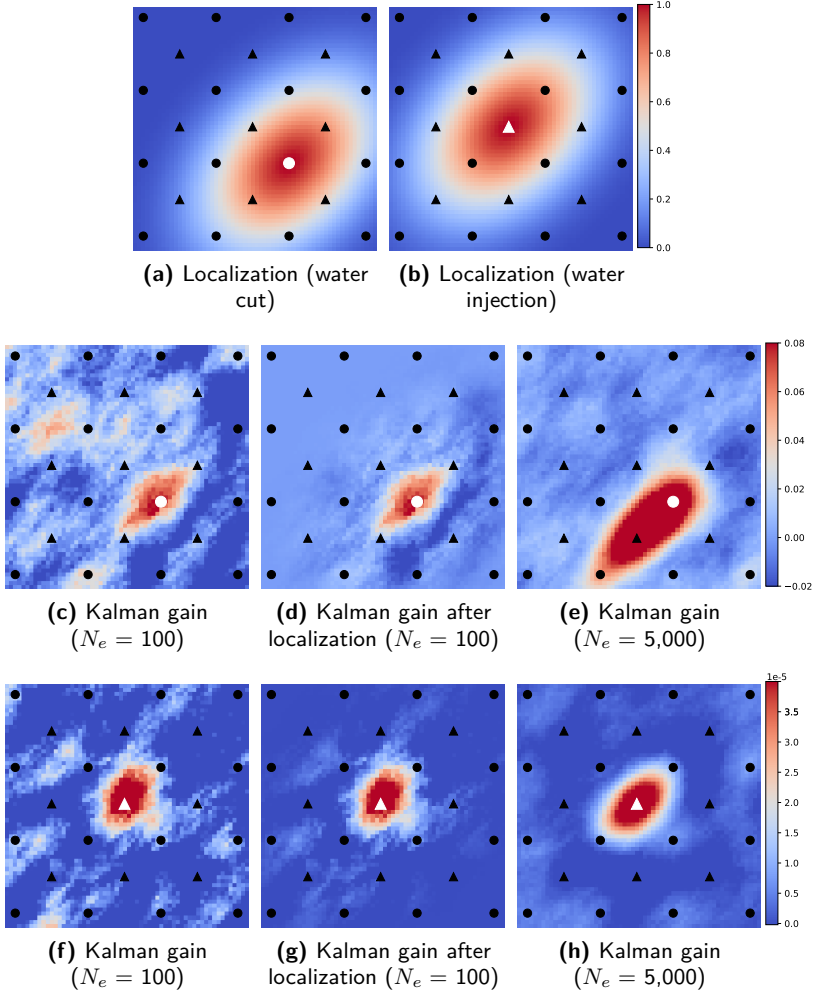


Fig. 7.8: Example illustrating the impact of distance-based localization on the Kalman gain. Panels (a) and (b) display localization values computed with the Gaspari-Cohn correlation function. Panels (c)–(e) display the Kalman gain estimates for water-cut data, while panels (f)–(h) show the estimates for water injection rate. Black dots represent the locations of production wells, and triangles indicate the positions of water injectors. White wells correspond to those where the Kalman calculation is performed. Example 7.5.3.1.

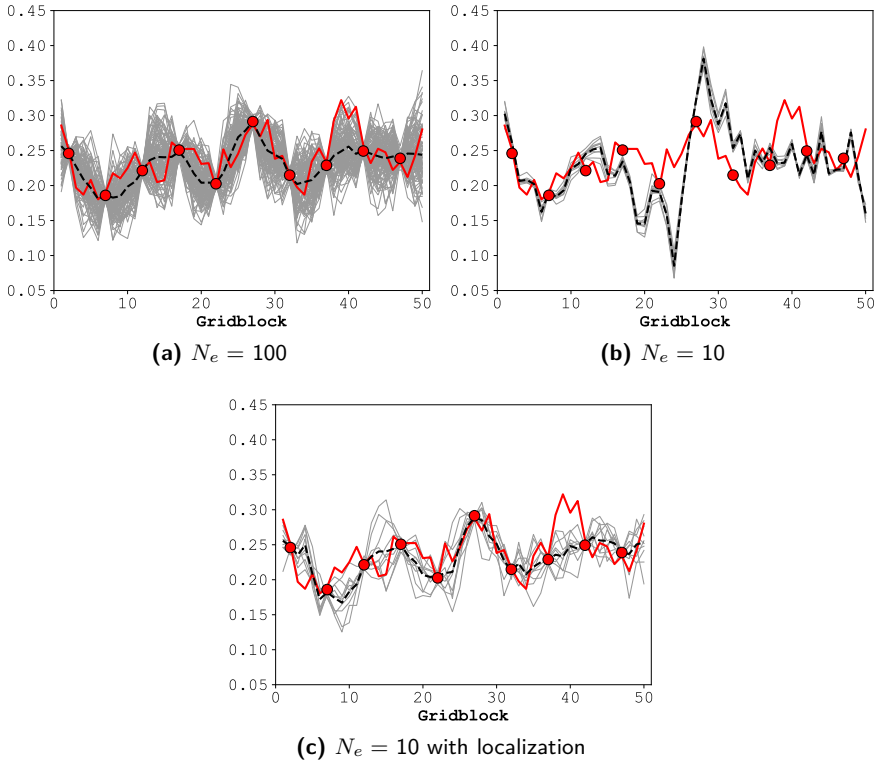


Fig. 7.9: Comparison of EnKF sampling with $N_e = 100$, $N_e = 10$ and $N_e = 10$ with localization for porosity in a core sample. The solid red line represents the ground truth. The red circles represent the observed data points. The gray lines represent samples obtained by EnKF for different numbers of observations. The dashed lines represent the ensemble mean. Example 7.5.3.2

observation is that Figs. 7.11e and 7.11f are almost identical, indicating a collapse of the ensemble, while the corresponding results with localization exhibit significant variability.

7.5.3.4 Selection of the Critical Length

Currently, distance-based localization is the most effective method for mitigating the negative effects of small ensembles. The “optimal” localization function depends on the specific problem, varying with the type of data, parameter, and ensemble size. However, choosing reasonable localization regions is not overly challenging. The Gaspari-Cohn correlation function is widely used, but the critical aspect is selecting the appropriate critical

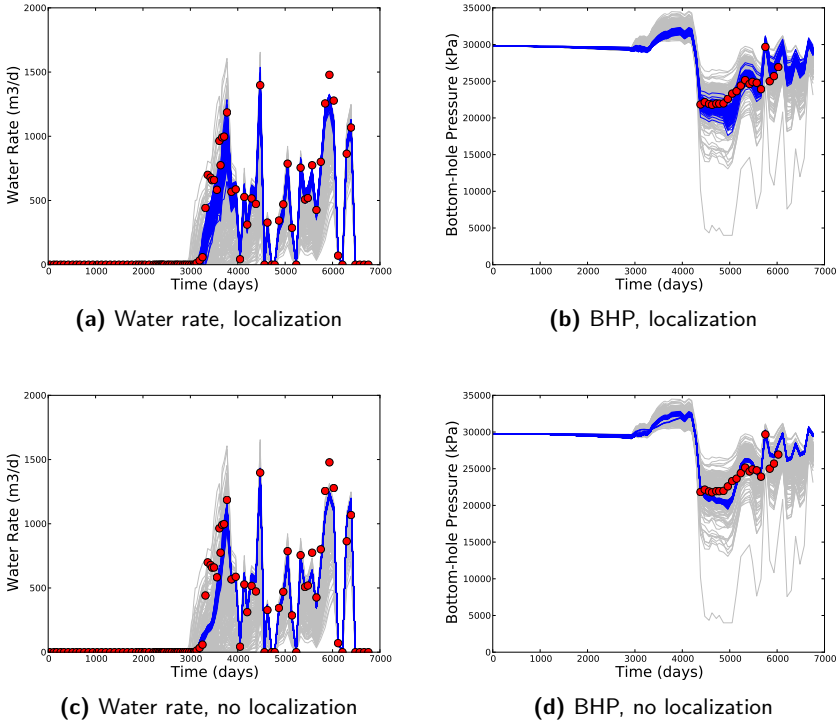


Fig. 7.10: Predicted data for two wells obtained by assimilating data with ES-MDA, with and without localization. Red circles are the observed data, gray lines the prior and blue lines the posterior ensemble. Example 7.5.3.3. Reproduced from Emerick [118] with permission from Elsevier.

lengths. Generally, underestimating the size of the localization region is more detrimental than overestimating it.

Emerick and Reynolds [128] argue that these critical lengths should be determined based on a combination of the data's sensitive region and the underlying correlation lengths of the geological model. Sensitivity analyses using the adjoint method [467, 259] have shown that each type of data exhibits a unique region of influence. For instance, bottom-hole pressure data from a flowing well are highly sensitive to gridblock permeabilities and porosities near the well, while water-cut data from a producing well are sensitive to porosities and permeabilities along streamlines connecting the producing well to a water injection well. Therefore, it is logical to consider the region of influence of the data when determining the critical lengths for localization. Using a localization region significantly smaller than the actual influence region may lead to disproportionately large local adjustments in model parameters.

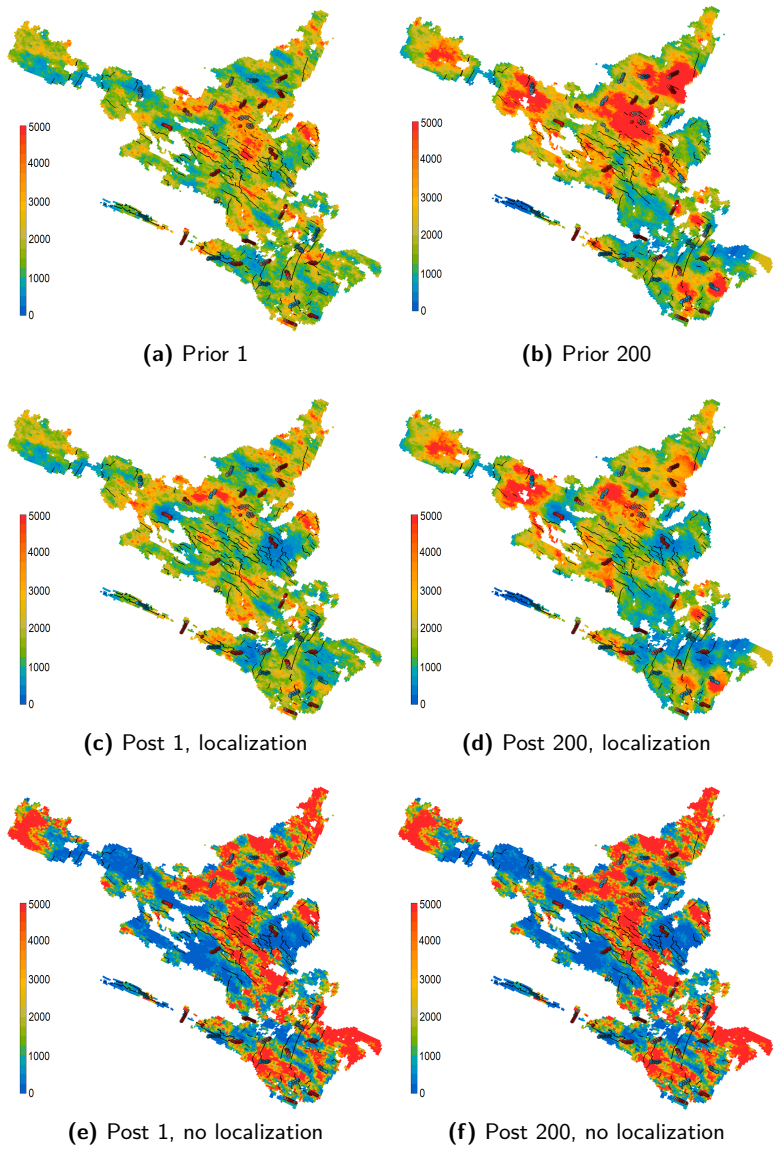


Fig. 7.11: Permeability realizations before (prior) and after data assimilation with ES-MDA, with and without localization. Example 7.5.3.3. Reproduced from Emerick [118] with permission from Elsevier.

Additionally, the localization region is influenced by geological structures, such as the covariance functions used in geostatistical modeling. Confining changes in model parameters to a region smaller than that dictated by the geological model's correlation lengths may introduce alterations that compromise the geological realism of the final ensemble.

7.5.4 Non-Distance Dependent Localization

Crucial parameters used in reservoir data assimilation, such as rock property multipliers, relative permeability curves, analytical aquifer parameters, and rock and fluid compressibilities, may lack a direct spatial relationship with observations. Unlike typical reservoir rock properties, which are discretized on the grid, these parameters are usually represented by scalar values with regional or even global impacts on model behavior. We refer to these scalar parameters as “non-local” parameters.

Additionally, we may encounter non-local data. For instance, one might be interested in assimilating indirect measurements such as estimates of the global reservoir pressure or the results of some data reparametrization [280, 281]. In these cases, a non-distance-dependent localization scheme becomes essential⁶.

There are several methods proposed in the literature for constructing localization matrices without relying on the assumption that covariances (or the Kalman gain) are distance-dependent [156, 14, 38, 482, 280, 452]. Lacerda et al. [245] reviewed some of these methods and compared their performance in updating non-local parameters in a synthetic reservoir data assimilation problem. Here, we review two of these methods.

7.5.4.1 Furrer-Bengtsson Taper Function

Furrer and Bengtsson [156] derived a localization function for general covariance structures by minimizing the expectation of the Frobenius norm of the difference between a covariance matrix, $\mathbf{C}$, and its localized estimate, $\mathbf{R} \circ \tilde{\mathbf{C}}$:

$$\mathbb{E} \left[\|\mathbf{C} - \mathbf{R} \circ \tilde{\mathbf{C}}\|_F^2 \right] = \text{tr}(\mathbf{C}^2) - 2\mathbb{E} \left[\text{tr} \left(\mathbf{C} \left(\mathbf{R} \circ \tilde{\mathbf{C}} \right) \right) \right] + \mathbb{E} \left[\left(\mathbf{R} \circ \tilde{\mathbf{C}} \right)^2 \right]. \quad (7.31)$$

⁶ The term “localization” was introduced by Houtekamer and Mitchell [198] with the concept of restricting the spatial extent of updates to eliminate long-distance correlations. Consequently, for non-local model parameters or data points, the term localization may seem less fitting. Nevertheless, we continue to use the term localization for historical reasons and because it is widely accepted in the literature.

They minimized Eq. (7.31) term-by-term ignoring the positive-definiteness constraint. The resulting expression depends on the true covariance:

$$r_{ik} = \frac{c_{ik}^2}{c_{ik}^2 + \frac{c_{ii}c_{kk}}{N_e}}. \quad (7.32)$$

Here, r_{ik} denotes the (i, k) th entry of the localization matrix $\mathbf{R}$ and c_{ik} denotes the corresponding covariance value. Furrer and Bengtsson [156] proposed to replace c_{ik} by the ensemble estimate, $\tilde{c}_{ik}$, forming a simple and computationally inexpensive non-distance-dependent localization procedure. The authors also suggested that sparseness can be introduced by zeroing small values of c_{ik} , which can be done setting $r_{ik} = 0$ if

$$\frac{|c_{ik}|}{\sqrt{c_{ii}c_{kk}}} = |\rho_{ik}| < \tau, \quad (7.33)$$

where $\tau \geq 0$ is a user-defined threshold. Instead of covariances, Eq. (7.32) can be rewritten in terms correlation coefficients as

$$r_{ik} = \frac{\tilde{\rho}_{ik}^2}{\tilde{\rho}_{ik}^2 + \frac{1 + \tilde{\rho}_{ikk}^2}{N_e}}, \quad (7.34)$$

where $\tilde{\rho}_{ik}$ is the correlation coefficient between the i th model parameter and the k th predicted data, which can be estimated from the current ensemble as

$$\tilde{\rho}_{ik} = \frac{\sum_{j=1}^{N_e} (m_{ij} - \bar{m}_i) (g_k(\mathbf{m}_j) - \overline{g_k(\mathbf{m})})}{\sqrt{\sum_{j=1}^{N_e} (m_{ij} - \bar{m}_i)^2} \sqrt{\sum_{j=1}^{N_e} (g_k(\mathbf{m}_j) - \overline{g_k(\mathbf{m})})^2}}. \quad (7.35)$$

Regrettably, the application of ensemble estimates of covariances in Eq. (7.32) for the estimation of localization coefficients suffers from spurious correlations, leading to a decrease in method effectiveness. This issue is illustrated in Fig. 7.12, which revisits the example presented in Section 7.5.1. In Fig. 7.12a, we observe the localization values calculated with the Furrer-Bengtsson taper using both true covariances and the covariance values estimated from an ensemble of 50 members. While the method utilizing ensemble estimates manages to identify regions with large actual covariances, it, unfortunately, fails to diminish at distant locations. This failure is a direct consequence of the challenge in estimating small covariance values, as illustrated in the example of Fig. 7.2. Consequently, the estimated covariances after localization were not significantly improved, as indicated in Fig. 7.12b.

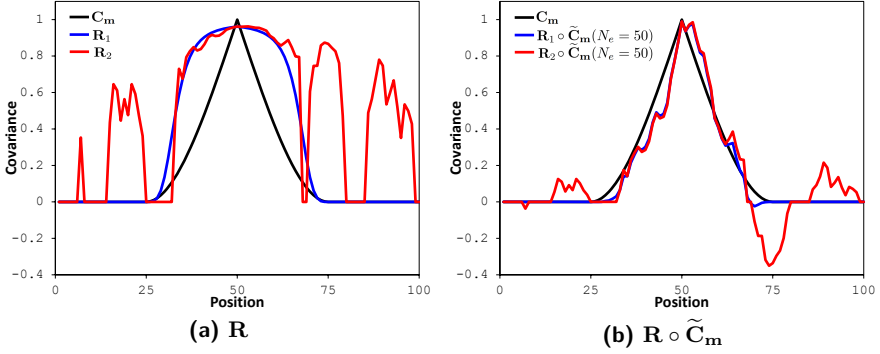


Fig. 7.12: Illustration of the Furrer-Bengtsson taper function. (a) Furrer-Bengtsson tapers estimated with the true covariance ($\mathbf{R}_1$) and covariance estimated with 50 ensemble members ($\mathbf{R}_2$). (b) Covariance estimates with $N_e = 50$ after localization. A threshold of $\tau = 0.1$ was adopted in the example. Note that the Schur product with $\mathbf{R}_1$ results in a significant improvement of the covariance estimate. However, the true covariances are typically unknown. $\mathbf{R}_2$, on the other hand, fails to eliminate distant spurious correlations.

7.5.4.2 Correlation-Based Localization

Luo et al. [280] proposed a localization approach based on the correlation coefficients estimated from the ensemble. In its simplest form, this method assigns a localization coefficient of one if the absolute value of the corresponding correlation coefficient exceeds a user-defined threshold (τ):

$$r_{ik} = \begin{cases} 1, & \text{if } |\tilde{\rho}_{ik}| > \tau \\ 0, & \text{otherwise.} \end{cases} \quad (7.36)$$

One alternative is to select τ proportional to the estimated error in the correlation coefficient, i.e.,

$$\tau = \gamma \sigma_\epsilon, \quad (7.37)$$

where $\gamma \in (0, 3]$ is a scaling factor, and σ_ϵ represents the standard deviation of the error in the correlation coefficient, which can be estimated as [407]:

$$\sigma_\epsilon = \frac{1 - \tilde{\rho}_{ik}^2}{\sqrt{N_e - 1}}. \quad (7.38)$$

Fig. 7.13 illustrates the localization threshold resulting from Eqs. (7.37) and (7.38) as a function of the estimated correlation coefficient for different ensemble sizes. Smaller values of $|\tilde{\rho}_{ik}|$ and N_e require higher truncation values.

Luo et al. [280] proposed a method for selecting τ for grid-based model parameters. They suggested assuming that the sample correlation can be decomposed as

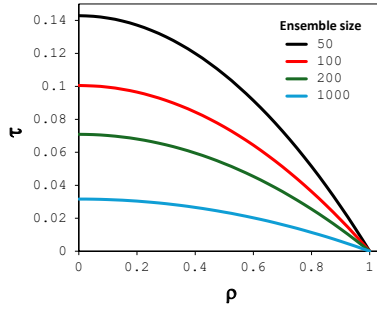


Fig. 7.13: Localization threshold as a function of the estimated correlation coefficient ($\gamma = 1$).

$$\tilde{\rho}_{ik} = \rho_{ik} + \epsilon_{ik}, \quad (7.39)$$

where ρ_{ik} is the true correlation and ϵ_{ik} represents the noise due to sampling error, with $\epsilon_{ik} \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The authors proposed dividing the grid parameters into groups. For example, one group assigned to each petrophysical property. For each group, they defined the random vector ϵ of dimension N_ϵ containing the noise components of the estimated correlation coefficients. The parameter σ_ϵ is then computed using the median absolute deviation (MAD) estimator [113]

$$\sigma_\epsilon = \frac{\text{med}(\text{abs}(\epsilon))}{0.6745}, \quad (7.40)$$

where $\text{med}(\text{abs}(\epsilon))$ denotes the median of the absolute values of the components of the vector ϵ . The threshold τ is computed using the universal rule [112]:

$$\tau = \sqrt{2 \ln(N_\epsilon)} \sigma_\epsilon. \quad (7.41)$$

This procedure relies on a proper assessment of the noise level in the correlation estimates or, at least, a good estimate of its standard deviation σ_ϵ . Luo et al. [280] discuss two approaches to estimate ϵ . The first approach, termed as “ideal,” proposes the use of a large ensemble, typically with a few thousand realizations, to estimate the true correlation values ρ_{ik} in Eq. (7.39), and then, by difference, estimate the noise. Evidently, this approach has the drawback of requiring a large ensemble, which prevents the use of the method in practice. The second approach, termed as “practical,” suggests using techniques from image denoising to separate the noise from the sample correlation values. Later, Luo and Bhakta [278] proposed a simpler procedure to estimate the noise level in the correlation estimates. They suggested randomly shuffling the indices of the ensemble members to create a new ensemble statistically independent of the predicted data.

They assumed that any correlation coefficient between model parameters and data in the shuffled ensemble is due to sampling errors and use these values as a proxy for ϵ . Luo and Bhakta [278] also proposed to replace the hard threshold scheme of Eq. (7.36) by a continuous tapering rule

$$r_{ik} = r\left(\frac{1 - |\tilde{\rho}_{ik}|}{1 - \tau}\right), \quad (7.42)$$

where $r(\cdot)$ is a correlation function. For example, the Gaspari-Cohn correlation function was adopted in [278].

Fig. 7.14 illustrates the effect of correlation-based localization on covariance estimates in the example presented in Section 7.5.1. The localization coefficients were computed using Eq. (7.42) with a threshold of $\tau = 0.1$. The results in Fig. 7.14 show that correlation-based localization was efficient in removing long-distance spurious covariances; it also removed some actual covariances, especially for the case with $N_e = 1,000$.

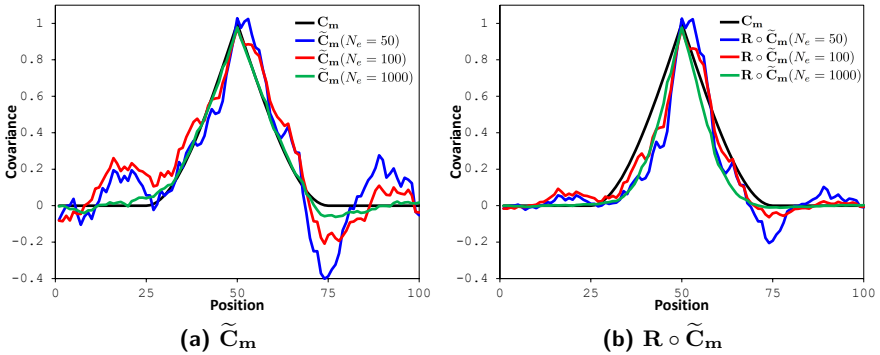


Fig. 7.14: Illustration of the effect of localization in removing spurious correlations. (a) Covariance estimates with ensembles of 50, 100, and 1,000 samples. (b) Covariance estimates after correlation-based localization. A threshold of $\tau = 0.1$ was adopted in the example.

Recently, Vishny et al. [452] introduced a method called noise-informed covariance estimation (NICE), which combines similar ideas used in correlation-based localization and the Furrer-Bengtsson taper.

7.5.4.3 Example: Correlation-Based Localization of the Kalman Gain

Fig. 7.15 demonstrates the impact of correlation-based localization on the Kalman gain for the same reservoir model of Example 7.5.3.1. Panels (a) and (d) show the correlation coefficient values estimated using Eq. (7.35), while panels (b) and (e) display the corresponding localization values calculated using Eq. (7.42) with the Gaspari-Cohn function for a threshold of

$\tau = 0.1$. Panels (c) and (f) present the Kalman gain values after applying localization. These results highlight that correlation-based localization significantly reduced spurious Kalman gain values in regions distant from the well locations. However, it also removed actual Kalman gain values close to the well, underestimating the region under actual Kalman gain influence (compare Figs. 7.15c and 7.8e).

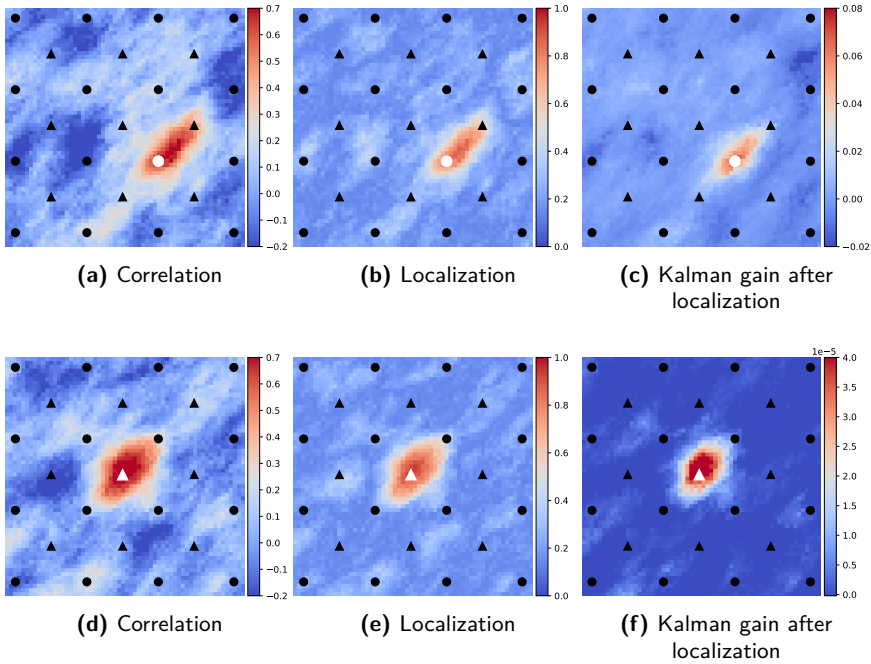


Fig. 7.15: Example illustrating correlation-based localization of the Kalman gain. The first row shows results for water cut data, and the second row shows results for water injection rate. Panels (a) and (d) display correlation coefficients, while panels (b) and (e) show localization values computed using Eq. (7.42) for $\tau = 0.1$ and the Gaspari-Cohn correlation function. Panels (c) and (f) show the Kalman gain values after localization. Black dots represent the locations of production wells, and triangles indicate the positions of water injectors. White wells correspond to those where the Kalman calculation is performed. Example 7.5.4.3.

7.5.4.4 Sensitivity Analysis to Define Localization Coefficients for Scalar Parameters

Experienced practitioners often select localization coefficients based on their knowledge of the field, for example, by assigning zero values to the co-

efficients of the localization matrix corresponding to parameters that are recognized as insensitive to the data. This process can be automated by performing a sensitivity analysis study. Pseudo-code 7.1 describes a one-at-a-time sensitivity analysis procedure to select localization coefficients for scalar parameters. In this approach, $r = 0$ indicates that the parameter is not sensitive to data, implying it should not be updated. Conversely, $r = 1$ is assigned if the parameter is sensitive to at least one time step, n , of the data time series. The experiments reported in Lacerda et al. [245] suggest that it is safer to set $r = 1$ for the entire time series rather than determining a value for r for each individual time step.

This procedure requires $2N_{m,s}$ reservoir simulations, where $N_{m,s}$ is the number of scalar parameters. However, two important considerations should be noted: firstly, if $N_{m,s}$ is large, this procedure can become computationally intensive; secondly, this method does not account for the interaction between parameters. During each simulation, only the value of the relevant parameter is changed, while the remaining parameters are kept at their base (mean) values.

Pseudo-code 7.1: One-at-a-time sensitivity analysis for localization of scalar parameters

For each scalar parameter ($i = 1, \dots, N_{m,s}$):

1. Run two reservoir simulations: one with the parameter at its minimum and another at its maximum, keeping all other parameters at their mean value:

$$\mathbf{d}_{\min} = \mathbf{g}(m_i = m_{i,\min}) \quad (7.43)$$

$$\mathbf{d}_{\max} = \mathbf{g}(m_i = m_{i,\max}) \quad (7.44)$$

2. Set $k = 0$
3. For each data source:
 - For each time step ($n = 1, \dots, N_t$):
 - $k = k + 1$
 - Set $r_{ik} = 0$
 - If the following condition is true

$$\frac{|d_{\max}^n - d_{\min}^n|}{\sigma_{e^n}} > \tau, \quad (7.45)$$

set $r_{ik} = 1$ for the entire time series ($\tau \geq 0$, say $\tau = 10^{-2}$).

7.5.4.5 Example: Localization of Scalar Parameters

Fig. 7.16 presents boxplots of scalar parameter values for a modified version of the PUNQ-S3 problem. This is the same problem considered in [245]. Each plot shows the actual parameter value (horizontal line in each plot), the prior distribution, and the posterior distributions obtained under different conditions: without localization (labeled as “No-local”), with localization coefficients computed using Eq. (7.34) (labeled as “FB”), and with a combination of Eq. (7.34) and the sensitivity analysis procedure described in Pseudo-code 7.1 (labeled as “FB+SA”). All data assimilation cases were based on ES-MDA with $N_e = 100$ and $N_a = 4$ with constant inflation factors. Each plot also includes results from a large ensemble, $N_e = 5,000$, without localization as a reference.

The localization coefficients computed using Eq. (7.34) were determined solely from the prior ensemble and remained fixed throughout the iterations. The results presented in [245] indicate that this procedure led to slightly better results than updating the localization coefficients at every ES-MDA iteration.

Two parameters are presented in Fig. 7.16. The first parameter corresponds to a change in the original water-oil contact, which significantly impacts the model predictions, with the true value being 1 meter. The data assimilation with the large ensemble resulted in a narrow posterior distribution centered around the true value. However, data assimilation without localization led to biased estimates of the water-oil contact. Both FB and FB+SA corrected this bias, accurately recovering the actual parameter value.

The second parameter is a dummy variable with a prior distribution sampled from a standard Gaussian. Since this parameter does not affect the simulation model, no updates during data assimilation are expected. However, the results in Fig. 7.16b show that the case without localization significantly reduced the posterior distribution, indicating updates solely due to sampling errors. While the FB case increased the posterior variability, it could not maintain the correct level of variance and resulted in biased final estimates. This poor performance is partly due to weak correlations being more prone to sampling errors. The introduction of the sensitivity analysis resulted in a zero localization coefficient for the dummy parameter, preventing any unnecessary updates.

7.5.5 Domain Localization

Domain localization, also known as local analysis, involves dividing the analysis into a collection of independent local analyses, each assimilating a subset of the observations (Fig. 7.17). These local analyses are conducted by assimilating only the observations within the vicinity of the parameter location. It is expected that the number of observations in these local regions

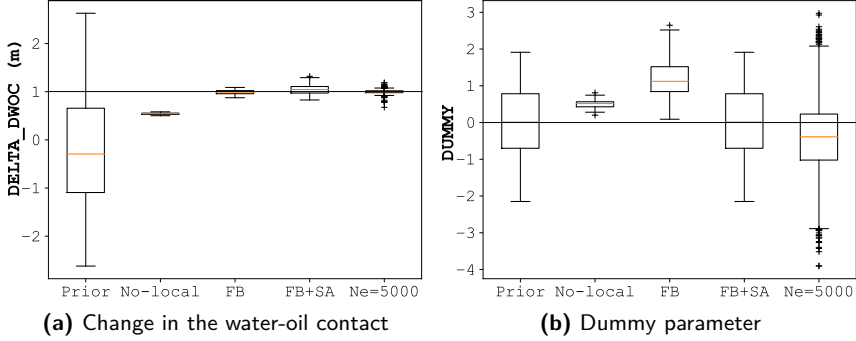


Fig. 7.16: Distribution of scalar parameter values for a modified version of the PUNQ-S3 problem with different localization setups. The horizontal line in each plot corresponds to the true parameter value. Example 7.5.4.5.

is small enough so that $N_e > N_d$ (resulting in locally full-rank matrices). Since the local regions overlap, a smooth transition between analyses is expected. One method to enhance these transitions is by obtaining a local $\mathbf{C}_e$ through tapering (Schur product) its inverse with a correlation function, which increases the variance for distant observations.

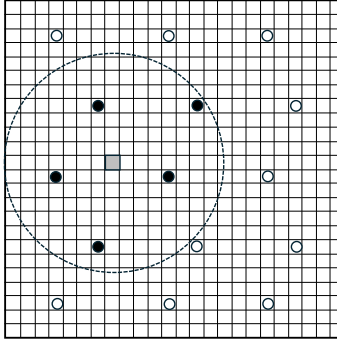


Fig. 7.17: Domain localization for updating the value of the parameter located at the highlighted gridblock. Only observations located within the localization region (depicted as the dashed circle) are used in the analysis.

Sakov and Bertino [372] showed that covariance localization and domain localization are not mathematically equivalent, but they tend to yield similar results when the analysis is primarily influenced by the prior. Chen and Oliver [81] concluded that LM-EnRML converges faster with domain localization compared to Kalman gain localization, particularly when the amount of data is large relative to the ensemble size. They also noted that

the optimal localization range for domain localization is shorter than that for Kalman gain localization. Additionally, domain localization can be applied to methods such as square root filters and SIES, where Schur product covariance or Kalman gain localization is not possible.

Computational Implementation of the Analysis

Abstract: *This chapter examines the matrix operations essential for efficiently implementing the analysis step, with a focus on optimizing memory usage. Special attention is given to ES-MDA with Kalman gain localization.*

8.1 Introduction

Ensemble-based methods are appreciated for their simplicity of implementation. However, naive approaches to implementing the analysis can quickly become computationally prohibitive. For example, consider a scenario with $N_m = 10^6$ parameters and $N_d = 10^4$ data points. In this case, the resulting Kalman gain matrix contains 10^{10} entries. Assuming double precision numbers (8 bytes), storing the Kalman gain requires 80 GB of RAM. Moreover, proper implementation of the inversion step of the analysis is critical, both in terms of performance and the quality of results.

This chapter delves into the matrix operations essential for efficiently implementing the analysis step, with a specific focus on robust inversion schemes and minimizing memory usage. While the discussion primarily centers on ES-MDA with Kalman gain localization, many of the concepts discussed are transferable to other variants of ensemble methods.

The ES-MDA update equation can be written as

$$\mathbf{m}_j^{\ell+1} = \mathbf{m}_j^\ell + \mathbf{R}_{\text{md}} \circ \left[\tilde{\mathbf{C}}_{\text{md}}^\ell \left(\tilde{\mathbf{C}}_{\text{dd}}^\ell + \alpha_{\ell+1} \mathbf{C}_e \right)^{-1} \right] \times (\mathbf{d}_{\text{obs},j}^\ell + \sqrt{\alpha_{\ell+1}} \mathbf{e}_j^\ell - \mathbf{g}(\mathbf{m}_j^\ell)), \quad (8.1)$$

for $j = 1, \dots, N_e$, where ℓ is the data assimilation index. In the following, we write the analysis in matrix form and drop the superscript ℓ , noting that all matrices are computed at the same data assimilation step.

8.2 Analysis in Matrix Form

We start defining the following matrices:

$$\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{N_e}], \quad (8.2)$$

with the j th column corresponding to the j th model vector $\mathbf{m}_j$, and

$$\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_{N_e}], \quad (8.3)$$

where $\mathbf{d}_j = \mathbf{g}(\mathbf{m}_j)$, and

$$\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N_e}], \quad (8.4)$$

with $\mathbf{e}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_e)$.

Then, we define the matrices of ensemble anomalies $\Delta \mathbf{M}$ and $\Delta \mathbf{D}$ by subtracting the ensemble mean vector from each column of $\mathbf{M}$ and $\mathbf{D}$, i.e.,

$$\begin{aligned} \Delta \mathbf{M} &= \frac{1}{\sqrt{N_e - 1}} (\mathbf{M} - \overline{\mathbf{M}}) \\ &= \frac{1}{\sqrt{N_e - 1}} [\mathbf{m}_1 - \overline{\mathbf{m}}, \dots, \mathbf{m}_{N_e} - \overline{\mathbf{m}}] \\ &= \mathbf{M} \mathbf{A} \end{aligned} \quad (8.5)$$

and

$$\begin{aligned} \Delta \mathbf{D} &= \frac{1}{\sqrt{N_e - 1}} (\mathbf{D} - \overline{\mathbf{D}}) \\ &= \frac{1}{\sqrt{N_e - 1}} [\mathbf{d}_1 - \overline{\mathbf{d}}, \dots, \mathbf{d}_{N_e} - \overline{\mathbf{d}}] \\ &= \mathbf{D} \mathbf{A}, \end{aligned} \quad (8.6)$$

where $\mathbf{A}$ is a centering matrix given by

$$\mathbf{A} = \frac{1}{\sqrt{N_e - 1}} \left(\mathbf{I} - \frac{1}{N_e} \mathbf{1} \mathbf{1}^\top \right), \quad (8.7)$$

with $\mathbf{1} = [1, \dots, 1]^\top$ with dimension N_e .

We also define an innovation matrix $\mathbf{J}$ as

$$\mathbf{J} = \mathbf{d}_{\text{obs}} \mathbf{1}^\top + \sqrt{\alpha} \mathbf{E} - \mathbf{D}. \quad (8.8)$$

With these matrices, we can express the ES-MDA analysis equation in the following form:

$$\mathbf{M}_{\text{updated}} = \mathbf{M} + \mathbf{R}_{\text{md}} \circ \left[\Delta \mathbf{M} \Delta \mathbf{D}^\top (\Delta \mathbf{D} \Delta \mathbf{D}^\top + \alpha \mathbf{C}_e)^{-1} \right] \mathbf{J}. \quad (8.9)$$

Table 8.1 summarizes the dimension of the matrices included in Eq. (8.9).

Table 8.1: Summary of matrices dimensions in Eq. (8.9).

Matrix	Dimension
$\mathbf{M}$	$N_m \times N_e$
$\mathbf{D}$	$N_d \times N_e$
$\mathbf{R}_{\text{md}}$	$N_m \times N_d$
$\mathbf{C}_e$	$N_d \times N_d$
$\mathbf{E}$	$N_d \times N_e$
$\mathbf{d}_{\text{obs}}$	$N_d \times 1$
$\Delta \mathbf{M}$	$N_m \times N_e$
$\Delta \mathbf{D}$	$N_d \times N_e$
$\mathbf{J}$	$N_d \times N_e$

8.3 Inversion

Eq. (8.9) involves the inversion of a $N_d \times N_d$ matrix $\mathbf{C}$ given by

$$\mathbf{C} \equiv \Delta \mathbf{D} \Delta \mathbf{D}^\top + \alpha \mathbf{C}_e. \quad (8.10)$$

Since the product $\Delta \mathbf{D} \Delta \mathbf{D}^\top$ forms a real-symmetric positive semidefinite matrix, $\mathbf{C}$ is real-symmetric and positive definite, at least when $\mathbf{C}_e$ is positive definite. Therefore, $\mathbf{C}$ has a real inverse. However, it may be poorly conditioned [139]. Hence, for implementation, we utilize a pseudoinverse of $\mathbf{C}$ computed through a truncated singular value decomposition (TSVD).

8.3.1 Pseudoinverse

Even though $\mathbf{C}$ is typically positive definite, it may be poorly scaled, particularly if it is constructed based on data with varying magnitudes. In this case, computing the pseudoinverse of $\mathbf{C}$ using TSVD may result in the loss of information necessary for data matching during truncation of small singular values.

Therefore, it is crucial to rescale the components of the matrix $\mathbf{C}$ before applying the TSVD. One approach is to rescale $\mathbf{C}$ using the diagonal elements of $\mathbf{C}_e$:

$$\mathbf{C}_e = \mathbf{S} \hat{\mathbf{C}}_e \mathbf{S}, \quad (8.11)$$

where $\mathbf{S}$ is the $N_d \times N_d$ diagonal matrix containing the square root of the diagonal components of $\mathbf{C}_e$, i.e., $\mathbf{S} = [\text{diag}(\mathbf{C}_e)]^{1/2}$. $\hat{\mathbf{C}}_e$ is the correlation matrix of the data errors.

Then,

$$\mathbf{C} = \mathbf{S} \hat{\mathbf{C}} \mathbf{S} = \mathbf{S} \left[\mathbf{S}^{-1} \Delta \mathbf{D} \Delta \mathbf{D}^\top \mathbf{S}^{-1} + \alpha \hat{\mathbf{C}}_e \right] \mathbf{S}. \quad (8.12)$$

Note that for the case with uncorrelated data-errors, $\hat{\mathbf{C}}_e = \mathbf{I}$.

Applying TSVD to $\hat{\mathbf{C}}$

$$\hat{\mathbf{C}} \approx \Phi_r \Lambda_r \Phi_r^\top, \quad (8.13)$$

where Φ_r is the $N_d \times N_r$ matrix with its j th column equal to the left singular vector of $\hat{\mathbf{C}}$ corresponding to the j th singular value. Λ_r is a diagonal matrix containing the N_r largest nonzero singular values of $\mathbf{C}$. We introduced the approximation sign because we keep only the $N_r \leq \min\{N_d, N_e - 1\}$ largest singular values of $\hat{\mathbf{C}}$.

The pseudoinverse of $\mathbf{C}$ assumes the form

$$\mathbf{C}^+ = \mathbf{S}^{-1} \hat{\mathbf{C}}^+ \mathbf{S}^{-1} = \mathbf{S}^{-1} \Phi_r \Lambda_r^{-1} \Phi_r^\top \mathbf{S}^{-1}. \quad (8.14)$$

Pseudo-code 8.1 summarizes the pseudoinverse procedure for the case with uncorrelated data errors.

Pseudo-code 8.1: Pseudoinverse with diagonal $\mathbf{C}_e$

1. $\mathbf{S} = [\text{diag}(\mathbf{C}_e)]^{1/2}$ {Data-error standard deviations}
2. $\hat{\Delta \mathbf{D}} = \mathbf{S}^{-1} \Delta \mathbf{D}$
3. $\hat{\mathbf{C}} = \hat{\Delta \mathbf{D}} \hat{\Delta \mathbf{D}}^\top + \alpha \mathbf{I}$
4. $\Phi_r \Lambda_r \Phi_r^\top = \hat{\mathbf{C}}$ {TSVD}
5. $\Omega = \mathbf{S}^{-1} \Phi_r$
6. $\Upsilon = \Lambda_r^{-1}$
7. $\mathbf{C}^+ = \Omega \Upsilon \Omega^\top$

8.3.2 Subspace Inversion

The computation cost of the TSVD procedure becomes impractical when the number of data points is large. Evensen [138] proposed a more computationally efficient approach known as subspace inversion. This method is particularly advantageous when the number of data points greatly exceeds the number of ensemble members, i.e., $N_d \gg N_e$.

Instead of directly computing the pseudoinverse of $\mathbf{C}$ through TSVD, subspace inversion involves performing SVD on $\widehat{\Delta\mathbf{D}} = \mathbf{S}^{-1}\Delta\mathbf{D}$ and truncate with the N_r largest singular values. The computational gain comes from the fact that the number of columns of $\widehat{\Delta\mathbf{D}}$ is N_e .

We start by writing the TSVD of $\widehat{\Delta\mathbf{D}}$ as

$$\widehat{\Delta\mathbf{D}} = \mathbf{S}^{-1}\Delta\mathbf{D} \approx \mathbf{U}_r \boldsymbol{\Sigma}_r \mathbf{V}_r^\top. \quad (8.15)$$

Again, the approximation is because we truncate small singular values. We write $\mathbf{C}$ as

$$\begin{aligned} \mathbf{C} &= \mathbf{S} \left[\mathbf{S}^{-1}\Delta\mathbf{D}\Delta\mathbf{D}^\top\mathbf{S}^{-1} + \alpha\widehat{\mathbf{C}}_e \right] \mathbf{S} \\ &\approx \mathbf{S} \left[\mathbf{U}_r \boldsymbol{\Sigma}_r \mathbf{V}_r^\top \mathbf{V}_r \boldsymbol{\Sigma}_r \mathbf{U}_r^\top + \alpha\widehat{\mathbf{C}}_e \right] \mathbf{S} \\ &\approx \mathbf{S} \mathbf{U}_r \boldsymbol{\Sigma}_r \left[\mathbf{I} + \alpha \boldsymbol{\Sigma}_r^{-1} \mathbf{U}_r^\top \widehat{\mathbf{C}}_e \mathbf{U}_r \boldsymbol{\Sigma}_r^{-1} \right] \boldsymbol{\Sigma}_r \mathbf{U}_r^\top \mathbf{S} \\ &= (\mathbf{S} \mathbf{U}_r \boldsymbol{\Sigma}_r) [\mathbf{I} + \boldsymbol{\Theta}] (\mathbf{S} \mathbf{U}_r \boldsymbol{\Sigma}_r)^\top. \end{aligned} \quad (8.16)$$

We introduced another approximation in the third line of the expression because $\mathbf{U}_r \mathbf{U}_r^\top \neq \mathbf{I}$. To justify this approximation, note first that $\mathbf{U}_r \mathbf{U}_r^\top$ is an orthogonal matrix that projects vectors onto the space spanned by the columns of $\mathbf{U}_r$. Let $\mathbf{x}$ be two vector in $\mathbb{R}^{N_d}$ and $\tilde{\mathbf{x}} = \mathbf{U}_r \mathbf{U}_r^\top \mathbf{x}$ its corresponding projection. Then

$$\begin{aligned} \tilde{\mathbf{x}}^\top \mathbf{C} \tilde{\mathbf{x}} &= \tilde{\mathbf{x}}^\top \mathbf{S} \left[\mathbf{U}_r \boldsymbol{\Sigma}_r \mathbf{V}_r^\top \mathbf{V}_r \boldsymbol{\Sigma}_r \mathbf{U}_r^\top + \alpha\widehat{\mathbf{C}}_e \right] \mathbf{S} \tilde{\mathbf{x}} \\ &= \mathbf{x}^\top \mathbf{S} \mathbf{U}_r \mathbf{U}_r^\top \left[\mathbf{U}_r \boldsymbol{\Sigma}_r \mathbf{V}_r^\top \mathbf{V}_r \boldsymbol{\Sigma}_r \mathbf{U}_r^\top + \alpha\widehat{\mathbf{C}}_e \right] \mathbf{U}_r \mathbf{U}_r^\top \mathbf{S} \mathbf{x} \\ &= \mathbf{x}^\top \mathbf{S} \mathbf{U}_r \boldsymbol{\Sigma}_r \left[\mathbf{I} + \alpha \boldsymbol{\Sigma}_r^{-1} \mathbf{U}_r^\top \widehat{\mathbf{C}}_e \mathbf{U}_r \boldsymbol{\Sigma}_r^{-1} \right] \boldsymbol{\Sigma}_r \mathbf{U}_r^\top \mathbf{S} \mathbf{x} \\ &= \mathbf{x}^\top \tilde{\mathbf{C}} \mathbf{x}. \end{aligned} \quad (8.17)$$

This means that the quadratic form induced by $\tilde{\mathbf{C}}$ corresponds to the quadratic form induced by $\mathbf{C}$ in the space spanned by the columns of $\mathbf{U}_r$.

In Eq. (8.16), we introduced the $N_r \times N_r$ matrix $\boldsymbol{\Theta}$ defined as

$$\boldsymbol{\Theta} \equiv \alpha \boldsymbol{\Sigma}_r^{-1} \mathbf{U}_r^\top \widehat{\mathbf{C}}_e \mathbf{U}_r \boldsymbol{\Sigma}_r^{-1}. \quad (8.18)$$

The matrix Θ is real-symmetric positive definite (at least when we choose $\mathbf{C}_e$ positive definite). Hence, we can use SVD to express the eigen-decomposition of Θ as

$$\Theta = \Psi \Gamma \Psi^\top. \quad (8.19)$$

Note that at this point, the rank of Θ is N_r and there is no reason for another truncation of singular values. Therefore, we can write $\mathbf{C}$ as

$$\mathbf{C} \approx (\mathbf{S} \mathbf{U}_r \Sigma_r \Psi) [\mathbf{I} + \Gamma] (\mathbf{S} \mathbf{U}_r \Sigma_r \Psi)^\top. \quad (8.20)$$

Thus, the pseudoinverse of $\mathbf{C}$ becomes

$$\mathbf{C}^+ = (\mathbf{S}^{-1} \mathbf{U}_r \Sigma_r^{-1} \Psi) [\mathbf{I} + \Gamma]^{-1} (\mathbf{S}^{-1} \mathbf{U}_r \Sigma_r^{-1} \Psi)^\top. \quad (8.21)$$

Note that Σ_r and $\mathbf{I} + \Gamma$ are diagonal matrices. Hence, the inverses are trivial to compute.

8.3.2.1 Case with Diagonal Data-Error Covariance

When we assume uncorrelated data errors, we have a diagonal matrix $\mathbf{C}_e$. In this situation, it suffices to use $\hat{\mathbf{C}}_e = \mathbf{I}$ and note that

$$\begin{aligned} \Theta &\equiv \alpha \Sigma_r^{-1} \mathbf{U}_r^\top \hat{\mathbf{C}}_e \mathbf{U}_r \Sigma_r^{-1} \\ &= \alpha \Sigma_r^{-1} \mathbf{U}_r^\top \mathbf{U}_r \Sigma_r^{-1} \\ &= \alpha \Sigma_r^{-2}, \end{aligned} \quad (8.22)$$

which is a diagonal matrix. Therefore, it is not necessary to do the SVD of Θ . Instead, write the pseudoinverse as

$$\mathbf{C}^+ = (\mathbf{S}^{-1} \mathbf{U}_r \Sigma_r^{-1}) [\mathbf{I} + \Theta]^{-1} (\mathbf{S}^{-1} \mathbf{U}_r \Sigma_r^{-1})^\top. \quad (8.23)$$

Pseudo-code 8.2 summarizes the subspace inversion procedure for the case with uncorrelated data errors.

Pseudo-code 8.2: Subspace inversion with diagonal $\mathbf{C}_e$

1. $\mathbf{S} = [\text{diag}(\mathbf{C}_e)]^{1/2}$ {Data-error standard deviations}
2. $\widehat{\Delta \mathbf{D}} = \mathbf{S}^{-1} \Delta \mathbf{D}$
3. $\mathbf{U}_r \Sigma_r \mathbf{V}_r^\top = \widehat{\Delta \mathbf{D}}$ {TSVD}
4. $\Theta = \alpha \Sigma_r^{-2}$
5. $\Omega = \mathbf{S}^{-1} \mathbf{U}_r \Sigma_r^{-1}$
6. $\Upsilon = (\mathbf{I} + \Theta)^{-1}$
7. $\mathbf{C}^+ = \Omega \Upsilon \Omega^\top$

8.3.2.2 Case with Cholesky Factorization of the Data-Error Covariance

For the case with correlated data errors, the matrix $\mathbf{C}_e$ is not diagonal. In this case, we typically use a square root of $\mathbf{C}_e$ to compute the data perturbations, $\mathbf{e}_j$'s. If the number of data points is not extremely large, the Cholesky decomposition can be used to compute the square root of $\mathbf{C}_e$. In this case, we write the Cholesky decomposition of the rescaled matrix $\mathbf{C}_e$ as

$$\hat{\mathbf{C}}_e = \mathbf{L}_e \mathbf{L}_e^\top, \quad (8.24)$$

where $\mathbf{L}_e$ is a lower-diagonal matrix. Then, the matrix Θ becomes

$$\Theta \equiv \alpha (\Sigma_r^{-1} \mathbf{U}_r^\top \mathbf{L}_e) (\Sigma_r^{-1} \mathbf{U}_r^\top \mathbf{L}_e)^\top. \quad (8.25)$$

The rest of the inversion remains unchanged. Pseudo-code 8.3 summarizes the subspace inversion procedure for the case with correlated data errors and Cholesky of $\mathbf{C}_e$.

Pseudo-code 8.3: Subspace inversion with Cholesky of $\mathbf{C}_e$

1. $\mathbf{S} = [\text{diag}(\mathbf{C}_e)]^{1/2}$ {Data-error standard deviations}
2. $\hat{\mathbf{C}}_e = \mathbf{S}^{-1} \mathbf{C}_e \mathbf{S}^{-1}$ {Correlation matrix}
3. $\mathbf{L}_e \mathbf{L}_e^\top = \hat{\mathbf{C}}_e$ {Cholesky}
4. $\hat{\Delta \mathbf{D}} = \mathbf{S}^{-1} \Delta \mathbf{D}$
5. $\mathbf{U}_r \Sigma_r \mathbf{V}_r^\top = \hat{\Delta \mathbf{D}}$ {TSVD}
6. $\Theta = \alpha (\Sigma_r^{-1} \mathbf{U}_r^\top \mathbf{L}_e) (\Sigma_r^{-1} \mathbf{U}_r^\top \mathbf{L}_e)^\top$
7. $\Psi \Gamma \Psi^\top = \Theta$ {SVD}
8. $\Omega = \mathbf{S}^{-1} \mathbf{U}_r \Sigma_r^{-1} \Psi$
9. $\Upsilon = (\mathbf{I} + \Gamma)^{-1}$
10. $\mathbf{C}^+ = \Omega \Upsilon \Omega^\top$

8.3.2.3 Case with Large Non-Diagonal Data-Error Covariance

For a case with a very large number of data points, it may not be feasible to store and factorize a large non-diagonal matrix $\mathbf{C}_e$. In these cases, one alternative is to construct a low-rank approximation of $\mathbf{C}_e$ using an ensemble of perturbed observations, i.e.,

$$\tilde{\mathbf{C}}_e \approx \frac{1}{N_e - 1} \mathbf{E} \mathbf{E}^\top. \quad (8.26)$$

However, instead of computing $\tilde{\mathbf{C}}_e$, we can apply the subspace inversion using $\mathbf{E}$. Before, we need to rescale $\mathbf{E}$ as

$$\widehat{\mathbf{E}} = \frac{1}{\sqrt{N_e - 1}} \mathbf{S}^{-1} \mathbf{E}. \quad (8.27)$$

Then, the matrix Θ is redefined as

$$\Theta \equiv \alpha \left(\Sigma_r^{-1} \mathbf{U}_r^\top \widehat{\mathbf{E}} \right) \left(\Sigma_r^{-1} \mathbf{U}_r^\top \widehat{\mathbf{E}} \right)^\top. \quad (8.28)$$

The rest of the inversion remains unchanged. Pseudo-code 8.4 summarizes this procedure.

Pseudo-code 8.4: Subspace inversion with ensemble representation of $\mathbf{C}_e$

1. $\mathbf{S} = [\text{diag}(\mathbf{C}_e)]^{1/2}$ {Data-error standard deviations}
2. $\widehat{\mathbf{E}} = \frac{1}{N_e - 1} \mathbf{S}^{-1} \mathbf{E}$
3. $\widehat{\Delta \mathbf{D}} = \mathbf{S}^{-1} \Delta \mathbf{D}$
4. $\mathbf{U}_r \Sigma_r \mathbf{V}_r^\top = \widehat{\Delta \mathbf{D}}$ {TSVD}
5. $\Theta = \alpha \left(\Sigma_r^{-1} \mathbf{U}_r^\top \widehat{\mathbf{E}} \right) \left(\Sigma_r^{-1} \mathbf{U}_r^\top \widehat{\mathbf{E}} \right)^\top$
6. $\Psi \Gamma \Psi^\top = \Theta$ {SVD}
7. $\Omega = \mathbf{S}^{-1} \mathbf{U}_r \Sigma_r^{-1} \Psi$
8. $\Upsilon = (\mathbf{I} + \Gamma)^{-1}$
9. $\mathbf{C}^+ = \Omega \Upsilon \Omega^\top$

8.3.3 Sherman-Morrison-Woodbury Pseudoinverse

For a case with a diagonal matrix $\mathbf{C}_e$ and $N_d \gg N_e$, we can use the Sherman-Morrison-Woodbury formula (A.9) to derive an efficient inversion strategy. In this case, we write

$$\mathbf{C} = \mathbf{S} \widehat{\mathbf{C}} \mathbf{S} = \mathbf{S} \left[\widehat{\Delta \mathbf{D}} \widehat{\Delta \mathbf{D}}^\top + \alpha \mathbf{I} \right] \mathbf{S}. \quad (8.29)$$

Using the Sherman-Morrison-Woodbury formula, we can write $\mathbf{C}^{-1}$ as

$$\mathbf{C}^{-1} = \frac{1}{\alpha} \left[\mathbf{I} - \widehat{\Delta \mathbf{D}} \left(\widehat{\Delta \mathbf{D}}^\top \widehat{\Delta \mathbf{D}} + \alpha \mathbf{I} \right)^{-1} \widehat{\Delta \mathbf{D}}^\top \right] \mathbf{S}^{-1}. \quad (8.30)$$

The matrix $\left(\widehat{\Delta \mathbf{D}}^\top \widehat{\Delta \mathbf{D}} + \alpha \mathbf{I} \right)$ has dimension $N_e \times N_e$. Therefore, if $N_e \ll N_d$ this is a convenient form to compute the inverse. Using TSVD, we write

$$\widehat{\Delta \mathbf{D}}^\top \widehat{\Delta \mathbf{D}} + \alpha \mathbf{I} = \mathbf{\Xi}_r \mathbf{\Lambda}_r \mathbf{\Xi}_r^\top. \quad (8.31)$$

The pseudoinverse of $\mathbf{C}$ becomes

$$\mathbf{C}^+ = \frac{1}{\alpha} \mathbf{S}^{-1} \left[\mathbf{I} - \left(\widehat{\Delta \mathbf{D}} \mathbf{\Xi}_r \right) \mathbf{\Lambda}_r^{-1} \left(\widehat{\Delta \mathbf{D}} \mathbf{\Xi}_r \right)^\top \right] \mathbf{S}^{-1}. \quad (8.32)$$

Pseudo-code 8.5 summarizes this procedure.

Pseudo-code 8.5: Sherman-Morrison-Woodbury pseudoinverse with diagonal $\mathbf{C}_e$

1. $\mathbf{S} = [\text{diag}(\mathbf{C}_e)]^{1/2}$ {Data-error standard deviations}
2. $\widehat{\Delta \mathbf{D}} = \mathbf{S}^{-1} \Delta \mathbf{D}$
3. $\mathbf{P} = \widehat{\Delta \mathbf{D}}^\top \widehat{\Delta \mathbf{D}} + \alpha \mathbf{I}$
4. $\mathbf{\Xi}_r \mathbf{\Lambda}_r \mathbf{\Xi}_r^\top = \mathbf{P}$ {TSVD}
5. $\mathbf{\Upsilon} = \alpha^{-1} \left[\mathbf{I} - \left(\widehat{\Delta \mathbf{D}} \mathbf{\Xi}_r \right) \mathbf{\Lambda}_r^{-1} \left(\widehat{\Delta \mathbf{D}} \mathbf{\Xi}_r \right)^\top \right]$
6. $\mathbf{C}^+ = \mathbf{S}^{-1} \mathbf{\Upsilon} \mathbf{S}^{-1}$

8.3.4 Truncation of Singular Values

The inversion methods described above involve truncating singular values. The most common approach for truncation is to determine the minimum number of singular values, N_r , such that

$$\frac{\sum_{i=1}^{N_r} \sigma_i}{\sum_{i=1}^{N_R} \sigma_i} \geq \xi, \quad (8.33)$$

where $\xi \in (0, 1]$ is referred to as the truncation energy. In most applications, high values like $\xi \geq 0.99$ are commonly used. However, in cases with a large number of redundant data points, using a lower value may be beneficial [78, 124].

Recently, Emerick and Neto [124] proposed using a method called “optimal hard thresholding” (OHT) [165] to determine the truncation level of $\widehat{\Delta \mathbf{D}}$ in subspace inversion. According to this method, a truncation threshold, τ_* , is computed as

$$\tau_* = \sqrt{2.858} \sigma_{\text{med}}, \quad (8.34)$$

where σ_{med} is the median singular value of $\widehat{\Delta \mathbf{D}}$.

8.3.4.1 Example: Truncation of Singular Values in the PUNQ-S3 Benchmark Problem

Table 8.2 presents the values of the data-mismatch objective function and normalized variance (NV) obtained by Emerick and Neto [124] for a modified version of the PUNQ-S3 benchmark problem [150]. The study utilized ES-MDA with subspace inversion and truncation based on the energy criterion ($\xi = 0.99$) and the OHT method. Fig. 8.1 displays the singular value spectrum, highlighting the number of singular values retained for each case. Fig. 8.2 illustrates the predicted water cut data for one well in the model. These results demonstrate that the OHT method retained fewer singular values, leading to a smaller reduction in the posterior variance of the ensemble without compromising the quality of the data match.

Table 8.2: Data-mismatch objective function and normalized variance for a modified version of the PUNQ-S3 problem with truncation of singular values based on the energy criterion $\xi = 0.99$ and OHT. Example 8.3.4.1.

Case	$\mathcal{O}_d(\mathbf{m})$ (mean $\pm$ std. dev.)	NV
$\xi = 0.99$	0.474 ± 0.015	0.230
OHT	0.494 ± 0.027	0.303

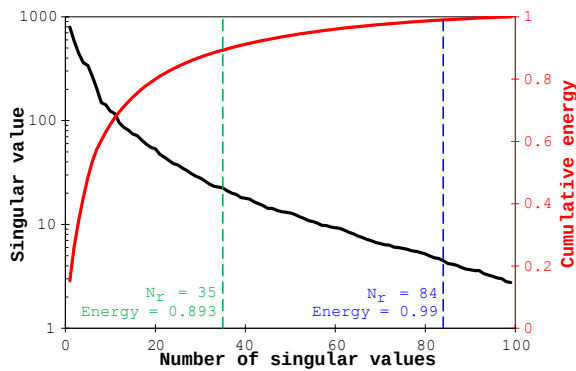


Fig. 8.1: Singular values of the matrix $\widehat{\Delta \mathbf{D}}$ (prior ensemble) for a modified version of the PUNQ-S3 problem. The green line represents the truncation level determined by the optimal hard threshold method, while the dashed line indicates the truncation level based on the energy criterion. Example 8.3.4.1. Reproduced from Emerick and Neto [124] with permission from Elsevier.

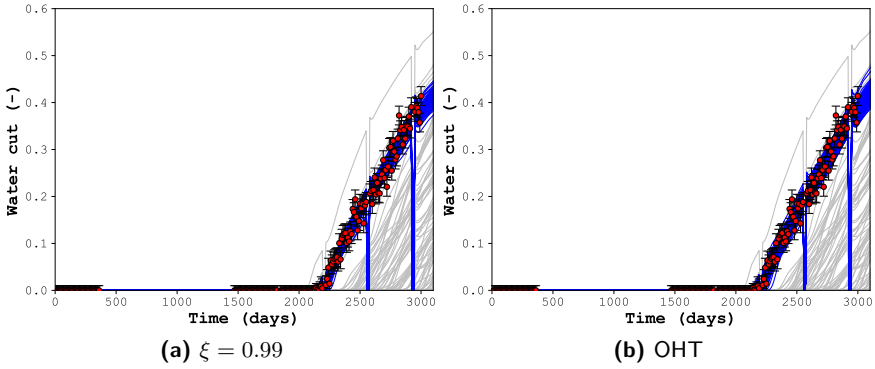


Fig. 8.2: Water cut of a well for a modified version of the PUNQ-S3 problem. Red dots represent the observed data points, with error bars corresponding to one standard deviation of the data error. The gray lines show the predicted data from the prior ensemble, while the blue lines depict the predictions from the posterior ensemble. Example 8.3.4.1. Reproduced from Emerick and Neto [124] with permission from Elsevier.

8.4 Matrix Multiplications

Regardless if we use pseudoinverse, subspace inversion or Sherman-Morrison-Woodbury pseudoinverse, the inverse of the matrix $\mathbf{C}$ can be written as

$$\mathbf{C}^{-1} \approx \mathbf{C}^+ = \mathbf{\Omega} \mathbf{\Upsilon} \mathbf{\Omega}^\top, \quad (8.35)$$

where $\mathbf{\Omega}$ is a $N_d \times N_r$ matrix and $\mathbf{\Upsilon}$ is a $N_r \times N_r$ diagonal matrix.

The definition of $\mathbf{\Omega}$ and $\mathbf{\Upsilon}$ depends on the inversion procedure used:

- **Pseudoinverse:**

$$\mathbf{\Omega} = \mathbf{S}^{-1} \mathbf{\Phi}_r \quad (8.36)$$

and

$$\mathbf{\Upsilon} = \mathbf{\Lambda}_r^{-1}. \quad (8.37)$$

- **Subspace inversion:**

$$\mathbf{\Omega} = \mathbf{S}^{-1} \mathbf{U}_r \mathbf{\Sigma}_r^{-1} \mathbf{\Psi} \quad (8.38)$$

and

$$\mathbf{\Upsilon} = (\mathbf{I} + \mathbf{\Gamma})^{-1}. \quad (8.39)$$

- **Sherman-Morrison-Woodbury Pseudoinverse:**

$$\mathbf{\Omega} = \mathbf{S}^{-1} \quad (8.40)$$

and

$$\mathbf{\Upsilon} = \alpha^{-1} \left[\mathbf{I} - \left(\widehat{\Delta \mathbf{D} \Xi_r} \right) \mathbf{\Lambda}_r^{-1} \left(\widehat{\Delta \mathbf{D} \Xi_r} \right)^\top \right]. \quad (8.41)$$

For the case without localization, Evensen [137] presents the following order for performing the multiplications to update the ensemble of models in $\mathbf{M}$:

$$\mathbf{\Omega}_1 = \mathbf{\Upsilon}\mathbf{\Omega}^\top, \quad (N_r \times N_d) \quad (8.42)$$

$$\mathbf{\Omega}_2 = \mathbf{\Omega}_1\mathbf{J}, \quad (N_r \times N_e) \quad (8.43)$$

$$\mathbf{\Omega}_3 = \mathbf{\Omega}\mathbf{\Omega}_2, \quad (N_d \times N_e) \quad (8.44)$$

$$\mathbf{\Omega}_4 = \Delta\mathbf{D}^\top\mathbf{\Omega}_3, \quad (N_e \times N_e) \quad (8.45)$$

and finally

$$\mathbf{M}_{\text{updated}} = \mathbf{M} + \Delta\mathbf{M}\mathbf{\Omega}_4. \quad (N_m \times N_e) \quad (8.46)$$

The resulting matrix dimensions are indicated alongside each multiplication to emphasize the sizes involved. This order of multiplication is particularly efficient because, typically, we have $N_m \gg N_e$. Moreover, these multiplications are optimized for the case where $N_d > N_e$, which is the typical situation for smoothers.

If we have $N_d \ll N_e$, it is better to introduce a slight modification in the last two multiplications. Instead of computing $\mathbf{\Omega}_4$, we compute

$$\mathbf{\Omega}_5 = \Delta\mathbf{M}\Delta\mathbf{D}^\top. \quad (N_m \times N_d) \quad (8.47)$$

Then, we update the ensemble using

$$\mathbf{M}_{\text{updated}} = \mathbf{M} + \mathbf{\Omega}_5\mathbf{\Omega}_3. \quad (N_m \times N_e) \quad (8.48)$$

For the case with Kalman gain localization, the best multiplication order is

$$\mathbf{\Omega}_1 = \mathbf{\Upsilon}\mathbf{\Omega}^\top, \quad (N_r \times N_d) \quad (8.49)$$

$$\mathbf{\Omega}_6 = \Delta\mathbf{D}^\top\mathbf{\Omega}, \quad (N_e \times N_d) \quad (8.50)$$

$$\mathbf{\Omega}_7 = \mathbf{\Omega}_6\mathbf{\Omega}_1, \quad (N_e \times N_d) \quad (8.51)$$

$$\mathbf{K} = \Delta\mathbf{M}\mathbf{\Omega}_7, \quad (N_m \times N_d) \quad (8.52)$$

$$\mathbf{K}_{\text{loc}} = \mathbf{R}_{\text{md}} \circ \mathbf{K}, \quad (N_m \times N_d) \quad (8.53)$$

$$\mathbf{\Omega}_8 = \mathbf{K}_{\text{loc}}\mathbf{J}, \quad (N_m \times N_e) \quad (8.54)$$

and finally

$$\mathbf{M}_{\text{updated}} = \mathbf{M} + \mathbf{\Omega}_8. \quad (N_m \times N_e) \quad (8.55)$$

However, when N_d is very large, the main challenge arises from the Kalman gain matrix, which has dimensions $N_m \times N_d$. In such situations, an alternative approach is to update the matrix $\mathbf{M}$ in groups of rows. For each update, a smaller Kalman gain matrix is computed. In the limit, each row of $\mathbf{M}$ can be updated independently, significantly reducing the memory requirements of the analysis. Not only does this approach save memory by avoiding the allocation of the entire Kalman gain matrix, but it can also be

accelerated through parallel processing, with each group of rows of $\mathbf{M}$ being updated independently.

8.4.1 Example: Parallelization of the Analysis

Compared to the time required to execute the ensemble of reservoir simulations, the time required for the computation of analysis is relatively insignificant, especially if the number of model parameters and data points is not too large. Typical reservoir data assimilation problems involving production data have $O(10^2)$ to $O(10^7)$ model parameters and $O(10^2)$ to $O(10^3)$ data points. In these cases, a standard implementation of the analysis is sufficient, particularly if up-to-date linear algebra packages are used to perform matrix multiplications, taking advantage of the multiple CPUs available in modern workstations.

However, this scenario changes significantly when the number of data points increases, such as in cases involving 4D seismic data. Fortunately, the calculations required for the analysis are highly parallelizable. To illustrate that, consider the data assimilation exercise presented by Exterkoetter et al. [146], with a reservoir model discretized in $150 \times 150 \times 50$ gridblocks operating with 36 production wells and 25 water injection placed in five-spot patterns as illustrated in Fig. 8.3a. The data assimilation parameters are the porosity and log-permeability of each gridblock, resulting in $N_m = 2,250,000$. The observed data correspond to a 4D seismic dataset represented as the water saturation changes in every gridblock after 10 years of production, which corresponds to $N_d = 1,125,000$ (Fig. 8.3b).

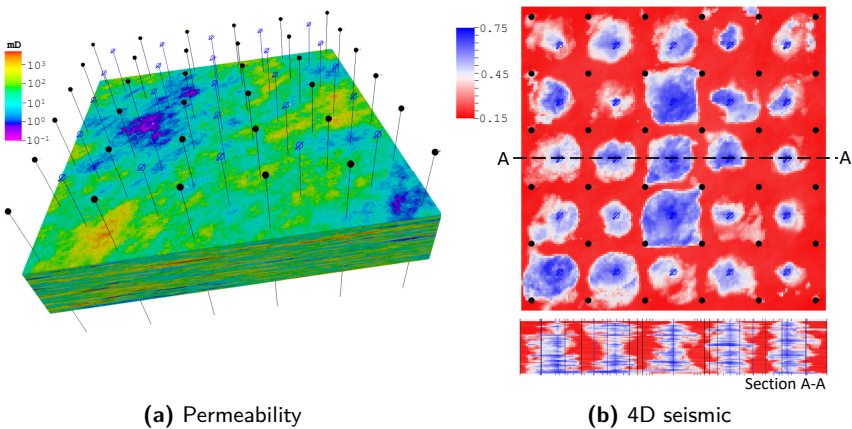


Fig. 8.3: Hypothetical data assimilation problem used to test the computational performance of the analysis in CPU and GPU implementations [146]. Example 8.4.1.

Exterkoetter et al. [146] compared the computational time required for the analysis step of the problem using CPU and GPU implementations. Both implementations utilized a distance-based Kalman gain localization with subspace inversion, as described in Section 8.3.2, and a diagonal $\mathbf{C}_e$. The ensemble size was 200. The CPU implementation was tested on an i7-9700K CPU with eight cores, while the GPU implementation used an NVIDIA RTX 3090 with 10,496 cores running CUDA 10.1. Table 8.3 shows the computational time required by both implementations. Notably, the standard CPU implementation took more than 2.3 days to complete the analysis calculation. This computational time was significantly reduced by leveraging the parallelism offered by GPUs. The time reduction is primarily due to the acceleration in matrix multiplications and the calculation of the localization coefficients. Kalman gain localization requires computing the distance between each data point and each model parameter, resulting in 1.125 million data points times 1.125 million gridblocks.

Table 8.3: Computational time of the analysis for a data assimilation problem with $N_m = 2,250,000$, $N_d = 1,125,000$ and $N_e = 200$.

Case	Time (hours)
CPU	57.15
GPU	4.19

Practical Aspects and Field Examples

Abstract: *This chapter discusses key practical considerations for applying iterative ensemble smoothers in reservoir data assimilation, offering guidelines for researchers and practitioners. The content is organized into three main sections: parametrization, data handling, and result evaluation. The chapter concludes with three field examples that demonstrate data assimilation in petroleum reservoirs.*

9.1 Introduction

Understanding the practical aspects of real-world data assimilation applications is essential for bridging the gap between theoretical models and their implementation in the field.

This chapter discusses key practical considerations in applying iterative ensemble smoothers for reservoir data assimilation and offers guidelines for researchers and practitioners. The discussion is organized into three main parts: parametrization, data handling, and evaluation of results. The chapter concludes with the presentation of three field examples that illustrate data assimilation in petroleum reservoirs.

9.2 Parametrization

One of the most notable differences between ensemble and optimization-based methods for data assimilation is the number of parameters each method can handle. In ensemble-based methods, the computational cost

is primarily determined by the number of model realizations. Conversely, in optimization-based methods, especially the ones not relying on derivative calculation, the computational cost is driven by the number of parameters. In fact, much of the literature on reservoir data assimilation using optimization-based methods focuses on developing efficient parametrizations to reduce the dimension of the parameter space, thereby enabling more efficient optimizations [331].

In contrast, ensemble methods are well-suited for handling problems with a vast number of model parameters. For example, these methods commonly allow reservoir rock properties to vary on a gridblock-by-gridblock basis, resulting in situations with easily $O(10^7)$ uncertain variables. The key element of these methods lies in shifting the search for solutions from the “model space” to the “ensemble space.” In fact, the size of the ensemble can be interpreted as the number of “effective parameters.” Recall from Section 7.2 that in cases without localization, the posterior ensemble is a linear combination of the prior ensemble members, meaning that the data assimilation has N_e effective parameters (coefficients of the linear combinations). Although localization increases the dimensionality of this space, the overall performance of data assimilation still heavily depends on the prior ensemble. While the size of the prior ensemble has an impact, practical applications typically limit the number of ensemble members to $O(10^2)$. Therefore, more crucial than the ensemble size is the quality of the prior models in terms of accurately capturing the reservoir’s behavior and adequately spanning the prior uncertainty.

Section 3.7 introduced parametrization with an emphasis on finding a reduced representation of the parameter space to enable efficient solution finding. In this section, we shift focus and discuss parametrization from the perspective of selecting relevant reservoir uncertainty properties to ensure a robust prior ensemble. A key principle here is to avoid overconfidence in parameter selection. Neglecting a parameter implies assuming its value is known with certainty, which can lead to significant issues. Overlooking relevant parameters often forces other parameters to compensate, resulting in values that fall outside their expected ranges [330] and possibly introducing bias in the model forecasts.

Before delving into the main types of parameters used in reservoir data assimilation, we will briefly discuss two elements: the use of model realizations versus scenarios and the use of integrated workflows for creating and updating reservoir models.

9.2.1 Realizations versus Scenarios to Describe Uncertainty

The current practice in reservoir modeling involves combining scenarios and stochastic realizations to describe uncertainty. Scenarios are constructed to capture higher-level uncertainties, typically modeling only a few scenarios

(e.g., pessimistic, base, optimistic). Realizations, on the other hand, are generated using geostatistics to populate models with rock properties consistent with the data. Regarding these processes, Ringrose and Bentley [365] identifies three differing approaches to uncertainty handling in reservoir modeling:

1. **Best guess approach:** The preferred model is chosen as the best guess and uncertainty is represented by adding a range around the best guess.
2. **Multiple stochastic approach:** A large number of models are probabilistically generated by geostatistical simulation.
3. **Multiple deterministic approach:** A smaller number of models are built, each one reflecting a complete reservoir concept.

According to the authors, the “best-guess approach” puts faith in the ability of an individual or team to make a reasonably precise judgment. This process tends to generate the so-called anchoring effect¹—“*once anchored, the adjustment away from the initial best guess is too limited as the outcome is overly influenced by the anchor point*” [445].

The “multiple stochastic approach” refers to the use of realizations, while the “multiple deterministic approach” refers to scenarios. Ringrose and Bentley [365] argue that the sole use of realizations to describe uncertainty may not be able to escape from anchoring around the “most likely” probabilistic outcome, highlighting the importance of scenarios in capturing higher-level uncertainties to represent plausible modeling hypotheses.

Since ensemble data assimilation methods rely on Gaussian assumptions for the prior, they are typically more effective at updating realizations. Mixing scenarios and realizations within a single ensemble may not yield optimal performance, particularly when the scenarios are highly distinct. In such cases, the formulation relying on the first two moments of the distribution (mean and covariance) becomes insufficient for updating the combined scenario-realization ensemble. This often results in the inability to generate plausible posterior models, with the components of the ensemble lacking the required geological realism or failing to match observations.

A more pragmatic approach involves generating an ensemble of realizations for each discrete scenario and performing separate data assimilation for each. During this process, certain scenarios may prove unfeasible and need to be discarded; a procedure sometimes referred to as prior model falsification [53].

However, this strategy is not without challenges. One significant issue is determining how to properly weigh the posterior scenarios and realizations for uncertainty evaluation in reservoir performance predictions. Methods for weighting posterior models, such as importance sampling [135], Bayesian model averaging [193], and Bayesian stacking [472], all suffer from the curse of dimensionality [4, 5].

¹ Essentially, the anchoring effect refers to our tendency to rely too heavily on the information offered, introducing a bias in the model-construction process. In practice, we may neglect relevant scenarios that were not revealed at the time of the model construction because they did not fit the idea of the best explanation.

9.2.2 Integrated Workflows

Integrated workflows for model updating have gained significant attention due to their automation of the reservoir model generation process. This integration combines geological modeling tools and reservoir simulators within workflows orchestrated by dedicated software that samples and updates parameters. Known as the “big-loop approach” [461], this process aims to make the model update process repeatable, thereby reducing time and inconsistencies. A notable example is Equinor’s Fast Model Update (FMU) [476, 401], which uses the open-source software **ERT**².

Integrated workflows offer substantial benefits for reservoirs with sparse data and significant uncertainties. In such cases, these workflows enable the evaluation of diverse modeling scenarios by updating high-level uncertainty parameters that have a global impact on the models. However, reservoirs with numerous wells and extensive operational history often require local updates in the model realizations to achieve acceptable data matches. In these cases, the big-loop approach with global parameters can be used to discard unfeasible modeling scenarios and generate consistent ensembles for plausible ones. These ensembles can then serve as prior models for data assimilation, ensuring that data is honored on a well-by-well basis.

9.2.3 Main Types of Reservoir Parameters

Internally, ensemble methods do not distinguish between different types of reservoir properties used as model parameters. This flexibility is one of their key advantages, allowing them to be easily coupled with complex forward models and parametrizations without the need for specific implementations to calculate partial derivatives, unlike adjoint-based methods.

Nevertheless, it is useful to categorize reservoir parameters into three main types: grid parameters, scalar parameters, and structural parameters. This categorization facilitates discussion, as each type may require different strategies for prior ensemble generation and localization.

9.2.3.1 Grid Parameters

Grid parameters encompass rock properties such as porosity and permeability, including special parametrizations to represent facies. These parameters are characterized by the fact that they vary on a gridblock-by-gridblock basis. Therefore, the number of grid parameters is in the same order as the number of active gridblocks in the model.

² **ERT** (Ensemble Reservoir Tool) is a GPL software available at <https://github.com/equinor/ert>

The prior ensemble of grid parameters is almost invariably generated using geostatistics. During data assimilation, the lower and upper limits on the parameter values are typically enforced through truncation. These limits can be established either globally, regionally, or even on a gridblock-by-gridblock basis. Setting limits on a gridblock-by-gridblock basis offers more detailed control over property values. Fig. 9.1 shows an example where using gridblock-by-gridblock limits resulted in a better estimation of the final porosity distribution in a model, preventing the occurrence of high-porosity values on the left side of the model. However, overly strict limits can constrain changes in the model. As a practical consequence, this may prevent the identification of areas with poor parametrization, limiting valuable feedback for improving the prior model [223].

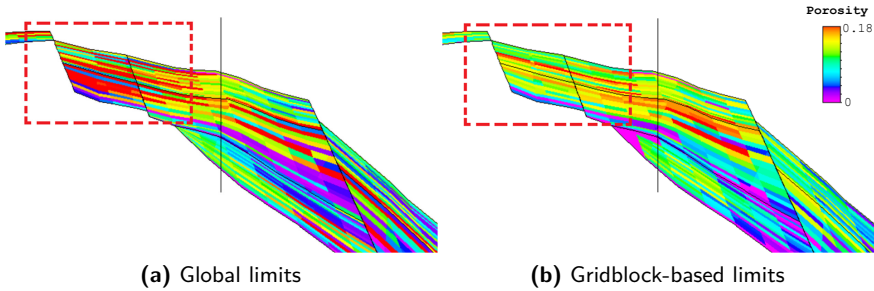


Fig. 9.1: Example with two posterior realizations of porosity obtained with global and gridblock-based limits after data assimilation with ES-MDA for a field in the Santos Basin.

One interesting aspect of ensemble methods is the fact that they use the prior ensemble to construct the uncertainty space. This means that if a parameter value does not vary in the ensemble, it implies no associated uncertainty. A practical consequence is that “hard data” used in the prior geostatistical modeling are retained during data assimilation. Conversely, underestimating prior uncertainty will invariably lead to underestimation in the posterior.

Data assimilation with grid parameters invariably requires the use of a localization strategy. Currently, distance-based localization is the most effective approach. The choice of the “optimal” localization function depends on the specific problem, varying with the type of data, type of parameter, and ensemble size. Nonetheless, selecting reasonable localization regions is not overly challenging. The compact function of Gaspari and Cohn [162] is the most commonly used correlation function, and the selection of the localization radius (critical length) is the most relevant aspect.

In general, underestimating the size of the localization region is more detrimental than overestimating it. In reservoir applications, it is recommended to align the localization with the model’s anisotropic direction and

choose regions larger than the prior model ranges [77, 126]. Additionally, the localization should encompass distances greater than those between injector-producer pairs to cover the entire sensitivity region.

Porosity

Porosity is a primary grid parameter used in reservoir data assimilation. It directly impacts reservoir volume and significantly influences overall reservoir pressure. Near wells, porosity is generally well-characterized due to the availability of well-logging data (hard data).

Permeability

Permeability is the most crucial and widely used grid parameter in reservoir data assimilation, as it defines flow patterns within the reservoir and influences both pressure and saturation matching. However, permeability data are scarcer than porosity data, leading to greater uncertainties. Additionally, significant discrepancies often exist between permeability values measured in laboratory tests on rock cores and those estimated in the field through pressure transient analysis, sometimes differing by orders of magnitude. Typically, permeability is correlated with porosity, and these correlations are often used to construct well profiles that serve as conditioning data in geological models. Since permeability frequently exhibits a log-normal distribution, it is advisable to adjust the log-transformed values during data assimilation.

Net-to-Gross Ratio

The net-to-gross ratio (NTG) is sometimes used as a grid parameter in data assimilation. NTG is a value between zero and one, indicating the fraction of reservoir rock in each gridblock of the model. NTG realizations are typically generated from the upscaling process because, in the fine-scale geological model, NTG is often modeled with binary values (reservoir/non-reservoir). In the coarse-scale flow simulation model, NTG serves as a simplified representation of facies. NTG impacts the porous volume and horizontal transmissibility in reservoir simulation, thereby redundantly affecting porosity and horizontal permeability during data assimilation.

Vertical Transmissibility Multipliers

Vertical transmissibility multipliers are also very common to represent vertical barriers to flow, such as cementation or thin shale layers that have

been “coarsened” during upscaling [345]. These barriers are often detected in well logs, but their spatial distribution is uncertain, especially because they are difficult to detect using seismic data. Similarly to permeability, adjusting the log-transformed values of transmissibility multipliers during data assimilation is recommended.

Facies

Facies are categorical variables that represent different rock types, posing a challenge for their update using ensemble methods. To address this complexity, facies realizations are typically re-parameterized into variables that assume continuous values, ideally following distributions close to Gaussian. Currently, one of the most widely adopted approaches for re-parameterizing facies in data assimilation with ensemble methods is the truncated pluri-gaussian (TPG) method [157, 17, 35].

In the TPG method, latent grid parameters used for generating facies realizations are updated simultaneously with other reservoir rock properties. During prior modeling, distinct geostatistical properties are assigned to each facies to characterize rock attributes such as porosity and permeability. Consequently, updating parameters associated with each facies type becomes necessary. The final realizations of the simulation are synthesized by combining the property values attributed to each facies, a technique often referred to as the “cookie-cutter approach” [104].

Fig. 9.2 shows an example of facies and permeability realization for an oil-field in the Campos Basin. The facies realization was generated using TPG by truncating two Gaussian random realizations, $\mathbf{z}_1$ and $\mathbf{z}_2$. The final model contains four facies, with only two (R1 and R2) assumed to contribute to fluid flow. The permeability of these facies was modeled using SGS, and the final permeability model was constructed using the cookie-cutter approach. During data assimilation, the latent Gaussians, $\mathbf{z}_1$ and $\mathbf{z}_2$, and the permeability of facies R1 and R2 are updated as reservoir parameters.

9.2.3.2 Scalar Parameters

Scalar parameters are single-value parameters that have a broad influence on the model, either globally or regionally. Examples encompass parameters defining relative permeability and capillary pressure tables, transmissibility multipliers, rock compressibility, parameters for aquifer influx modeling, and initial fluid contacts.

The prior ensemble of scalar parameters is typically constructed by independently sampling univariate distributions, assuming uncorrelated prior uncertainty. Commonly used distributions include normal and log-normal. However, practitioners often favor distributions with compact support, such

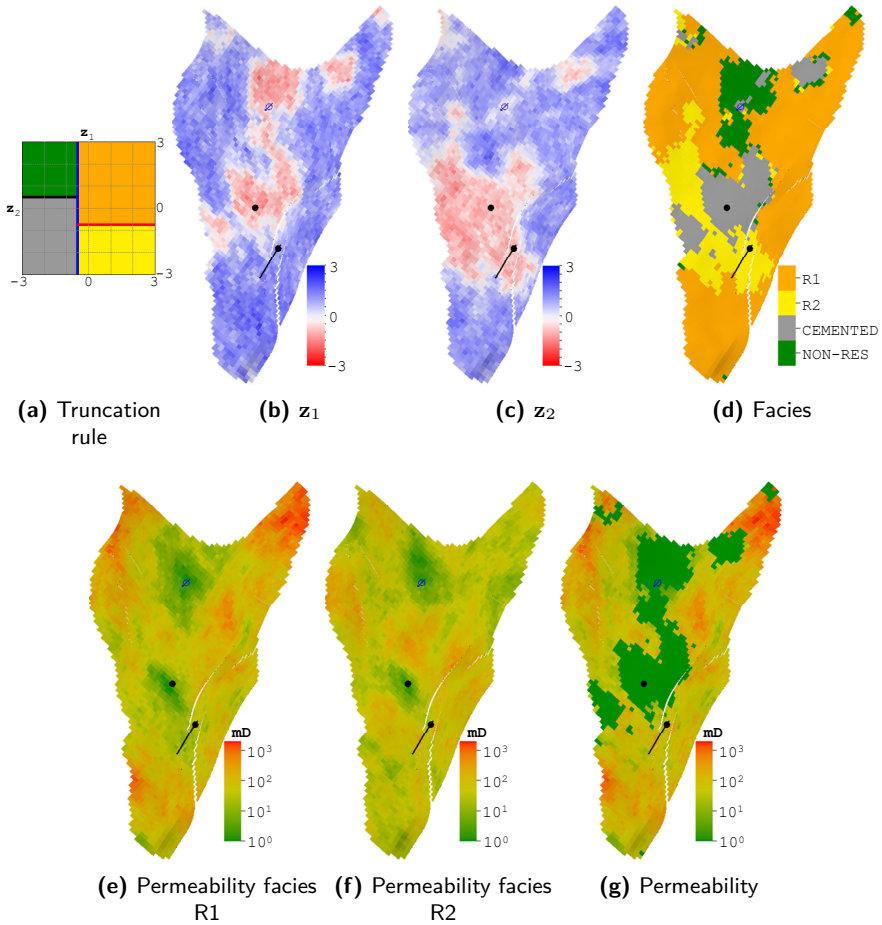


Fig. 9.2: Facies realization generated using TPG for an oilfield in the Campos Basin. Panel (a) shows the truncation rule. Panels (b) and (c) display the Gaussian random fields, while panel (d) presents the resulting facies realization. Panels (e) and (f) depict the permeability realizations for facies R1 and R2, respectively, and panel (g) shows the final permeability realization.

as uniform, triangular, and PERT distributions (see Appendix C, Section C.1.18).

When dealing with correlated prior uncertainty, sampling can be achieved using the Cholesky decomposition of the prior covariance matrix, assuming all scalar parameters follow normal distributions. However, if parameters have different distributions, more advanced sampling strategies are needed, such as employing copulas [318].

Selecting coefficients for the localization matrix for scalar parameters poses a greater challenge because these parameters lack an associated spa-

tial location, making it impossible to calculate spatial distances between parameters and data points. Despite this limitation, it is still feasible to introduce a degree of “localization” based on field knowledge. For instance, one approach involves setting the coefficients of the localization matrix to zero for parameters identified as insensitive to the data. Alternatively, the non-distance-dependent localization schemes discussed in Section 7.5.4 can be employed.

Relative Permeability

Relative permeability curves govern the displacement speed of different phases within the reservoir, making the associated parameters highly utilized in reservoir data assimilation problems with multiphase flow. These curves are initially derived from laboratory displacement tests on core samples. In these tests, determining the endpoints is generally more reliable than defining the curve shapes. A key consideration for using relative permeability curves in simulations is their dependence on the model scale; identical curves can lead to different water and gas breakthrough times based on the dimensions of the model gridblocks. To address this, the curves are often adjusted during upscaling to account for the absence of heterogeneities in coarse simulation grids and to compensate for numerical dispersion effects in multiphase fluid flow. These adjusted curves are sometimes called pseudo-curves [29, 116]. Consequently, uncertainty in relative permeability curves is typically much larger than the variability observed in laboratory measurements, especially in fields with limited production history where fluid breakthroughs have not been fully observed.

Relative permeability curves are provided as tables in most reservoir simulators. For more effective data assimilation, these curves are re-parameterized using power-law models (Fig. 9.3). The most common model used is the Brooks-Corey model, which can be expressed as

$$\kappa_{ri} = \kappa_{ri,\max} s_{ni}^{e_i} \quad (9.1)$$

with

$$s_{ni} = \frac{s_i - s_{i,\min}}{s_{i,\max} - s_{i,\min}}, \quad (9.2)$$

where κ_{ri} is the relative permeability of the phase i at saturation s_i . The exponent e_i determines the shape of the curve. s_{ni} is the normalized saturation with $s_{i,\min}$ and $s_{i,\max}$ denoting the minimum and maximum phase saturation, respectively. For example, the relative permeability for the water phase in an oil-water table we have $s_{i,\min} = s_{w,\text{cr}}$ and $s_{i,\max} = 1 - s_{o,\text{rw}}$, where $s_{w,\text{cr}}$ is the critical water saturation and $s_{o,\text{rw}}$ is the residual oil saturation.

Another frequently used model to parametrize relative permeabilities is the LET model [269]. Unlike the Corey, the LET model provides more de-

degrees of freedom, enabling the modeling of curves with varying concavity. The LET curves can be expressed as

$$\kappa_{ri} = \kappa_{ri,\max} \frac{s_{ni}^{L_i}}{s_{ni}^{L_i} + E_i (1 - s_{ni})^{T_i}}. \quad (9.3)$$

The parameters L_i , E_i , and T_i do not have direct physical interpretations. Parameter L_i governs the lower segment of the curve, akin to the Corey exponent, while T_i influences the upper segment. Parameter E_i determines the position of the curvature transition. A value of $E_i = 1$ signifies a neutral point, where the curve is influenced by L_i and T_i . Increasing E_i shifts the transition point towards the upper curve segment. According to Lomeland et al. [269], the conditions are $L_i \geq 1$, $E_i > 0$, and $T_i \geq 0.5$.

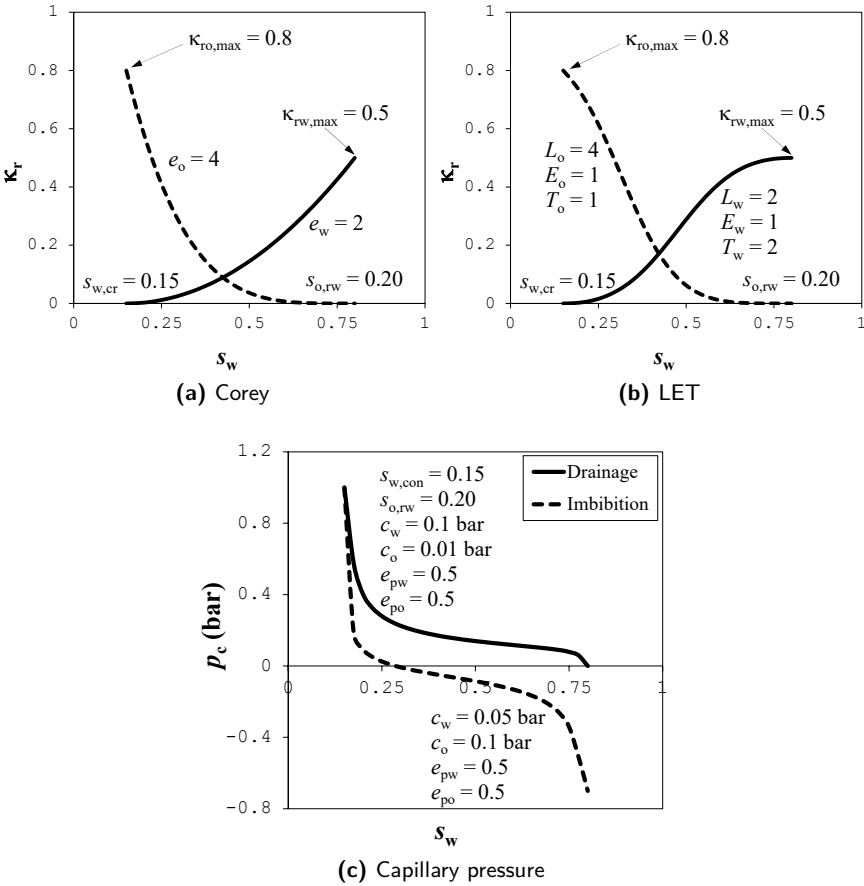


Fig. 9.3: Typical power-law models used to re-parametrize relative permeability and capillary pressure curves in reservoir models.

Capillary Pressure

Like relative permeability, capillary pressure in reservoir simulation models is typically represented in tabular form, requiring re-parametrization. One alternative, in this case, is to use the expressions proposed by Skjaeveland et al. [398]:

$$p_c = \frac{c_1}{\left(\frac{s_1 - s_{1,\min}}{1 - s_{1,\min}}\right)^{e_{p1}}} - \frac{c_2}{\left(\frac{s_2 - s_{2,\min}}{1 - s_{2,\min}}\right)^{e_{p2}}}, \quad (9.4)$$

where p_c is the capillary pressure defined as the difference between the pressure in the non-wetting phase, p_2 , and the pressure in the wetting phase, p_1 . For example, for a oil-water system with water as preferential wetting phase, we write $s_1 = s_w$ (water saturation), $s_2 = s_o$ (oil saturation), $s_{1,\min} = s_{w,\text{con}}$ (connate water saturation), and $s_{2,\min} = s_{o,\text{rw}}$ (residual oil saturation). The parameters c_1 , c_2 , e_{p1} , and e_{p2} control the shape of the curves. Fig. 9.3c shows an example of drainage and imbibition curves generated with Eq. (9.4).

One important aspect of the parametrization is the spatial distribution of relative permeability and capillary pressure properties in the model. There are several alternatives, but in practice, the most common approaches include dividing the model into fixed regions, such as partitioning the reservoir according to geological formations. For each region, a different parametrization of relative permeability and capillary pressure is adopted. Alternatively, relative permeability and capillary pressure can be assigned to each reservoir facies. A third option, often used in practical applications, is to divide the reservoir into “flow units,” where the goal is to group rock types with similar flow behavior. This is typically done based on porosity and permeability values, using methods like the flow zone indicator (FZI) [12]. One advantage of using facies or flow units is the ability to update these regions during data assimilation.

Faults

Scalar transmissibility multipliers across faults and between blocks play a crucial role in pressure and saturation matching in reservoir modeling. Fault positions are usually identified via seismic interpretation and remain static during data assimilation. These parameters consist of a single multiplier per fault between zero and one. Initializing these values with a log-normal distribution is often preferred, with adjustments made to the logarithm of the multiplier during data assimilation.

Although fault transmissibility multipliers are commonly treated as scalar parameters, they can alternatively be represented as grid parameters. For example, Emerick [123] presented a data assimilation example in which fault transmissibility multipliers were handled as grid parameters. This approach

improves flexibility by allowing multipliers to vary along faults and simplifying the implementation of distance-based localization.

Discrete Parameters

Ensemble methods are designed for real-valued parameters, and using parameters with discrete distributions is generally not recommended. However, practitioners sometimes employ them. In such cases, an alternative approach involves applying a normal scores transformation (see Fig. 9.4) to the discrete distribution. The transformed parameters are then updated during data assimilation. It is essential to ensure that the discrete parameter values are ordered to maintain a monotonic effect on the model predictions.

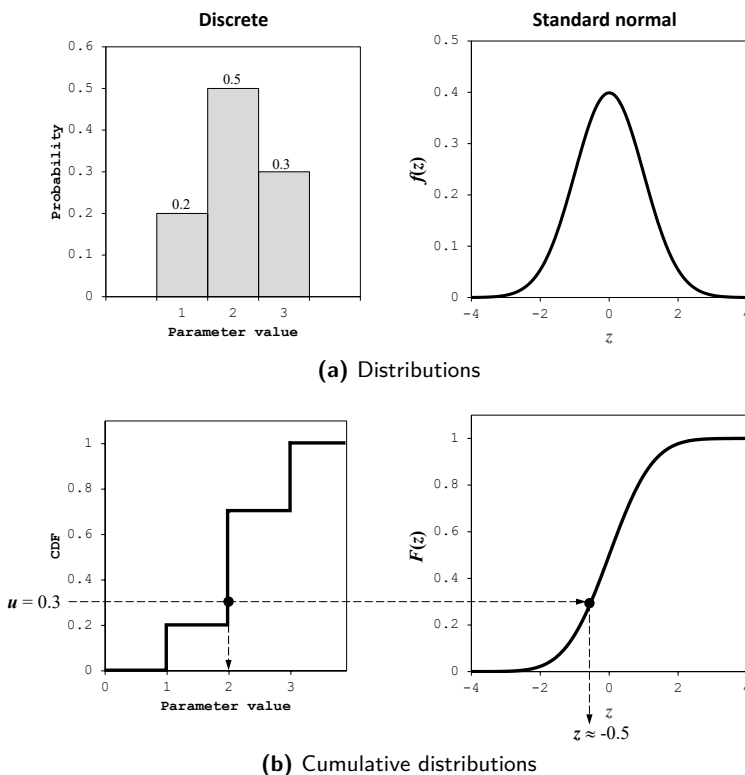


Fig. 9.4: Illustration of normal transformation and sampling of discrete parameters. During data assimilation, the normal-transformed values, z , are updated and the corresponding discrete values are obtained using the CDF.

9.2.4 Structural Parameters

Structural parameters determine the reservoir's geometry, thereby shaping the simulation model grid. Often, these parameters are relevant in new fields with limited wells and production history. The external and internal geometry of the reservoir can greatly influence flow behavior and the initial fluid volume.

One of the earliest studies to incorporate structural parameters into ensemble data assimilation was conducted by Seiler et al. [389], where they updated the top and bottom reservoir surfaces and deformed a corner-point grid to align with the updated surfaces. Today, commercial geomodeling software enables the specification of uncertainty in the input parameters of the structural model and the automatic reconstruction of the reservoir grid. These tools make it feasible to account for structural uncertainty in practical applications.

9.2.4.1 Example: Uncertainty in the Fault Position

Fig. 9.5 shows reservoir grids generated in a synthetic data assimilation exercise, where fault positions are estimated using data from a producer-injector well pair. The integration of geological modeling software with an ES-MDA-based data assimilation tool allows for grid updates during the process, leading to a significant improvement in the data match, as demonstrated in Fig. 9.6.

9.3 Observed Data

9.3.1 Production Data

Production data correspond to dynamic measurements gathered from production and injection wells during field operations, typically including fluid flow rates and pressure readings. These datasets are characterized by long time series. The use of production data in assimilation processes requires proper consideration of three interdependent aspects: data selection, frequency, and error covariance.

Data selection depends on the specific problem and should be guided by factors such as data acquisition methods and model limitations. A clear recommendation is to avoid unreliable data points (outliers). One straightforward method to identify outliers is to compare the mean innovation (the difference between the observed value and the ensemble mean prediction) to the combined standard deviations of predicted data, σ_d , and data-error, σ_e .

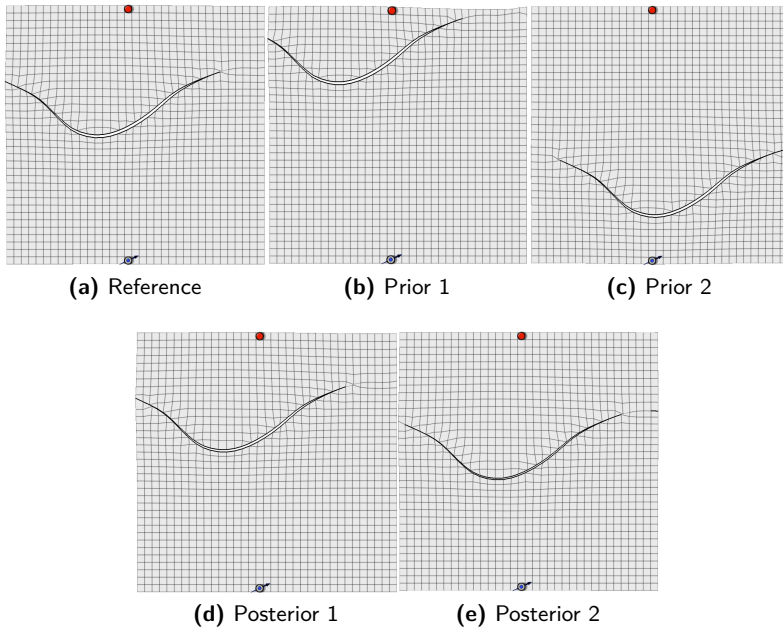


Fig. 9.5: Reservoir grids generated with varying fault positions. The red circle marks the location of the oil-producing well, while the blue circle marks the location of the gas injection well. Example 9.2.4.1.

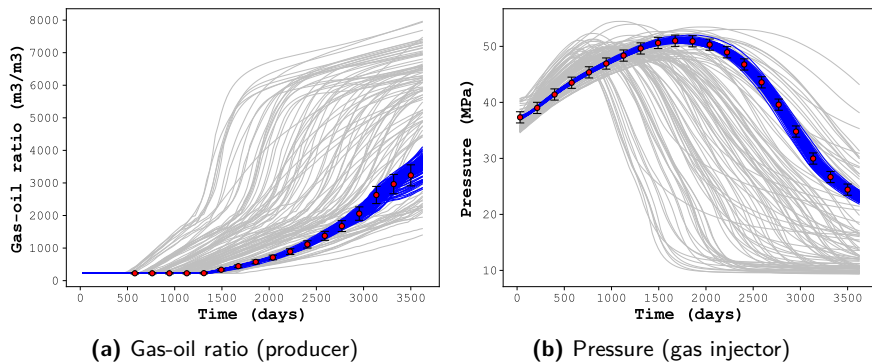


Fig. 9.6: Gas-oil ratio and bottom-hole pressure for a synthetic reservoir problem with uncertainty in the fault position. Red circles represent observed data, while gray and blue lines show predicted data from the prior and posterior ensembles, respectively. The error bars in the observed data points correspond to one standard deviation of the data error. Example 9.2.4.1.

If the innovation is significantly large compared to the sum of these standard deviations, the data point can be considered an outlier:

$$\text{if } \left| d_{\text{obs},i} - \overline{g_i(\mathbf{m})} \right| > 3(\sigma_{d,i} + \sigma_{e,i}), \text{ then } d_{\text{obs},i} \text{ is labeled as "outlier."} \quad (9.5)$$

Note that data points meeting the previous condition are not necessarily outliers (in the sense of being biased or severely corrupted by noise). They may also fall under this condition because the ensemble of models fails to capture the system's physical behavior, indicating model deficiency (model error). In both cases, removing these data points from the study may be beneficial. Alternatively, instead of removing the data point, its weight can be reduced by inflating the corresponding error standard deviation. One option is to set:

$$\sigma_{e,i} = \frac{1}{\gamma} \left| d_{\text{obs},i} - \overline{g_i(\mathbf{m})} \right| - \sigma_{d,i}, \quad \gamma \in [1, 3]. \quad (9.6)$$

In petroleum reservoirs, production data are typically reported at high frequencies, often on a daily basis. However, much of this data is derived by distributing the total production among wells using allocation tables, which are updated less frequently during periodic separation tests. As a result, the high frequency of production data may not provide significant additional information.

Another aspect is the presence of redundant information in long time series. Consecutive data points are expected to provide similar information content because they are correlated due to the system's dynamic behavior. It is important to note, however, that this type of correlation differs from data-error correlation, which is handled by the matrix $\mathbf{C}_e$. Redundant information in the data is managed by the matrix $\tilde{\mathbf{C}}_{\text{dd}}$ in the ensemble smoother formulation. Unlike $\mathbf{C}_e$, which is predefined, $\tilde{\mathbf{C}}_{\text{dd}}$ is calculated based on ensemble predictions and, therefore, may be affected by sampling errors. Emerick and Neto [124] demonstrated through a series of test problems that reducing the frequency of production data can help mitigate variance loss in this case without compromising data-match quality.

The selection of the data-error covariance, $\mathbf{C}_e$, greatly influences the performance of data assimilation, as it determines the weight of each observation and the balance between the prior model and the data. Underestimating $\mathbf{C}_e$ leads to excessively strong updates in model parameters, while overestimating it prevents effective use of all available information. Unfortunately, choosing an appropriate $\mathbf{C}_e$ is challenging. Data often contain more than just random noise, and systematic errors in data acquisition are hard to characterize and may be unknown to users. Moreover, $\mathbf{C}_e$ is frequently used to address forward modeling deficiencies [124].

Usually, we assume independence in production data errors, resulting in a diagonal $\mathbf{C}_e$. However, this assumption is often incorrect. Production data is usually derived from rate allocation based on infrequent separator

tests, which can lead to time-correlated measurement errors [143]. Additionally, model errors often exhibit time and space correlations. Ideally, a non-diagonal $\mathbf{C_e}$ should be used in these cases, as ignoring the off-diagonal terms leads to overconfidence in the data. However, estimating these correlations for proper $\mathbf{C_e}$ specification is challenging, and handling a non-diagonal $\mathbf{C_e}$ becomes computationally expensive if N_d is large. The standard alternative is to assume a diagonal $\mathbf{C_e}$ and inflate the variances to downweight the data.

The selection of data-error variance depends on the specific problem. In the absence of additional information, however, the values in Table 9.1 are commonly used in practical applications.

Table 9.1: Typical values for data-error standard deviation of production data

Data	Standard deviation
Oil rate	5–10% of the data value (min. 10 m ³ /day)
Water rate	10–15% of the data value (min. 10 m ³ /day)
Water cut	10–15% of the data value (min. 0.01)
GOR	15–20% of the data value
Pressure (flowing)	5–10 bars
Pressure (static)	1–5 bars

9.3.1.1 Example: Water Production Data

In petroleum reservoirs, especially in fields with water injection or strong aquifer influence, the volume of produced water is one of the most commonly used data for assimilation. Reservoir engineers often utilize the water cut as observed data, representing the ratio between water production and total liquid production rates. Other options include using the water production rate or cumulative water production. Fig. 9.7 shows an example of water production for a well in an onshore oilfield, presented in terms of water cut (Fig. 9.7a), water rate (Fig. 9.7b), and cumulative water production (Fig. 9.7c). This figure also shows the predicted data from a model operated at a specified oil rate. From the viewpoint of information content, all three representations of water production data contain the same information. However, a visual inspection of the plots may lead to different conclusions regarding the quality of the data match, even though the predictions are from the same model. The water production plot gives the impression of a better data match compared to the water cut and cumulative production plots. The water cut data exhibit a more “erratic” behavior, with significant short-term variations highlighting mismatches with the predicted water cut. The cumulative water data, on the other hand, follow a monotonically increasing curve that smooths out these oscillations. Nonetheless, the initial

mismatch observed in water production propagates through the cumulative data.

Although using cumulative production data may seem attractive for data assimilation due to the continuous nature of the curves, in practice, it requires accounting for correlated errors when defining the matrix $\mathbf{C}_e$.

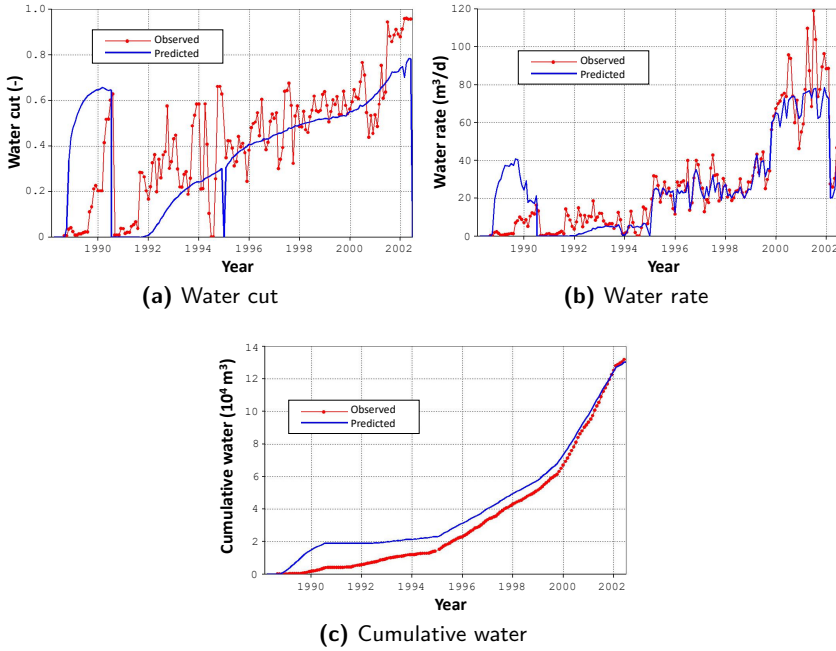


Fig. 9.7: Example of water production data for a well in an onshore oilfield. All panels show the same data but with different formats and the prediction from a model. Example 9.3.1.1.

9.3.1.2 Example: Well Pressure Data Sensitivity

Well-pressure measurements acquired with permanent downhole gauges (PDGs) are important sources of information for reservoir data assimilation. These data are typically divided into flowing and static pressures. Flowing pressure refers to the pressure measured while a well is producing or injecting fluids, whereas static pressure is measured when the well is shut in, allowing pressures to stabilize and reflect the reservoir's equilibrium state. Flowing and static pressures typically contain different information content. Flowing pressure is highly sensitive to the near-wellbore region and is affected by short-term fluctuations in flow rates, while static pressure data capture the

overall reservoir energy and are sensitive to large-scale reservoir properties that reflect long-term behavior.

The sensitivities of flowing and static pressures are illustrated in Fig. 9.8, which shows the partial derivative values of bottom-hole pressure data with respect to gridblock permeability in a single-phase linear model, computed using the adjoint method. A single well is completed in the first gridblock. Fig. 9.8a shows the predicted pressure in the model, which undergoes a 30-day drawdown period followed by a 30-day build-up. Figs. 9.8b and 9.8c display the sensitivity values at different times. The flowing pressure sensitivity values with respect to the permeability of the gridblock containing the well are ten times larger than any other sensitivity values.

Due to the high sensitivity of flowing pressure data to the well productivity/injectivity index, practitioners often prefer the assimilation of static pressure measurements. After conditioning the models to static pressure, the match for flowing pressure is achieved by adjusting well productivity/injectivity index multipliers.

9.3.1.3 Example: Rate versus Ratio Data

Identifying problematic data points is often easier when analyzing ratio data, such as water cut and gas-oil ratio (GOR), compared to flow rate data. Fig. 9.9 shows this with actual data from two wells in an oilfield. For the first well, we observe periods with perfectly constant water cut (Fig. 9.9b), which is highly unlikely to reflect actual measurements. However, these constant water cut values are somewhat obscured when analyzing only the flow rate data (Fig. 9.9a). A similar issue arises with gas rate data; the gas rate plot (Fig. 9.9c) does not clearly indicate suspicious data points. In the GOR plot, however, it is easy to spot data points that deviate from the general trend. For example, Fig. 9.9d highlights two suspicious data points: the first with a large GOR value and another with GOR values below the initial value of the gas-oil solubility ratio.

9.3.1.4 Example: Use of Rate Allocation Tables

In offshore petroleum fields, the production data for individual wells are calculated by dividing the total platform production using rate allocation tables. These tables are updated based on measured well flow rates obtained from separation tests. Fig. 9.10 depicts the GOR for a well in the Santos Basin. In the figure, the green circles represent GOR values computed from allocation tables, while the red circles show GOR values obtained from separation tests. It is interesting to note that only a few separation tests are conducted annually, approximately one every two to three months. Significant changes in GOR values are observed at the time of these separation tests, suggesting that the intermediate GOR values calculated between tests

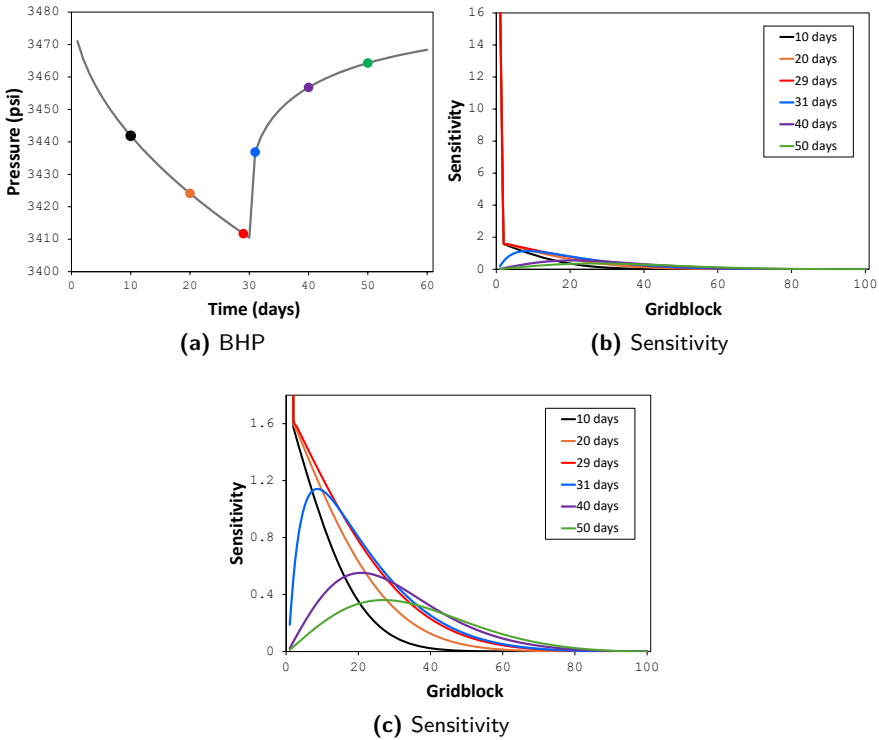


Fig. 9.8: Sensitivity values of bottom-hole pressure (BHP) with respect to permeability for a single-phase problem. Panel (a) shows the BHP prediction, indicating a drawdown period of 30 days followed by a 30-day build-up. Panel (b) displays the sensitivity values of BHP with respect to the permeability of each gridblock at different times. Panel (c) presents the same sensitivity values with an adjusted vertical scale. Example 9.3.1.2.

may be compromised due to outdated rate allocation factors. As a result, data errors are likely correlated over time for at least the duration between separation tests. Although this effect can be modeled using a non-diagonal $\mathbf{C}_e$ matrix, the simplest approach is to use only the data points measured during test times.

9.3.1.5 Example: Errors in Injection Rates

The historical fluid injection rates are often imposed as simulation constraints during data assimilation, meaning that reservoir models honor these data exactly. However, injection rates are subject to measurement errors. Consequently, forcing the models to use the observed rates precisely can introduce nonlinear errors in predictions. Fig. 9.11 illustrates this effect, showing a plot of predicted reservoir pressure for an oilfield in the Campos

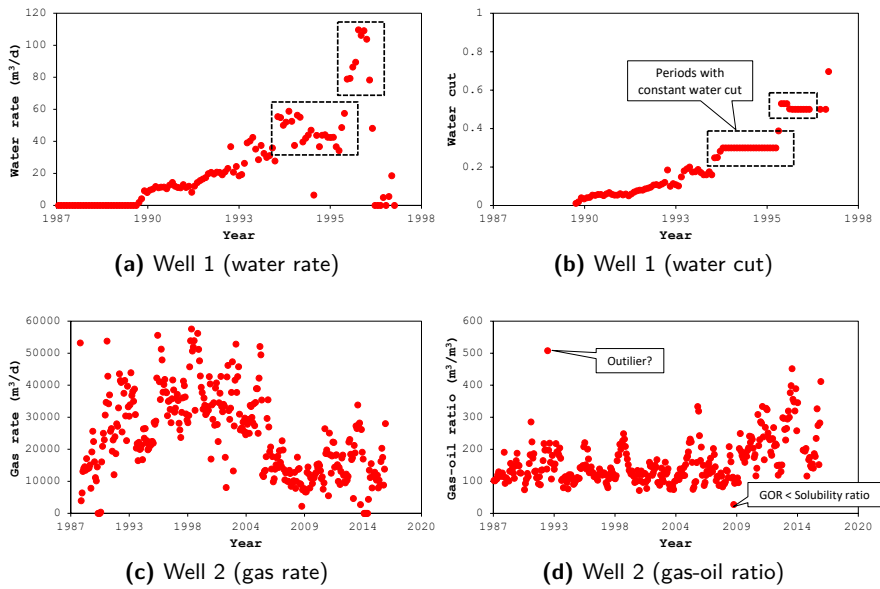


Fig. 9.9: Examples of data from an actual oilfield, illustrating how problematic measurements can be more easily identified in ratio data. Example 9.3.1.3.

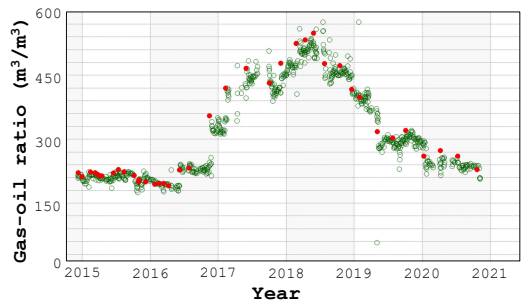


Fig. 9.10: Example of GOR data for a well in the Santos Basin. The red circles represent GOR measurements obtained from separation tests, while the green circles indicate GOR data calculated based on allocation tables. Example 9.3.1.4. Adapted from Emerick [123].

Basin. Field engineers identified a severe corrosion problem with the orifice plates used to measure total water injection rates, resulting in systematic errors in all water injection measurements. To investigate the effect of these errors on model predictions, the reservoir engineers ran simulations adjusting the total water injection by 2%. They concluded that a systematic error of $\pm 2\%$ in the water injection affects the average reservoir pressure, equivalent to a $\pm 7.5\%$ variation in the original volume of oil in place.

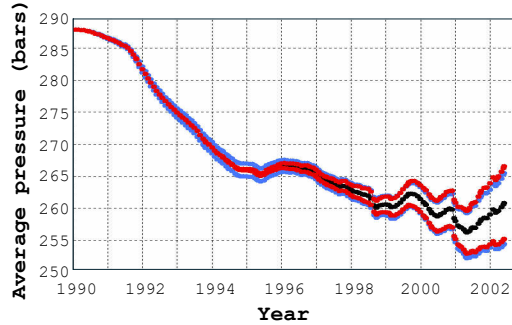


Fig. 9.11: Example of the effect of a 2% systematic error in the specified water injection rate on the reservoir pressure for a model of a field in the Campos Basin. The black dots correspond to the base prediction of the reservoir pressure. The red dots represent the predictions with a $\pm 2\%$ variation in the water injection rate, and the blue dots represent the predictions with a $\pm 7.5\%$ variation in the original volume of oil in place. Example 9.3.1.5.

9.3.1.6 Simulation Controls

During data assimilation, well controls in the models are set to honor the historical flow rates observed in the field. For oil-producing wells, there are three common control options: (1) impose the oil rate, (2) impose the liquid rate (oil plus water), and (3) compute the total fluid rate at reservoir condition and use this as the well-control. All three approaches are used in practice. The first approach is sometimes preferred because the simulator prioritizes the primary fluid of interest (oil). However, the second and third approaches have the advantage of honoring the historical fluid balance, often making it easier to match the average field pressure. For injection wells, the typical approach involves controlling the wells based on the observed injection rates.

The data used as well-controls are typically not included as observations in data assimilation. Ideally, we expect all models to honor the imposed control data. However, sometimes the models are unable to meet the specified constraints. Fig. 9.12 shows an example where prior models from an oilfield could not honor the specified liquid production rate, causing the bottom-hole pressure to reach a lower limit. This situation is undesirable and can negatively affect data assimilation performance. When the well-controls switch from a liquid rate to a pressure constraint, the ensemble loses the ability to estimate covariances between pressure data and model parameters, preventing the assimilation of these data. This situation commonly occurs in injection data when the simulation reaches the maximum allowable well pressure. In such cases, a possible approach is to use the injection rates as observed data for assimilation, even though they are also prescribed as control variables.

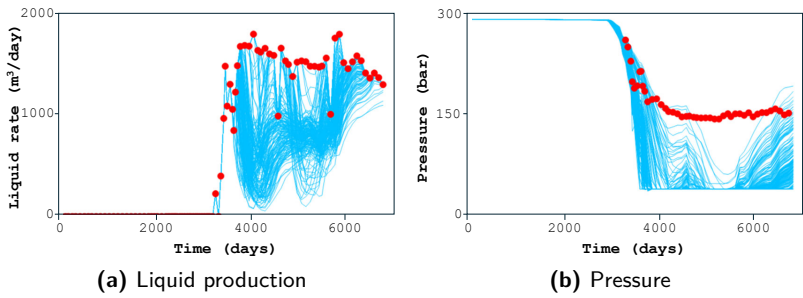


Fig. 9.12: Example of a well where the prior models were not able to honor the specified historical liquid production rate. The simulator switched the controls to minimum bottom-hole pressure.

9.3.2 4D Seismic Data

4D seismic, involving repeated seismic surveys over time, provides detailed insights into fluid movements and pressure changes within reservoirs. However, integrating 4D seismic data into reservoir models is more complex than incorporating production data. This complexity is due to several factors. The large volume of data points significantly increases computational demands, necessitating more degrees of freedom to accurately capture changes in rock properties while maintaining geological plausibility. Additionally, errors, biases, and inaccuracies in the forward model tend to be more pronounced in 4D seismic datasets, presenting substantial challenges that are difficult to identify and correct during the assimilation process.

4D seismic data assimilation is generally performed in one of three domains: the seismic domain, which includes attributes like amplitudes and time shifts; the elastic domain, which involves pressure and shear-wave velocity or impedance data; and the simulation domain, which deals with pressure and phase saturation. The elastic domain is often favored because it offers a good balance between data preparation and forward modeling [336]. Alternative representations of 4D data in the literature include discrete images [438, 327], fluid fronts [443, 254], and coefficients related to data compression methods [280, 146].

In 4D seismic assimilation, observed and simulated seismic data must be compared on a consistent scale. This entails converting observed seismic data from time to depth and adjusting it to a grid that aligns with the simulator’s response. However, because seismic data typically have lower vertical resolution compared to the simulation’s vertical discretization, 4D seismic data are often interpreted as “maps” corresponding to multiple layers of the simulation model.

9.3.2.1 Petroelastic Modeling

A petroelastic model (PEM) is used to convert reservoir simulation results, such as pressure and fluid saturation, into elastic properties. This model is crucial for 4D seismic data assimilation, and extensive literature on PEMs can be found in works by Mavko et al. [298] and Grana et al. [175]. Most PEMs are based on the Gassmann fluid substitution model [163], often supplemented with (semi-)empirical correlations to estimate the elastic properties of rock and fluid.

Characterizing the elastic properties of the dry-rock frame of the reservoir is typically the most influential and uncertain aspect of PEM. This characterization involves developing correlations for porosity, lithology, and effective pressure effects derived from well-logging data and laboratory experiments with rock core samples. Among these factors, the pressure-dependent behavior of elastic properties is generally the most uncertain.

Fluid density and bulk modulus are also critical components in PEMs. Many applications use the correlations presented in Batzle and Wang [31] for these parameters. For more complex fluid descriptions, such as compositional modeling of volatile oils, miscible gas injection, and high CO₂ content, Neto et al. [319] provide a recent and detailed discussion advocating the use of equations of state calibrated with laboratory measurements.

9.3.2.2 Example: 4D Seismic Data

Fig. 9.13 shows an example of 4D seismic data used for data assimilation in an oilfield in the Campos Basin. The 4D data correspond to the difference in inverted P-impedance between two streamer surveys acquired 14 years apart, normalized by the impedance of the base survey. Fig. 9.13a shows the seismic data in a cross-section in the original seismic scale after the time-to-depth conversion. In this figure, the positive seismic anomalies (depicted in blue) correspond to an increase in P-impedance, primarily reflecting the effect of water replacing oil in the reservoir. The original position of the oil-water contact is indicated. In this field, multiple wells inject water below the original contact. Negative anomalies (depicted in tones from red to yellow) are also visible and are interpreted as artifacts (side lobes)³. Due to these artifacts and the limited vertical resolution of streamer surveys, the 4D seismic data were represented as maps computed over the positive seismic anomalies for two main geological formations in this reservoir. These maps were then upscaled to match the horizontal discretization of the simulation grid, resulting in the image presented in Fig. 9.13b, which was used for

³ Side lobes are unwanted artifacts that appear in seismic data images [233]. These artifacts are typically generated by the secondary lobes of the wavelet. Ideally, seismic inversion from amplitude to impedance should remove these artifacts by eliminating the wavelet's effect. However, in practice, side lobes are frequently observed in 4D impedance datasets.

data assimilation. Figs. 9.13c and 9.13d show the predicted 4D seismic data before and after data assimilation, where improvements in the predicted seismic data are notable.

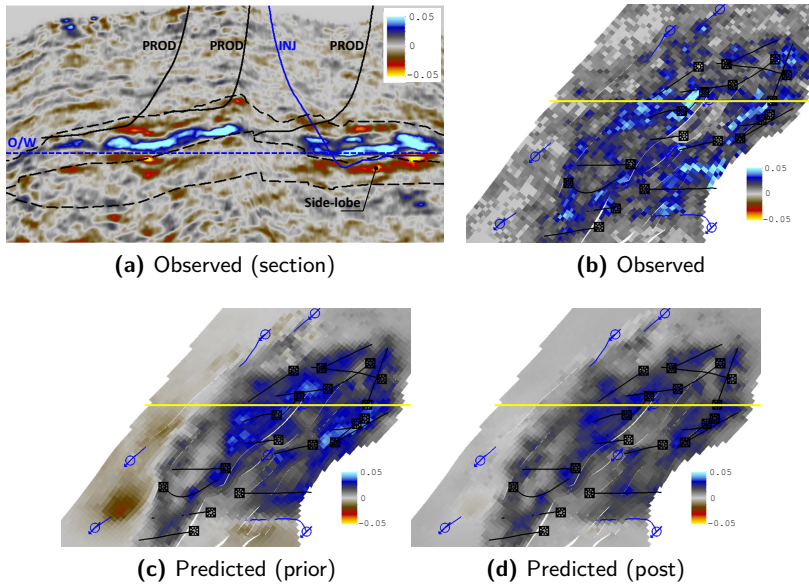


Fig. 9.13: 4D seismic data (normalized P-impedance changes) for oilfield in the Campos Basin. Panel (a) shows a cross-section of the observed 4D data in the seismic scale, indicating the original position of the oil-water contact (dashed horizontal blue line). Panel (b) shows the observed 4D data transferred to the reservoir simulation grid. Panel (c) shows the predicted 4D data with a prior model realization while Panel (d) shows the corresponding prediction after data assimilation. The horizontal yellow line in (b)–(d) indicates the position of the cross-section shown in Panel (a). Example 9.3.2.2.

9.3.2.3 4D Seismic Data-Error Covariance

Choosing the data-error covariance matrix, $\mathbf{C}_e$, for 4D seismic data presents a significantly greater challenge than production data. Firstly, 4D seismic datasets are prone to systematic errors due to inconsistent data acquisition conditions, extensive data processing, and uncertainties in time-to-depth conversion and amplitude-to-impedance inversion. Moreover, modeling errors in 4D seismic data, particularly issues related to scale transference and limitations of petroelastic models, are prevalent. Unlike production data, assuming uncorrelated data errors is often less realistic for seismic datasets [329]. However, using a non-diagonal $\mathbf{C}_e$ can be problematic for large datasets where memory storage of $\mathbf{C}_e$ may be impractical. In such

cases, the use of a low-rank representation of $\mathbf{C}_e$ in a viable alternative [138, 9].

As with other data types, removing or downweighting unreliable 4D seismic data points is crucial during assimilation. In this context, the normalized root-mean-squared error (NRMS) [241] is widely used to identify problematic data arising from survey repeatability issues. Several works propose strategies to estimate $\mathbf{C}_e$ in the context of 4D seismic data assimilation; see, e.g., Oliver [329] and references therein. One important aspect is to consider both data and model errors in this process. Alfonzo and Oliver [9] proposed estimating $\mathbf{C}_e$ based on the residuals of the data assimilation (the difference between observed and predicted 4D data from posterior realizations). This procedure is relatively general, simple to implement, and allows estimation of the combined effect of data and model errors. One drawback is that it requires repeating the data assimilation, increasing computational costs. Emerick and Neto [125] proposed estimating $\mathbf{C}_e$ based on the residuals between the observed seismic data and its projection onto the subspace defined by predicted data from the prior ensemble, eliminating the need for repetitions of the data assimilation. They used this procedure to identify and eliminate data points with significant discrepancies, compute error variance, and estimate error correlations across long distances.

9.3.3 Other Data

9.3.3.1 Tracer Data

Tracers are substances used to track fluid movement in applications involving multiphase flow. They are widely used for monitoring operations and reservoir characterization in applications including the injection and storage of waste fluids [87, 312], groundwater flow and contaminant transport in aquifers [351], and production in oil and gas fields [342, 477].

In practice, a tracer is injected into one well and monitored as it is recovered from one or more production wells. Data on tracer breakthrough time recovered mass and concentration over time provide information about inter-well pore volume and connectivity. There are also single-well tracer tests, which are used to estimate residual oil saturation [477].

There are two main types of tracers: radioactive and chemical. Radioactive tracers emit detectable radiation, offering high sensitivity and precision, while chemical tracers are versatile and can be tailored to specific reservoir conditions. Tracers can be further classified based on their behavior in the reservoir. Conservative tracers remain in the same phase throughout their movement, providing direct information about fluid flow paths and velocities. Partitioning tracers, on the other hand, shift between phases, revealing

interactions between fluids and reservoir rock, such as wettability and saturation.

Tracer modeling encompasses several physical and chemical effects, such as diffusion and dispersion, partitioning between fluid phases, adsorption onto reservoir rock, and radioactive decay. This modeling can be particularly sensitive to numerical dispersion, especially in large-scale 3D models where the volume of tracer pulse injections is negligible compared to the volumes of injected and produced fluids. Consequently, it is crucial to account for the model's limitations when incorporating tracers into the data assimilation process. For example, depending on the formulation, numerical scheme, and spatial and temporal discretizations, it may not be realistic to assimilate a full-time series of tracer concentration data. Instead, we might consider using only the tracer breakthrough time as the observed data and assume a relatively large data-error variance.

9.3.3.2 Example: Tracer and 4D Seismic Data

Fig. 9.14 shows a 4D seismic image and the interpreted pathways of chemical tracers with observed breakthroughs for the central area of a carbonate oil-field in the Santos Basin. The figure includes six oil-producing wells (labeled P1 to P6), two water injection wells (IW1 and IW2), four gas injection wells (IG1 to IG4), and two wells with water-alternating-gas injection (WAG1 and WAG2). The 4D image is a normalized map displaying hardening (blue) and softening (red) anomalies derived from time-lapse amplitude data. Hardening anomalies are interpreted as increases in reservoir impedance due to water injection, while softening anomalies are associated with gas injection. Notably, the 4D data is consistent with the tracer breakthrough information (indicated by blue and red arrows in Fig. 9.14b). This information helps geoscientists understand fluid distributions within the reservoir. For example, it is possible to identify preferential fluid paths for injected water aligned with geological faults in the field. Further details on this field can be found in [97] (see also Field 2 in Section 9.5.2).

9.3.3.3 Pressure Transient Data

Pressure transient analysis involves monitoring pressure changes in a well over a controlled period of operation. By analyzing the pressure response, engineers can interpret reservoir behavior using analytical models, such as the radial flow model, to estimate key reservoir parameters, including permeability, skin effects, reservoir boundaries, and fluid properties. Pressure transient data are also assimilated into numerical reservoir models, particularly in newer fields with limited well coverage or a brief production history. In such cases, special attention must be given to reducing time step sizes

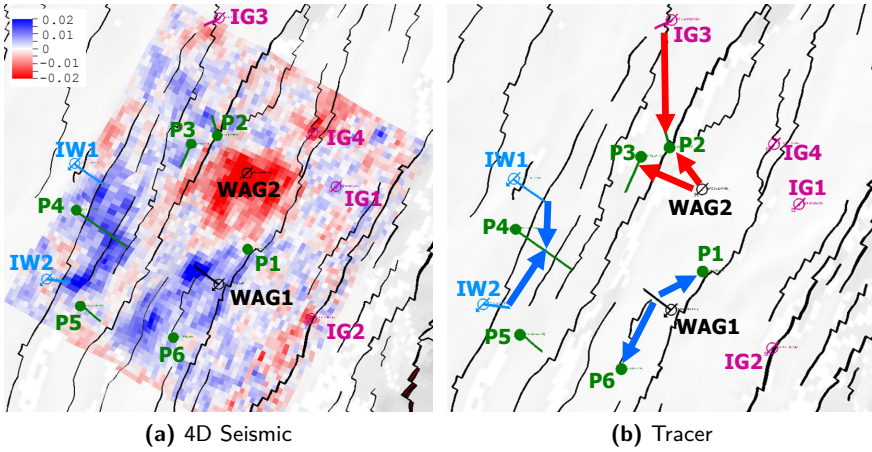


Fig. 9.14: Comparison between 4D seismic and tracer data for an oilfield in the Santos Basin. Panel (a) shows the 4D seismic data, while panel (b) indicates the interpreted pathways of chemical tracers with observed breakthroughs, injected in the water phase (blue arrows) and gas phase (red arrows). Example 9.3.3.2.

and enhancing grid discretization around the well to accurately capture the short-term transient behavior.

9.3.3.4 Well Profile Data

Well profile data comprise measurements of pressure and flow rate taken along the well trajectory. Among these, pressure profiles obtained using repeat formation testers (RFTs) [404] are invaluable for detecting reservoir compartmentalization and indicating fluid contact positions. Flow rate profiles acquired through production-logging tools (PLTs) [192] are frequently employed in data assimilation, especially for wells that have recently been drilled in the field. These profiles are instrumental in evaluating the relative contributions of various reservoir formations to the total flow rate of the well.

9.3.3.5 Example: Pressure Transient and PLT Data

Coutinho et al. [94] presents an example of PLT and pressure transient data assimilation using the standard EnKF for a vertical well in an oilfield in the Santos Basin. The model parameters correspond to permeability multipliers, one for each reservoir layer. Fig. 9.15a shows the pressure data and the corresponding predictions obtained from the prior and posterior ensembles, indicating a visible improvement in the data matches. Fig. 9.15b

presents a log-log diagnostic plot of the first build-up period, displaying pressure changes and pressure derivatives. These results indicate that, despite improvements in the predicted pressure data, the posterior ensemble was unable to capture the correct level of the pressure derivative. Note that pressure changes and pressure derivatives are derived from pressure measurements and are not used as data during assimilation; rather, they are presented in Fig. 9.15b for analysis of the results. Fig. 9.15c shows the PLT data, which corresponds to the cumulative contribution of each reservoir layer to total liquid production. Fig. 9.16 shows the prior and posterior distributions of permeability multipliers, indicating that only a few reservoir layers are effectively contributing to total well production.

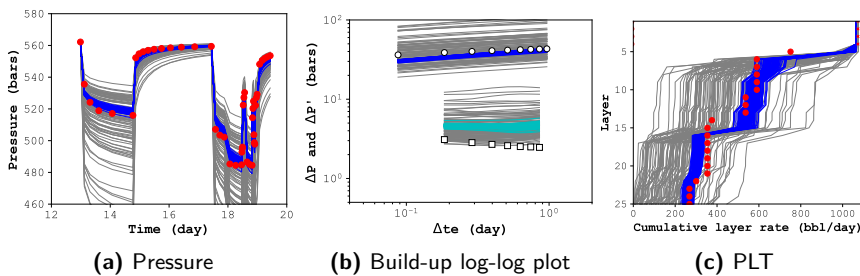


Fig. 9.15: Pressure and PLT for a well in an oilfield in the Santos Basin. Red circles represent observed data, while gray and blue lines show predicted data from the prior and posterior ensembles after data assimilation with EnKF, respectively. Panel (a) shows the pressure data. Panel (b) shows a log-log diagnostic plot of the first build-up period, where the white circles and squares represent the pressure changes and pressure derivatives, respectively. Panel (c) shows the PLT data in terms of the cumulative contribution of each reservoir layer to the total liquid production. Example 9.3.3.5. Adapted from Coutinho et al. [94] with permission from the authors.

9.3.3.6 Well Log Data

Similar to well profile data, well logging involves measurements taken along the well trajectory. Various methods and tools are used for well logging, including electric, acoustic, and radioactive logs. These logs provide indirect measurements of reservoir rock and fluid properties. Despite being susceptible to noise and interpretation errors, log data are generally considered very accurate. They represent a primary source of information for reservoir characterization and are used as conditioning data in geostatistics to generate prior realizations. Because these data are precisely honored in geostatistics, they are often referred to as “hard data.”

It is worth noting that ensemble methods use the prior ensemble to construct the uncertainty space for data assimilation. This means that if a

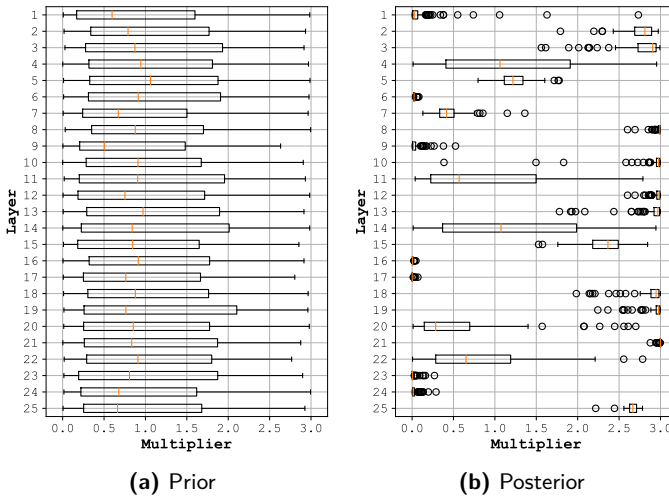


Fig. 9.16: Distribution of layer permeability multipliers. Example 9.3.3.5. Adapted from Coutinho et al. [94] with permission from the authors.

parameter value does not vary in the ensemble, it is interpreted as having no associated uncertainty. Consequently, hard data used in prior geostatistical modeling are retained during data assimilation. This is a nice feature of the method as it assures that prior data are preserved during the process. However, we should keep in mind that it also means that uncertainty underestimation in the prior modeling will invariably lead to uncertainty underestimation in the posterior.

9.4 Results Evaluation

The quality and adequacy of the initial ensemble are critical for successful data assimilation. The ensemble must accurately reflect uncertainties, ensuring it samples the prior uncertainty space adequately. Beginning data assimilation with an inadequate ensemble leads to well-known consequences: failure to match observed data, geological models lacking plausibility, and inaccurate uncertainty estimates in reservoir performance forecasts.

However, achieving this adequacy is often challenging in practical applications, as important uncertainty parameters are frequently overlooked during the problem setup. This oversight arises from inherent limitations in recognizing all relevant aspects of the problem, including biases from modeling assumptions and constraints in time and resources. To address this, it is crucial to thoroughly analyze processes and results, revisiting assump-

tions as necessary. In this context, Oliver and Alfonzo [330] outlines data assimilation in reservoir models as a five-step process:

1. **Model creation:** Select the main modeling hypotheses, including parametrization, observations, and error tolerances.
2. **Model criticism (*a priori* perspective):** Analyze the consistency of the prior models compared to observations.
3. **Model calibration:** Perform data assimilation to update the model parameters based on the observations.
4. **Model criticism (*a posteriori* perspective):** Analyze the consistency of the posterior models.
5. **Model improvements:** Revisit the model creation step if necessary.

These steps may require iteration. Both pre- and post-calibration model criticism should focus on three main aspects:

- Geological plausibility.
- Ability to reproduce the observed behavior of dynamic data.
- Ability of the ensemble to adequately estimate the range of uncertainties in predictions.

In Chapter 4, we derived the EnKF starting from the RML objective function. In Chapter 5, we similarly used RML to introduce ES-MDA. Although this is not the only way to present these methods, it highlights a key similarity between them: both aim to minimize an objective function consisting of two components—the prior and the likelihood. The likelihood term increases when predicted data diverge from observations, while the prior term increases as models deviate from the initial ones. Thus, the goal is to find the minimal adjustments to the prior models necessary to match the observations [328]. However, when initial model predictions differ significantly from the data, substantial changes to the models are required to minimize the likelihood term, often compromising geological plausibility.

For example, if none of the initial model realizations can replicate the production behavior of one or more wells, it indicates flaws in the prior modeling. In these cases, it is generally more beneficial and efficient to invest time in refining the initial modeling rather than attempting to correct it later during the calibration phase.

9.4.1 Example: Non-Representative Priors

Fig. 9.17 demonstrates an ES-MDA application where the initial modeling did not adequately represent key uncertainties. This is evident from the initial water production curves, which show minimal variation across the 100 prior realizations and a significant mismatch with the observations. Although ES-MDA managed to produce an ensemble that matched the water

production data, it required substantial adjustments in the permeability values beyond the expected variability range. Further analysis revealed the importance of incorporating multiple faults (barriers) in the model to achieve proper data assimilation. Fig. 9.18 illustrates the data assimilation results for the same field in a more recent study. In this case, transmissibility multipliers across several faults have been included as uncertainty parameters. As a result, only mild modifications are observed in the permeability distribution after data assimilation. Fig. 9.18e shows the water cut data for the same well presented in Fig. 9.17 but with a longer production history. Note that, unlike the previous example, the predictions of the prior ensemble span the observations, and excellent data matches were obtained.

Fig. 9.19 presents another field example highlighting discrepancies between observed data and prior modeling. Here, the observed pressure data for a particular well showed a significant deviation from the predicted range of the prior ensemble. As a result, ES-MDA led to an unrealistic reduction in permeability around the well location. Subsequent investigations identified inconsistencies in the specified water injection for a nearby well within the model.

9.4.2 Prior Observation Coverage and Mean Squared Error

The previous examples show that when simulated production profiles diverge significantly from observed production history, the initial models are not suitable for the calibration step. However, identifying this issue is not always straightforward through visual inspection alone, particularly in fields with many wells. In this section, we discuss metrics designed to evaluate the quality of the prior ensemble based on predicted production data. These metrics act as qualitative indicators, and failure to meet them indicates a need to revisit the prior model.

9.4.2.1 Observation Coverage (OC)

This metric assesses whether the range of predicted data is sufficiently broad to encompass the observations. We count the number of data points ($\pm\sigma_e$) that fall within the minimum and maximum predictions. Based on the value of OC, we classify the ensemble as follows:

- Excellent: $OC > 0.95$.
- Good: $0.8 < OC \leq 0.95$.
- Fair: $0.7 < OC \leq 0.8$.
- Poor: $OC \leq 0.7$.

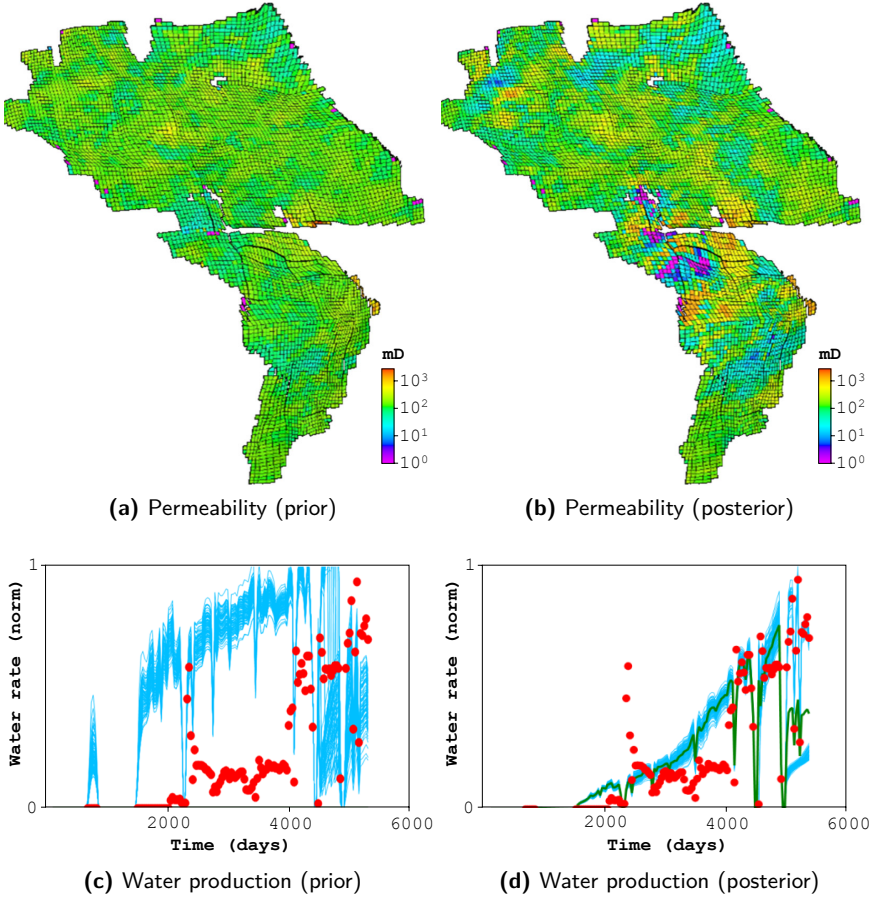


Fig. 9.17: Example of data assimilation using ES-MDA, where the prior model predictions have been improved by introducing transmissibility multipliers across several faults mapped from seismic data. This figure represents a later data assimilation study of the same field presented in Fig. 9.17. Well 1 is the same well shown in Fig. 9.17, but with a longer production history. Example 9.4.1. Adapted from Emerick [123].

9.4.2.2 Mean Squared Error (MSE)

This metric evaluates whether the predicted data from the prior ensemble are biased compared to the observations. The MSE is calculated using the square of the Mahalanobis distance relative to the covariance $\mathbf{C} = \Delta \mathbf{D} \Delta \mathbf{D}^\top + \mathbf{C}_e$:

$$\mathcal{D}(\mathbf{m}) = \left(\mathbf{d}_{\text{obs}} - \overline{\mathbf{g}(\mathbf{m})} \right)^\top \mathbf{C}^{-1} \left(\mathbf{d}_{\text{obs}} - \overline{\mathbf{g}(\mathbf{m})} \right). \quad (9.7)$$

The inverse of $\mathbf{C}$ can be computed using the procedures discussed in Chapter 8, Section 8.3. Alternatively, for the sole purpose of computing the MSE,

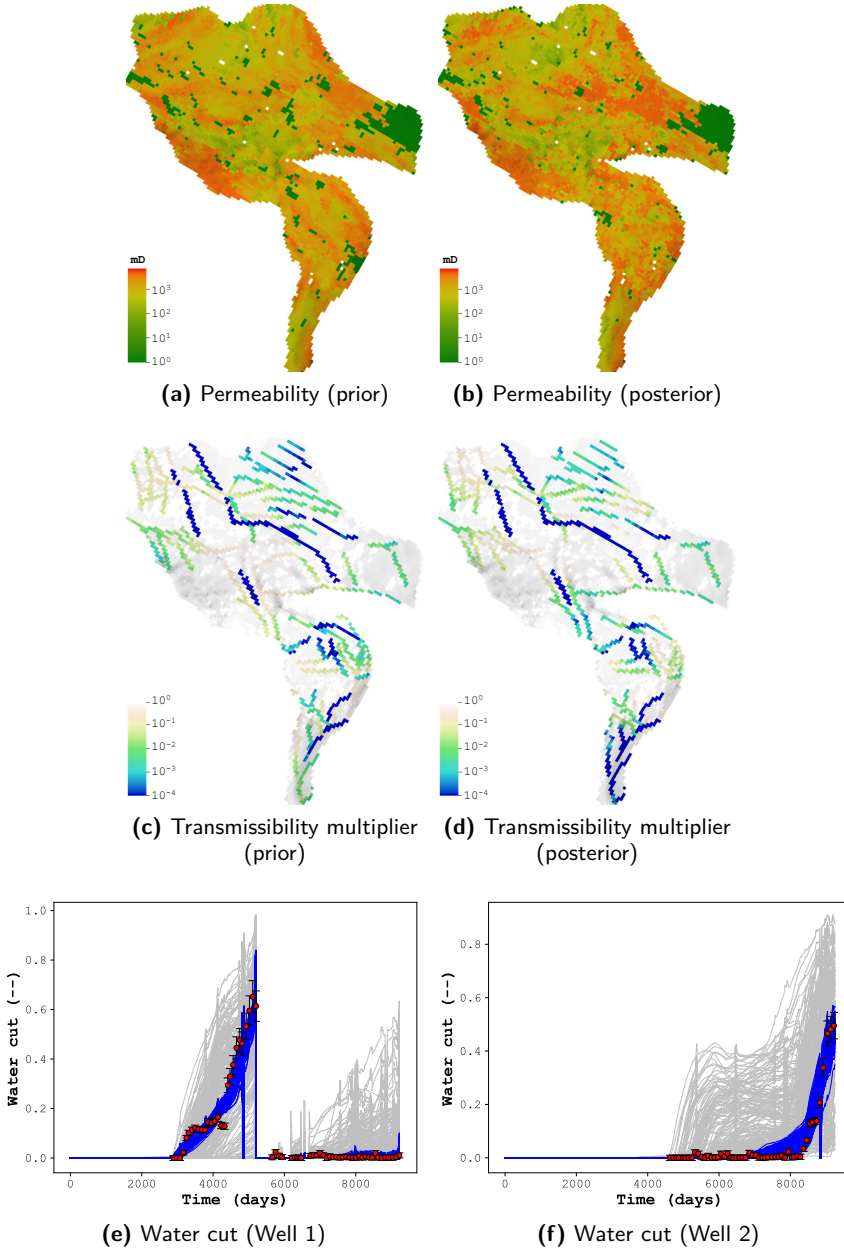


Fig. 9.18: Example of data assimilation with ES-MDA where the prior model predictions have been improved with the introduction of transmissibility multipliers across several faults mapped from seismic data. This is a later data assimilation study of the same field presented in Fig. 9.17. In this figure, Well 1 is the same well with data presented in Fig. 9.17, but with a longer production history. Red circles in (e) and (f) represent observed data, while gray and blue lines show predicted data from the prior and posterior ensembles, respectively. Example 9.4.1.

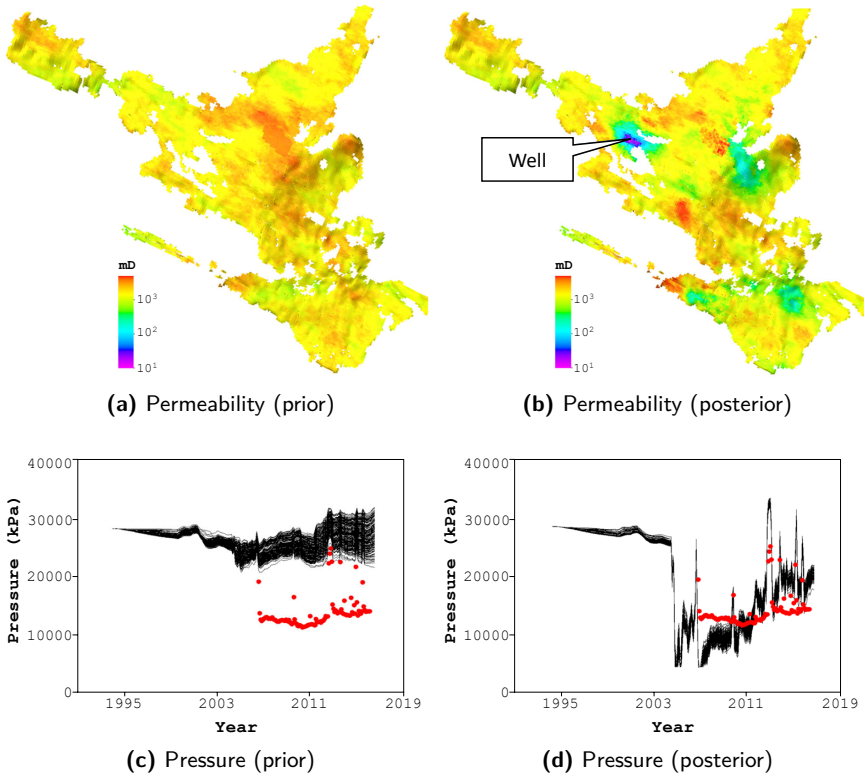


Fig. 9.19: Example of data assimilation with ES-MDA where the data are inconsistent with the prior ensemble. The observed bottom-hole pressure in a particular well fell completely outside the predicted range of the prior ensemble. Consequently, data assimilation led to an unrealistic reduction in permeability around the well location. Subsequent investigation revealed inconsistencies in the water injection specified for a nearby well in the model. Example 9.4.1.

it is sufficient to use only the diagonal terms

$$\mathbf{C} = \begin{bmatrix} \mathbb{V}[g_1(\mathbf{m})] + \sigma_{e,1}^2 & & 0 \\ & \ddots & \\ 0 & & \mathbb{V}[g_{N_d}(\mathbf{m})] + \sigma_{e,N_d}^2 \end{bmatrix} \quad (9.8)$$

and compute the MSE as

$$\text{MSE} = \frac{\mathcal{D}(\mathbf{m})}{N_d}. \quad (9.9)$$

Based on the value of MSE, we classify the ensemble as follows:

- OK: $\text{MSE} \leq 1$.
- Fair: $1 < \text{MSE} \leq 2$.
- Biased: $\text{MSE} > 2$.

Instead of using MSE, we can apply the chi-square hypothesis test. In this approach, we define the null hypothesis as “ $\mathbf{d}_{\text{obs}}$ is a sample from $\mathcal{N}(\mathbf{g}(\mathbf{m}), \mathbf{C})$.” The metric $\mathcal{D}(\mathbf{m})$ follows a chi-square distribution with N_d degrees of freedom, $\chi^2_{N_d}$. The CDF of $\chi^2_{N_d}$ is given by

$$P_\chi(\mathcal{D}(\mathbf{m})) = \frac{1}{\Gamma(N_d/2)} \gamma\left(\frac{N_d}{2}, \frac{\mathcal{D}(\mathbf{m})}{2}\right), \quad (9.10)$$

where $\Gamma(\cdot)$ is the Gamma function and $\gamma(\cdot, \cdot)$ is the lower incomplete Gamma function. The p -value of the null hypothesis is

$$p = 1 - P_\chi(\mathcal{D}(\mathbf{m})). \quad (9.11)$$

If $p > p_0$ (say $p_0 = 0.05$), the null hypothesis is true.

9.4.2.3 Example: OC and MSE

Fig. 9.20 displays the predicted water cut for three wells in an oilfield in the Campos Basin, along with their corresponding OC and MSE values. Strictly speaking, the prior ensemble only passes the quality test for the first well. However, in practice, it can be challenging and time-consuming to create a prior ensemble that meets the quality criteria for all wells. These metrics are not strict thresholds but rather indicators of ensemble quality. They can be computed on a well-by-well basis, as shown in Fig. 9.20, for a group of wells or even for the entire field, providing a global assessment. Using these metrics also helps to track improvements in prior modeling.

9.4.3 Plausibility of the Posterior Realizations

Validating the results can pose challenges, especially concerning the plausibility of the posterior models due to the numerous realizations involved. It is essential to evaluate the statistical properties of all data assimilation parameters by comparing them before and after assimilation. Calculating metrics that assess the overall changes in the models can be helpful in this regard. However, a comprehensive and thoughtful evaluation of the posterior models necessitates the collaboration of a multidisciplinary team.

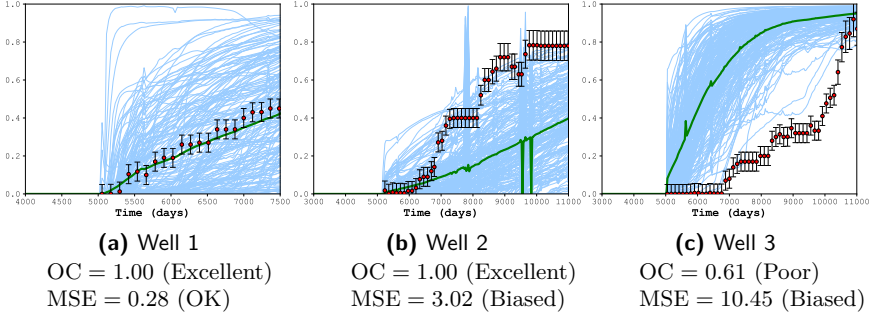


Fig. 9.20: Examples of predicted water cut for wells from a field in the Campos Basin with varying levels of OC and MSE. Red dots are the observed data points. The error bars in the observations correspond to one standard deviation of the data error. Light blue lines are the predicted data from the prior ensemble. The green line is the average prediction. Example 9.4.2.3.

9.4.3.1 Normalized Model Change (NMC)

We can calculate the differences between the posterior and prior realizations to identify areas where the model has undergone significant modifications. Standardizing these changes by dividing them by the prior standard deviation is recommended in this context. The NMC metric summarizes these model changes:

$$\text{NMC} = \frac{1}{N_m} \sum_{i=1}^{N_m} \left| \frac{m_{c,i} - m_i}{\sigma_{m,i}} \right|, \quad (9.12)$$

where m_i and $m_{c,i}$ represent the i th values of the model parameter before and after data assimilation, respectively. $\sigma_{m,i}$ denotes the corresponding prior standard deviation. A NMC value of 1 indicates an average change approximately equal to one standard deviation of the prior uncertainty.

9.4.3.2 Cosine Similarity (CS)

Another valuable measure for assessing the similarity between two models is the cosine similarity (CS). It calculates the cosine of the angle between two vectors, indicating whether these vectors are aligned in a similar direction, i.e.

$$\text{CS} = \frac{\mathbf{m}^\top \mathbf{m}_c}{\|\mathbf{m}\| \|\mathbf{m}_c\|} \in [-1, 1], \quad (9.13)$$

where $\mathbf{m}$ and $\mathbf{m}_c$ are the prior and posterior realizations, respectively. In this context, $\text{CS} = 1$ indicates that the two models are identical.

9.4.3.3 Facies Overlap Coefficient (FOC)

For categorical or discrete properties such as facies, a useful metric for evaluating the similarity between two models is the overlap coefficient:

$$\text{FOC}(m_i, m_j) = \begin{cases} 1 & \text{if } m_i = m_j \\ 0 & \text{otherwise.} \end{cases} \quad (9.14)$$

We can use the average FOC for the model as a measure of similarity between prior and posterior realizations.

9.4.3.4 Normalized Variance (NV)

For linear-Gaussian problems, it can be shown that the posterior covariance, $\mathbf{C}_{\mathbf{m}_c}$, has the form [334]:

$$\begin{aligned} \mathbf{C}_{\mathbf{m}_c} &= \mathbf{C}_{\mathbf{m}} - \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top (\mathbf{C}_e + \mathbf{G} \mathbf{C}_{\mathbf{m}} \mathbf{G}^\top)^{-1} \mathbf{G} \mathbf{C}_{\mathbf{m}} \\ &= \mathbf{C}_{\mathbf{m}} - \Delta \mathbf{C}_{\mathbf{m}}. \end{aligned} \quad (9.15)$$

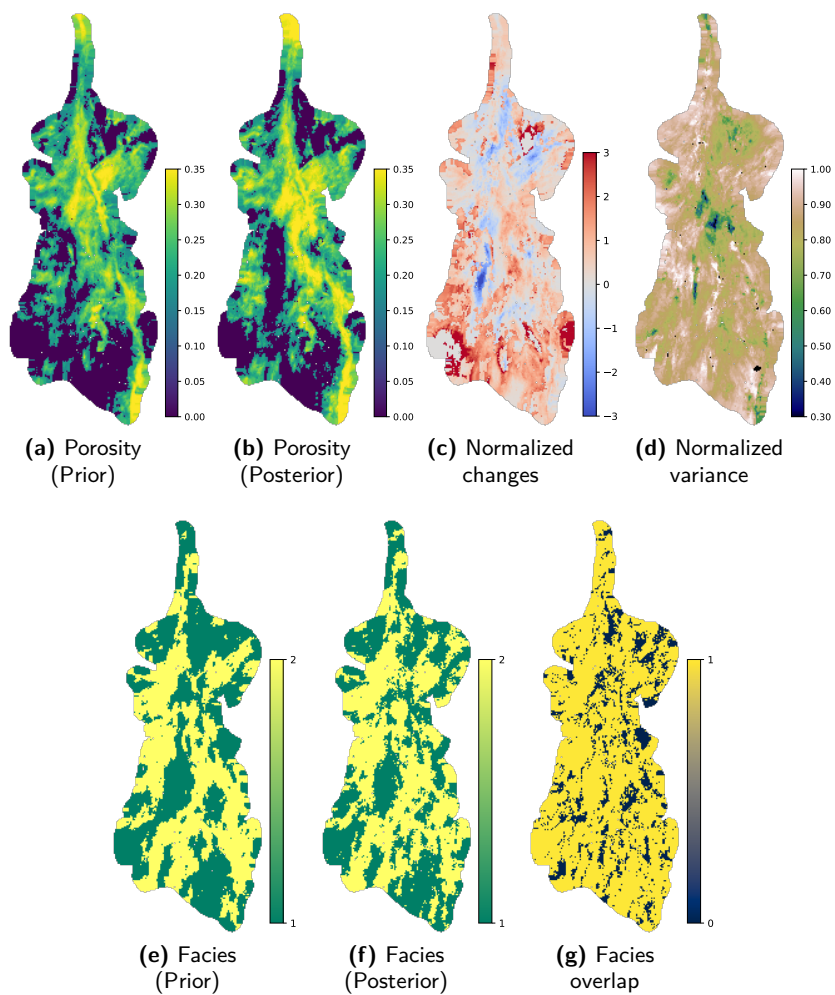
The matrix $\Delta \mathbf{C}_{\mathbf{m}}$ is real symmetric positive semidefinite, ensuring all diagonal entries are nonnegative. Let c_i , c'_i , and δc_i denote the i th diagonal entry of $\mathbf{C}_{\mathbf{m}}$, $\mathbf{C}_{\mathbf{m}_c}$, and $\Delta \mathbf{C}_{\mathbf{m}}$, respectively. The normalized variance, NV_i , of the i th model parameter m_i is defined as:

$$\text{NV}_i = \frac{c'_i}{c_i} = 1 - \frac{\delta c_i}{c_i}. \quad (9.16)$$

NV_i ranges between 0 and 1, providing a measure of the reduction in variance of m_i due to data assimilation [334]. A NV_i value of 1 indicates no reduction in uncertainty, while $\text{NV}_i = 0$ signifies a complete collapse in the uncertainty estimate.

9.4.3.5 Example: Model Changes

Fig. 9.21 depicts the model changes and corresponding metrics for a turbidite reservoir in the Campos Basin. This example shows that the overall changes in the model during data assimilation were moderate, preserving the general characteristics of the prior in the posterior model. The normalized changes image (Fig. 9.21c) and facies overlap (Fig. 9.21g) facilitate the identification of the primary changes induced by data assimilation. Additionally, the normalized variance image (Fig. 9.21d) highlights regions where the largest uncertainty reduction occurred due to the data assimilation.



Metric	Value
Normalized model change	0.739
Cosine similarity	0.968
Facies overlap coefficient	0.868
Average normalized variance	0.849

Fig. 9.21: Example of model changes after data assimilation with ES-MDA in a turbidite reservoir in the Campos Basin. Example 9.4.3.5.

9.4.4 Data-Mismatch Objective Function

The data-mismatch objective function divided by the number of data points serves as a useful metric to quantify the quality of data matching:

$$\mathcal{O}_{N,d}(\mathbf{m}) = \frac{1}{2N_d} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m}))^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{g}(\mathbf{m})). \quad (9.17)$$

For diagonal $\mathbf{C}_e$, $\mathcal{O}_{N,d}$ reduces to

$$\mathcal{O}_{N,d}(\mathbf{m}) = \frac{1}{2N_d} \sum_{i=1}^{N_d} \left(\frac{d_{\text{obs},i} - g_i(\mathbf{m})}{\sigma_{e,i}} \right)^2. \quad (9.18)$$

The average objective function for posterior samples in linear-Gaussian problems is expected to be around 1/2 [330]. However, achieving this level of objective function is typically observed only in controlled synthetic problems where model and data uncertainties are well characterized and significant model errors are absent. Establishing stringent criteria for data matching quality in real-world field applications can be challenging and subjective, as terms like “good” or “acceptable” data matches vary depending on context. Experienced practitioners emphasize the importance of aligning evaluation criteria with the study’s objectives when assessing model adequacy.

Nevertheless, the objective function defined in Eq. (9.18) provides a valuable reference point. One useful interpretation of $\mathcal{O}_{N,d}(\mathbf{m})$ values is how they compare to the standard deviation of the observation error. For instance, if the absolute difference $|d_{\text{obs},i} - g_i(\mathbf{m})|$ in (9.18) equals $\sigma_{e,i}$ for all data points, then $\mathcal{O}_{N,d}(\mathbf{m}) = 0.5$. Similarly, mismatches of $2\sigma_{e,i}$ or $3\sigma_{e,i}$ correspond to $\mathcal{O}_{N,d}(\mathbf{m}) = 2$ and $\mathcal{O}_{N,d}(\mathbf{m}) = 4.5$, respectively (see Fig. 2.17, Chapter 2). Higher values of $\mathcal{O}_{N,d}(\mathbf{m})$ may indicate the need to re-evaluate the validity of the data assimilation process, including prior modeling or the possibility that data-error covariances have been underestimated.

9.4.4.1 Example: Data Mismatch Objective Function

Fig. 9.22 presents examples of predicted water cut data for a well from a field in the Campos Basin, along with the corresponding values of $\mathcal{O}_{N,d}(\mathbf{m})$. In this figure, panels (a) and (b) depict scenarios where the data matches have average mismatches on the order of $\sigma_{e,i}$ and $3\sigma_{e,i}$, respectively.

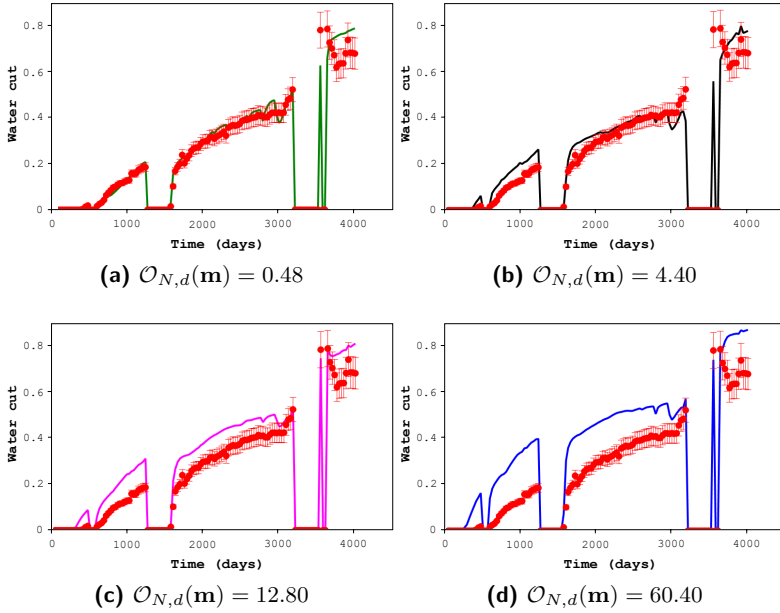


Fig. 9.22: Example of predicted water cut data for well in the Campos Basin with corresponding values of data mismatch objective function. The error bars correspond to $\pm\sigma_{e,i}$, which were defined as 10% of the data values (with a minimum of 0.01). Example 9.4.4.1. Adapted from Emerick [123].

9.5 Field Examples

This section presents three data assimilation case studies from real petroleum reservoirs. The first case involves a field with a long production history, demonstrating the effectiveness of data assimilation in problems with a large number of production data points. The second case focuses on a carbonate reservoir, highlighting the assimilation of 4D seismic data. The third case illustrates the data assimilation involving facies updates. It is important to note that these cases were not designed specifically as examples for this book. They represent actual applications conducted by teams of engineers and geoscientists. Therefore, these examples should not be viewed as ideal applications or for drawing generalized conclusions about performance. Instead, they should be considered in the context of real-world applications, which are often constrained by practical factors such as strict deadlines and specific study objectives.

9.5.1 Field 1

The first example pertains to a mature oilfield in the Campos Basin operating with water injection for more than 32 years. This field is characterized by a turbidite reservoir with poorly consolidated sandstones, low clay content, and high permeability [47, 221]. The simulation model consists of $332 \times 376 \times 18$ gridblocks (502,389 active). The reservoir fluid is modeled using a black-oil formulation. Production data includes quarterly measurements of water cut and gas-oil ratio from 121 wells, as well as pressure data collected during well shutdown periods. The data-error covariance matrix was assumed diagonal with standard deviations computed as 10% of the corresponding water cut values, with a minimum limit of 0.01. For gas-oil ratio data, an error standard deviation of 20% was assumed. Additionally, a constant value of 1,000 kPa was used for the pressure data. These data were assimilated using 10 ES-MDA iterations with constant inflation factors and Kalman gain localization to update 100 realizations.

The data assimilation parameters include porosity, horizontal, and vertical permeability for each gridblock of the model. Additionally, vertical transmissibility multipliers between the main geological formations were used to account for the presence of thin shale barriers. During data assimilation, the logarithm of permeabilities and transmissibility multipliers were adjusted. Other data assimilation parameters include rock compressibility, transmissibility multipliers across 50 faults, and Corey exponents of oil-water relative permeability curves.

Fig. 9.23 shows the evolution of the data-mismatch objective function throughout the ES-MDA iterations. The average objective function decreased significantly from an initial value of 47.18 to a final value of 5.14, indicating a substantial improvement in the data-match quality of the models. This improvement is further illustrated in Fig. 9.24, which presents the production data from four wells in the field. Each plot includes the observed data along with the predicted data from both the prior and posterior models.

Analyzing the metrics for the prior ensemble reveals that the observation coverage for the entire field is 0.84, indicating that 84% of the observed data points fall within the range of predicted data with the prior ensemble. A well-by-well analysis shows that 21% of the wells have poor observation coverage ($OC < 0.7$), while 56% exhibit excellent coverage ($OC > 0.95$). The overall mean squared error is 3.0, classifying the prior ensemble as biased, as discussed in Section 9.4.2.2.

Fig. 9.25 depicts the posterior mean values of the transmissibility multipliers across the 50 faults mapped in this field. The faults are ranked based on their multiplier values and categorized as closed, partially sealed, or open to flow.

Fig. 9.26 shows the first two realizations of porosity, horizontal permeability, and vertical transmissibility multipliers before and after data assimilation. Local changes in rock properties, particularly permeability and

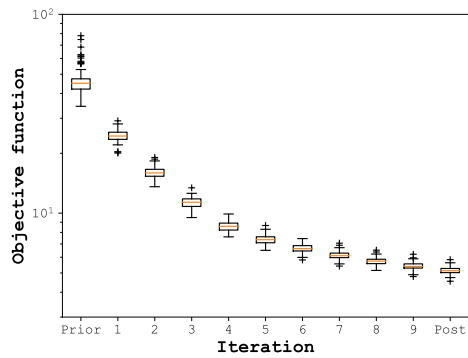


Fig. 9.23: Evolution of the data-mismatch objective function for Field 1.

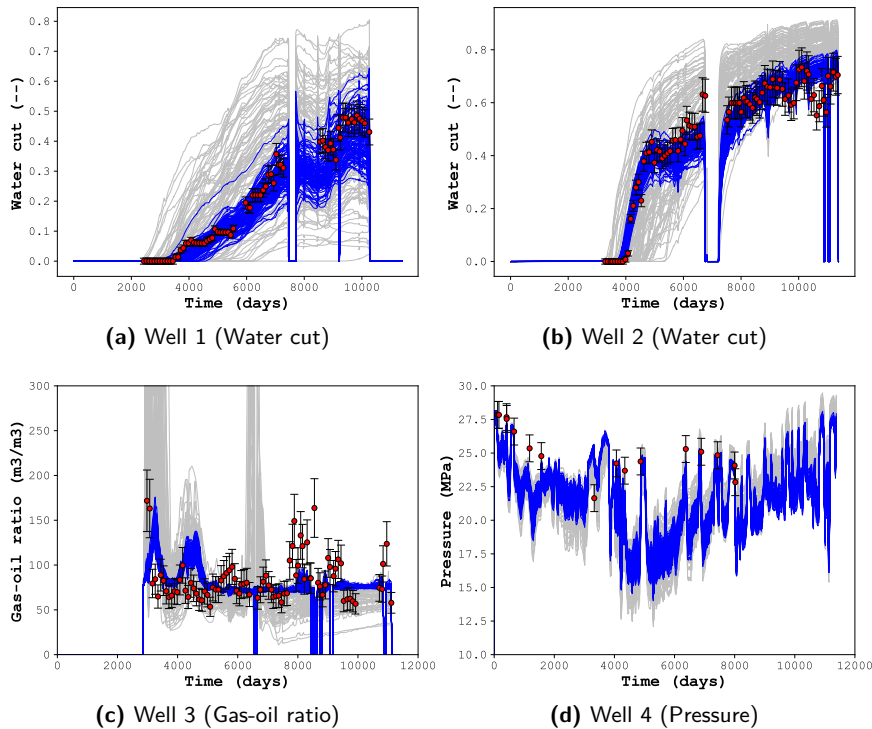


Fig. 9.24: Production data from four wells of Field 1. Red circles represent observed data, while gray and blue lines show predicted data from the prior and posterior ensembles, respectively. The error bars in the observed data points correspond to one standard deviation of the data error.

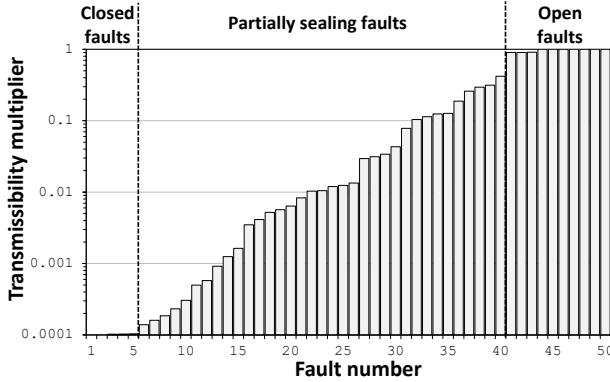


Fig. 9.25: Posterior mean transmissibility multipliers across faults for Field 1.

transmissibility multipliers, are evident. Despite these updates, the posterior realizations preserve the main characteristics of the corresponding prior models, which is often considered empirical evidence of successful data assimilation. Just for reference, the average values of normalized model change and cosine similarity for this case are 0.540 and 0.997, respectively. The average normalized variance for the grid parameters is 0.8, indicating that significant variability was still present in the posterior ensemble.

There are numerous other results and aspects of this case that could be explored further. The results presented in Figs. 9.23 through 9.26 were selected solely to illustrate the overall application.

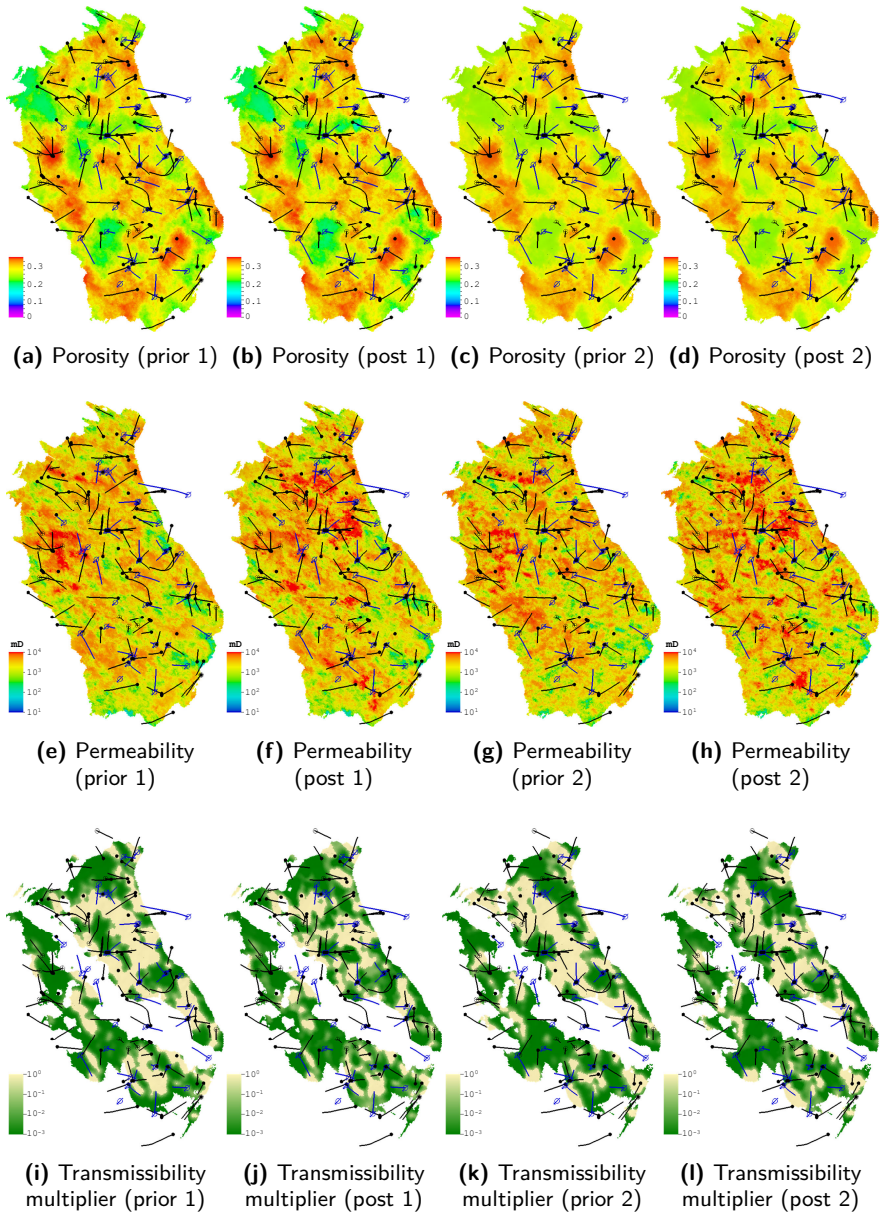


Fig. 9.26: First two realizations of porosity, horizontal permeability and vertical transmissibility multipliers for Field 1. Oil producing wells are indicated in black and water injection wells are indicated in blue.

9.5.2 Field 2

The second field example involves a giant carbonate reservoir in Brazil's pre-salt region. The reservoir model for this field is discretized into $217 \times 362 \times 71$ gridblocks, incorporating dual porosity and dual permeability to simulate flow in the fracture system. Including both matrix and fractures, the total number of active gridblocks is 2,150,510. The fluid model is compositional with nine pseudo-components. Data assimilation involved using 200 realizations of the geological model to calibrate the porosity of the rock matrix, as well as the horizontal and vertical permeability of both the matrix and fractures. Additionally, the parameters for the oil-water and gas-liquid relative permeability curves were adjusted during data assimilation. Production in this field commenced in 2009 through an extended well test, and currently, over 100 wells operate in this field. The production data used in the data assimilation consists of monthly measurements of water cut, gas-oil ratio, and pressure data during well shutdown periods. The data-error covariance matrix for production data was assumed to be diagonal, with a standard deviation of 10% for water cut and gas-oil ratio data values. For pressure measurements, a constant value of 200 kPa was assumed.

In 2015, this field became the focus of a pioneering pilot project to assess the feasibility of 4D seismic technology for monitoring water-alternating-gas (WAG) injection in ultradeep offshore reservoirs with complex overburden and low-compressibility rocks. Data collection for the project included base and monitor surveys acquired two years apart using ocean bottom nodes (OBN) technology. The results from this pilot project demonstrated the effectiveness of OBN surveys for production monitoring in pre-salt reservoirs [97].

Although the 4D data is limited to the central part of the field, the model of the entire field was updated by assimilating production data from all wells. However, here, the focus is on the results within the pilot area, which includes six oil-producing wells, two water injection wells, four gas injection wells, and two WAG wells, as illustrated in Fig. 9.27.

The 4D seismic data consist of a normalized map depicting hardening and softening anomalies. These data were compared to simulated maps of P-impedance differences between the monitor and base dates, normalized by the base survey. To achieve this, the simulation models were integrated with a PEM based on the Gassmann equation to compute the predicted impedance response. An upscaling process using weighted averaging across the reservoir thickness was employed to compute predicted seismic maps. The PEM utilizes empirical correlations derived from experimental data to estimate the elastic properties of the dry-rock frame (Fig. 9.28), and a cubic equation of state, calibrated with experimental data, to estimate the fluid properties.

Fig. 9.34a illustrates the observed seismic data, highlighting primary time-lapse anomalies. The majority of these anomalies fall within the range

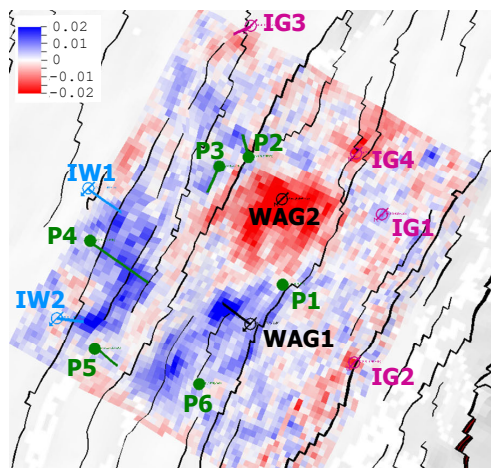


Fig. 9.27: Observed 4D seismic data showing normalized P-impedance changes for Field 2. Wells P1–P6 are producers, IW1 and IW2 are water injectors, IG1–IG4 are gas injectors, WAG1 and WAG2 are water-alternating-gas injectors.

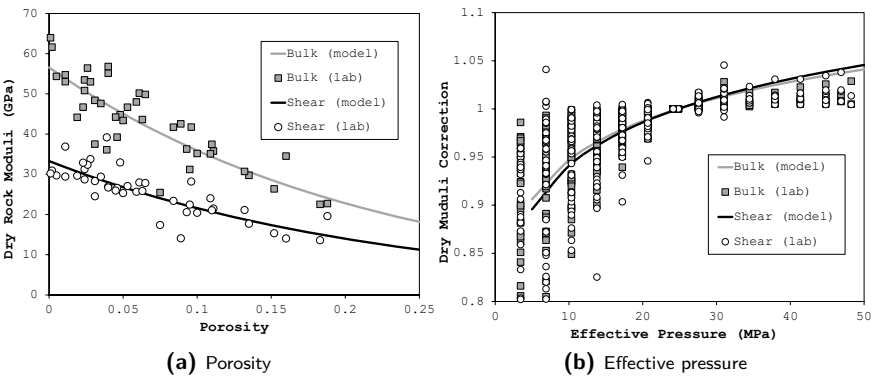


Fig. 9.28: Empirical correlations used in the PEM to estimate elastic properties of the dry rock frame for Field 2.

of $[-0.02, 0.02]$, which may appear small given the stiffness of carbonate rocks. However, the high-quality OBN acquisitions ensured well-defined time-lapse anomaly images. In the figure, hardening anomalies (blue scale) are primarily associated with water injection from wells IW1, IW2, and WAG1, while softening anomalies (red scale) relate to gas injection. Hardening around gas injector IG1 results from injected CO_2 , denser than reservoir fluid at current pressure. During data assimilation, seismic data-error covariance assumed a diagonal matrix with a constant standard deviation of 0.004. Emerick and Neto [125] considered the same field problem using the data projection method mentioned in Section 9.3.2 to estimate the data-

error covariance. For this particular example, the use of a diagonal matrix with constant variance resulted in similar results.

Fig. 9.29 shows the evolution of the data-mismatch objective functions for production and 4D seismic data. The average objective function for production data decreased from 120.0 to 13.0, while the objective function for 4D seismic data reduced from 6.5 to 2.0. It is noteworthy that the final production data objective function exceeds the 4.5 limit, corresponding to a three standard deviation, as discussed in Section 9.4.4. Additionally, tests with further ES-MDA iterations showed only marginal reductions. These results suggest that there is still room for improvement in the model by revisiting the initial steps of the process.

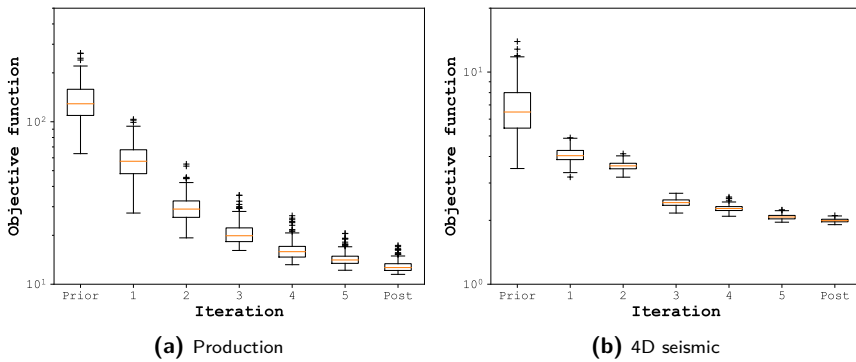


Fig. 9.29: Evolution of the data-mismatch objective function for Field 2.

Fig. 9.30 shows the porosity and horizontal permeability of the first layer in the first two realizations of the prior and posterior ensembles. Evident changes in the petrophysical properties are observed as a result of data assimilation, although the main characteristics of the prior realizations are preserved in the posterior ones. It is noteworthy that the two posterior models presented in this figure look remarkably similar to each other, indicating that a significant portion of the ensemble variability was lost in the process. Fig. 9.31 shows images of normalized variance, which corroborates this indication.

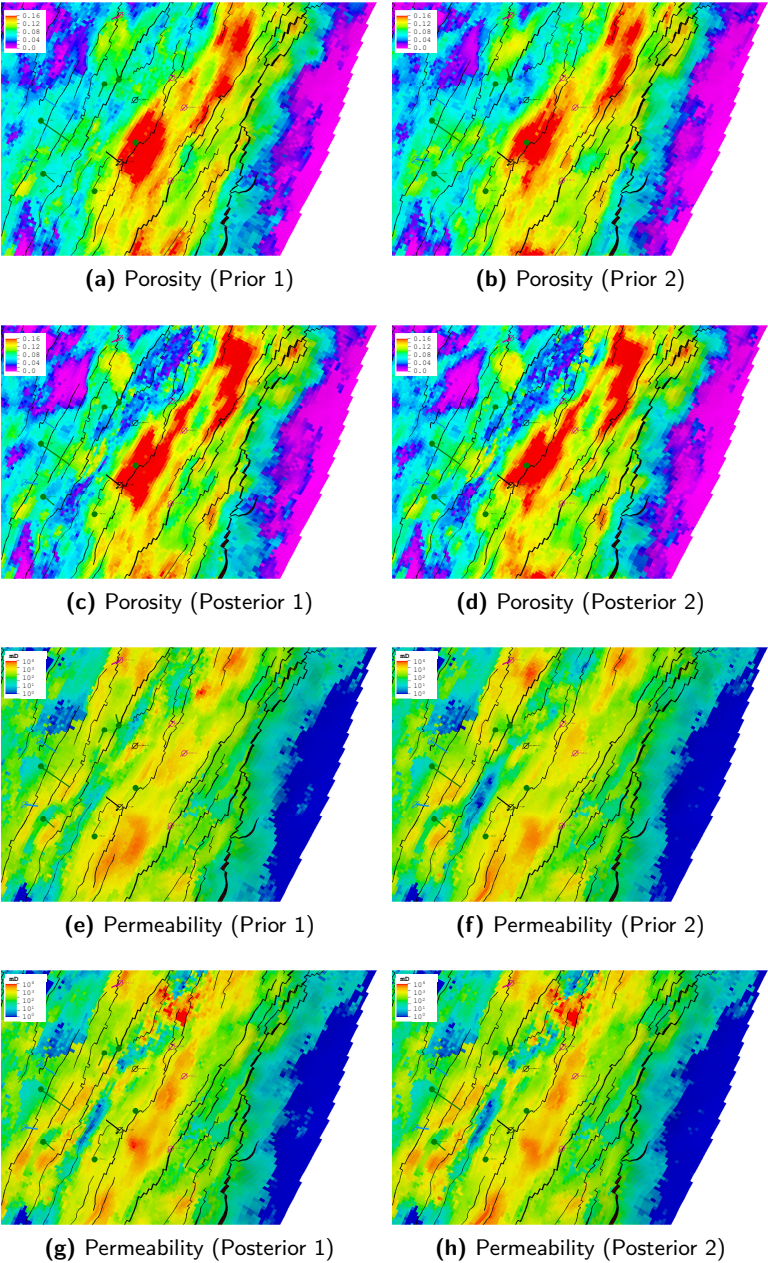


Fig. 9.30: First two realizations of porosity and horizontal permeability for Field 2.

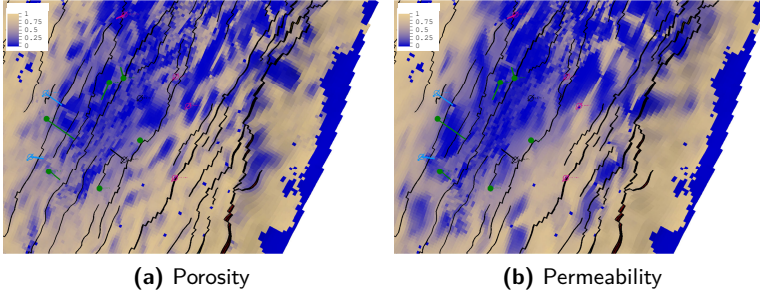


Fig. 9.31: Normalized variance of porosity and horizontal permeability for Field 2.

Fig. 9.32 illustrates two sets of oil-water relative permeability curves for this field. The final curves still show considerable variability compared to the prior ones. This result was achieved by applying an *ad hoc* localization coefficient of 0.1 for these parameters during data assimilation [123].

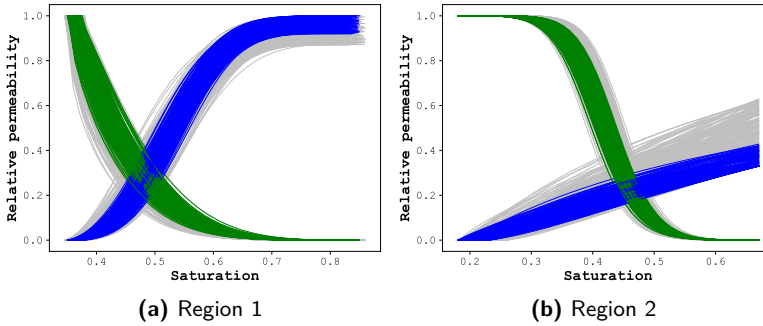


Fig. 9.32: Two oil-water relative permeability curves for Field 2. The blue and green lines represent the posterior water and oil relative permeability curves, respectively, while the grey lines indicate the prior estimates.

Fig. 9.33 presents production data from four wells alongside predictions using both prior and posterior ensembles, showing significant improvements in production data matches. Notably, the prior ensemble did not provide a particularly good starting point for the well data shown in Fig. 9.33. The overall observation coverage value, including all wells, is only 0.64, while the mean squared error is 9.97, indicating a biased prior ensemble. Analyzing individual wells, we observe that 40% have an observation coverage classified as “poor,” while 51% have a mean squared error classified as “biased.” These values suggest the need for a revision of the prior modeling. Despite these indications, ES-MDA was able to achieve reasonable data matches for the majority of wells in the field. The wells presented in Fig. 9.33 serves as a

good illustration of this achievement. Nevertheless, this does not imply that the entire data assimilation process is complete. In reality, reservoir modeling should be seen as a continuous process, with each cycle of assimilation revealing more about the field and generating feedback for the next round of models.

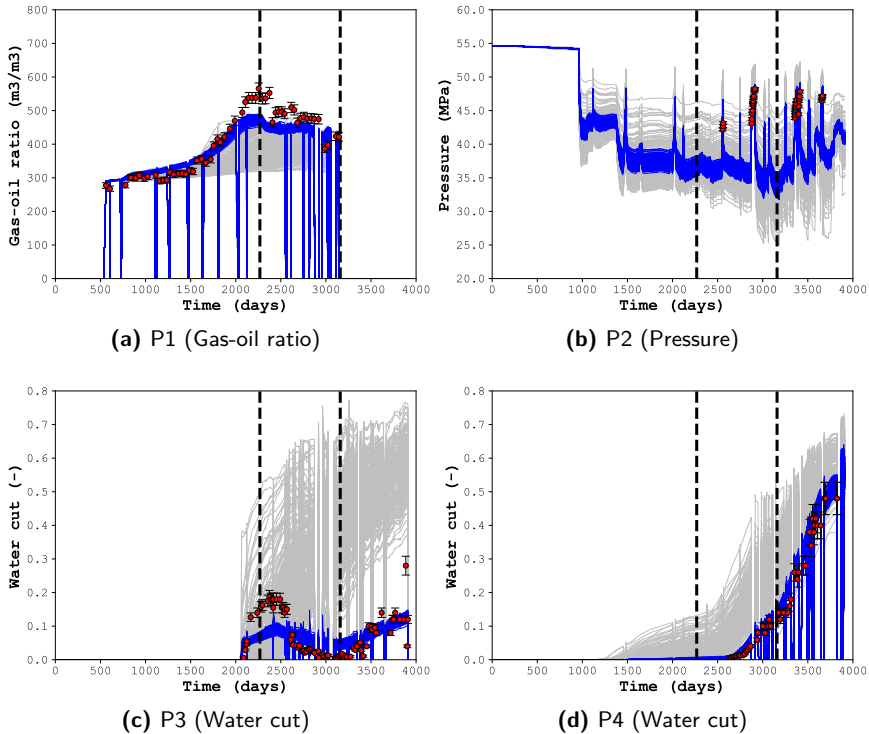


Fig. 9.33: Production data for four wells of Field 2. Red circles represent observed data, while gray and blue lines show predicted data from the prior and posterior ensembles, respectively. The error bars in the observed data points correspond to one standard deviation of the data error. Vertical dashed lines indicate the times of the base and monitor surveys.

Figs. 9.34b and 9.34c show the predicted 4D seismic data from the first two realizations of the ensemble before data assimilation, whereas Figs. 9.34d and 9.34e display the corresponding results after data assimilation. The posterior models exhibit significant improvements in the predicted 4D data, achieved through a better representation of overall pore pressure, a more accurate distribution of water-flooded areas, particularly near P4 and P5, and a better depiction of the region with increased gas saturation around well WAG2.

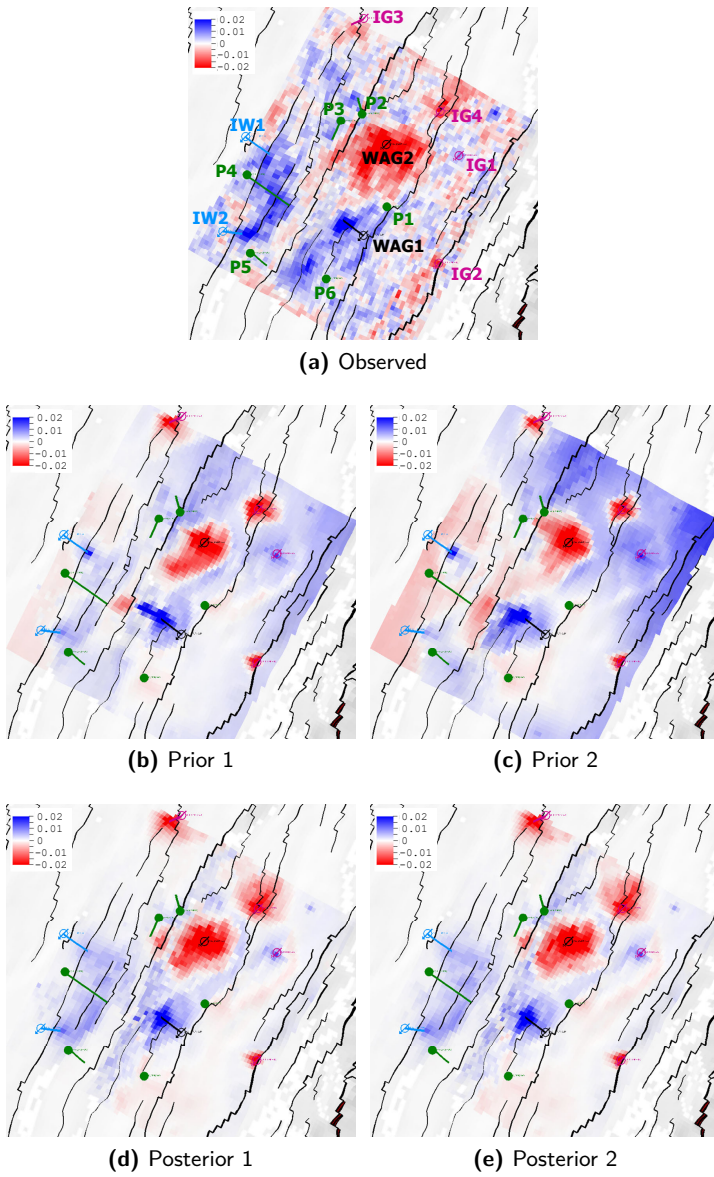


Fig. 9.34: Predicted P-impedance changes for the first two realizations before and after data assimilation in Field 2.

9.5.3 Field 3

The third field example is also an offshore oilfield located in the Campos Basin. In this field, the reservoir consists of a sequence of amalgamated channels deposited in a turbidite system, characterized by numerous faults and structural discontinuities. This field was selected as an example due to the inclusion of facies updates during data assimilation. The simulation model is composed of $147 \times 153 \times 47$ gridblocks, totaling 380,749 active gridblocks. The reservoir fluid is modeled using a black-oil formulation. Observed production data includes quarterly water cut measurements over a 38-year production period, along with a few isolated static pressure measurements.

The initial ensemble of facies was constructed by the field geologist using sequential indicator simulation (SIS) [7, 104] with locally varying azimuths [43]. These facies realizations were converted into a TPG parametrization following the procedure outlined in [123]. Fig. 9.35 illustrates the rock-type rule and the latent variables, $\mathbf{z}_1$ and $\mathbf{z}_2$, used to generate the facies realization. The truncation thresholds in the rock-type rule were determined by solving optimization problems to match the facies proportions of the original SIS realizations.

The data assimilation parameters include the latent variables $\mathbf{z}_1$ and $\mathbf{z}_2$, as well as the porosity, horizontal, and vertical permeability for the facies “coarse sand” and “sand.” The porosity and permeability for the facies “cemented” and “non-reservoir” are assumed to be zero. Other data assimilation parameters include scalar variables defining rock compressibility, the oil-water relative permeability curve, and the transmissibility multiplier across 126 faults.

The data assimilation used 16 ES-MDA iterations with constant inflation factors and Kalman gain localization to update an ensemble with 200 realizations. Fig. 9.36 shows the evolution of the data mismatch objective function. The average objective function reduced from an initial value of 18.4 to 3.1. Fig. 9.37 illustrates the improvements in the predicted water cut data for four wells in this field. For this field case, the overall observation coverage is 0.91, which is classified as “good” according to the criteria presented in Section 9.4.2. The mean squared error is 2.11, which is slightly above the threshold to be classified as “biased.” It is interesting to note that none of the three field examples presented in this chapter started the data assimilation with prior ensembles having observation coverage classified as “excellent” and mean squared error classified as “OK.” This underscores the challenges involved in modeling reservoirs in operational settings. In this third field example, the *ad hoc* procedure described in Section 9.3.1 was used to increase the data-error standard deviation. This effect is noticeable for Well 2 (Fig. 9.37b) in the period between 10,000 and 12,000 days, where larger error bars are shown.

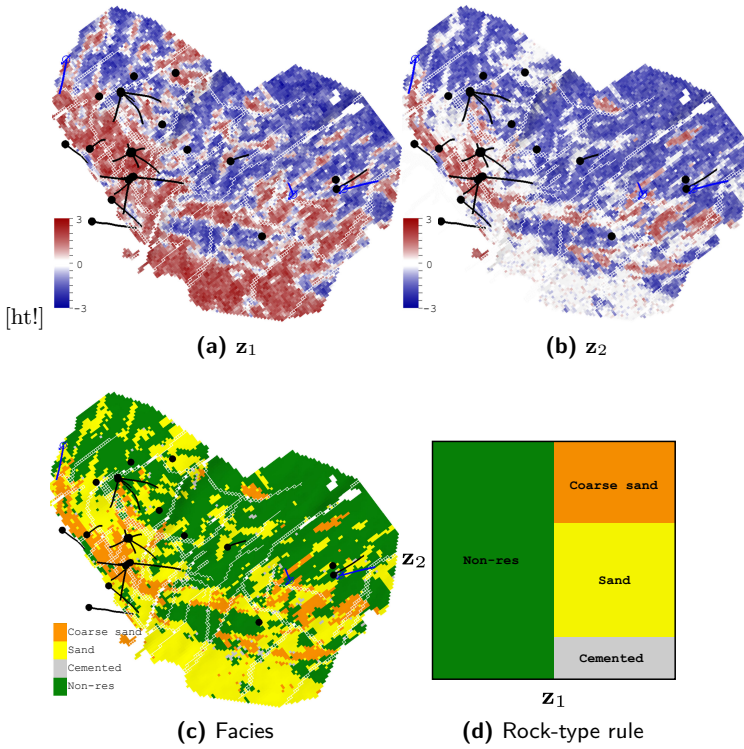


Fig. 9.35: Illustration of the TPG scheme used to parameterize the facies realizations with two latent variables, z_1 and z_2 . The rock-type rule in (d) is applied to truncate z_1 and z_2 , defining four facies types. The specific values of the truncation thresholds vary across the reservoir to reproduce a non-stationary distribution of facies proportions. Field 3.

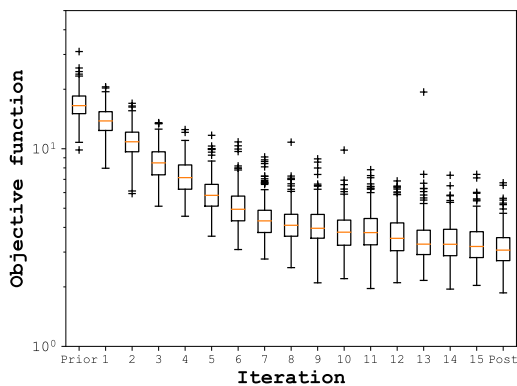


Fig. 9.36: Evolution of the data-mismatch objective function for Field 3.

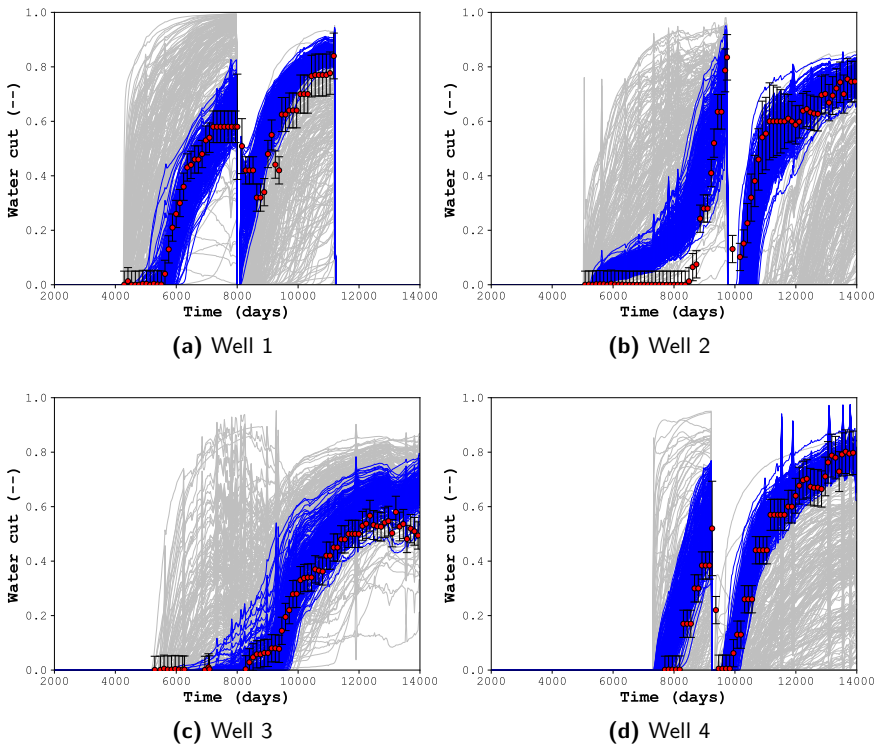


Fig. 9.37: Water cut data for four wells of Field 3. Red circles represent observed data, while gray and blue lines show predicted data from the prior and posterior ensembles, respectively. The error bars in the observed data points correspond to one standard deviation of the data error.

Fig. 9.38 displays the first two facies realizations before and after data assimilation. As with the other two field cases, data assimilation introduced noticeable changes in the realizations; however, the posterior models retain several characteristics of the prior ones. Notably, data assimilation increased the short-scale variation of the facies. Fig. 9.39 presents images showing the probability of occurrence of the facies “coarse sand” and “sand,” computed using the prior and posterior ensembles. This image shows a strong increase in the probability of occurrence of the facies “coarse sand” close to a cluster of wells on the left side of the model. This is an example of feedback from the calibration step that can be used to revise the prior modeling.

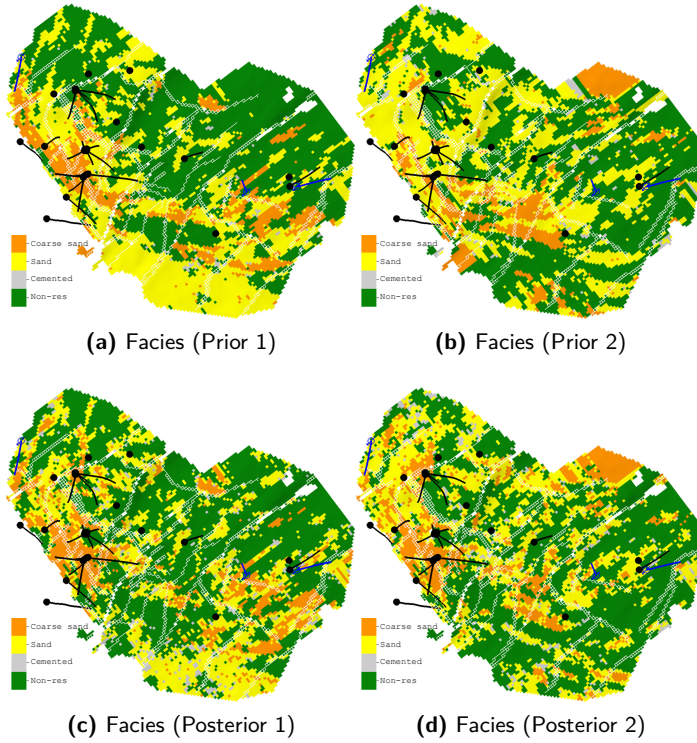


Fig. 9.38: First two facies realizations before and after data assimilation for Field 3.

Fig. 9.40 illustrates the relative permeability curves before and after data assimilation. The final ensemble essentially converged to nearly identical curves. This situation is common in reservoir data assimilation cases, as these parameters strongly impact scenarios involving breakthrough of injected fluids. In this particular field, this result does not seem surprising, given the 38-year production history with water breakthroughs in all oil-producing wells.

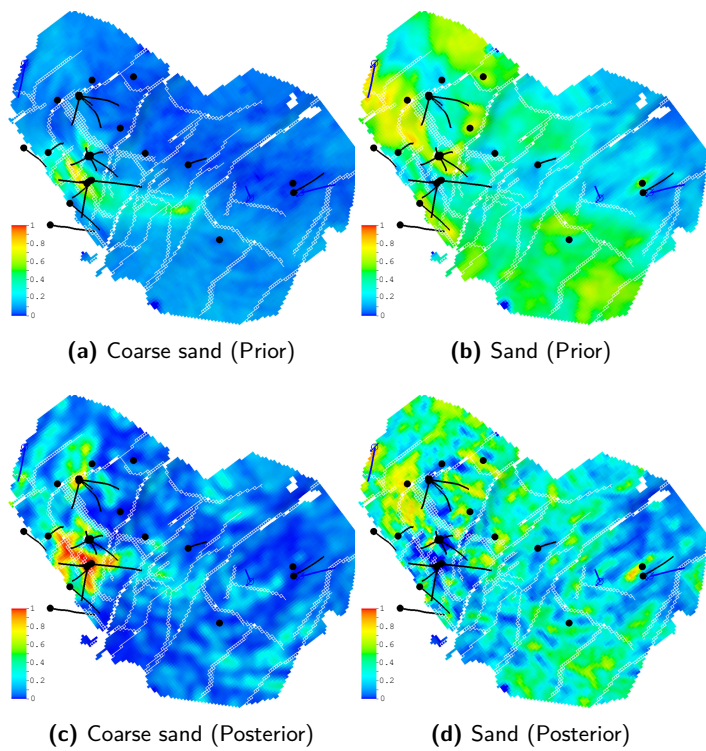


Fig. 9.39: Probability of occurrence for facies “coarse sand” and “sand” before and after data assimilation for Field 3.

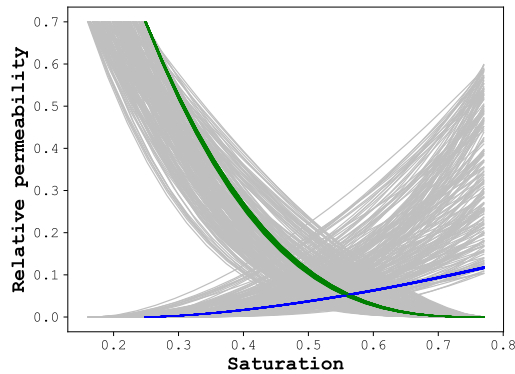


Fig. 9.40: Oil-water relative permeability curves for Field 3. The blue and green lines represent the posterior water and oil relative permeability curves, respectively, while the grey lines indicate the prior estimates.

9.6 Representative Models

As a final discussion in this chapter, it is important to address the use of representative models. A common practice in the industry involves selecting certain model realizations as “representative” of percentiles such as p10, p50, and p90 for decision-making.

While discussing results at the p10, p50, and p90 estimates is logical, the selection of representative models is highly questionable. The core issue with this approach is that different models can produce similar forecasts for one quantity while showing vastly different forecasts for other quantities of interest.

For example, Fig. 9.41a shows a plot of predicted cumulative oil production generated with an ensemble of model realizations for the PUNQ-S3 benchmark problem [150]. From this plot, three realizations were selected to represent the p10, p50, and p90 percentiles of cumulative production. The predicted cumulative productions of these models are highlighted in red, blue, and green in this plot. However, these realizations are no longer representative when analyzing another quantity of interest. For example, Fig. 9.41b shows a histogram of the original volume of oil in place (OIP) for the same ensemble. In this figure, the OIP values of the realizations representing cumulative production percentiles are indicated with colored vertical lines. Clearly, these realizations are not representative of the p10, p50, and p90 percentiles of OIP.

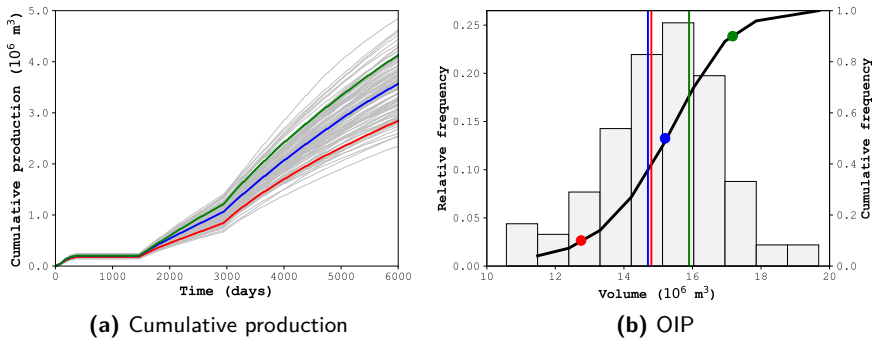


Fig. 9.41: Cumulative oil production and OIP for the PUNQ-S3 model. The red, blue and green lines in Panel (a) represents the predicted cumulative oil production from representative models of the percentiles p10, p50 and p90, respectively. The OIP of these models are indicated in the histogram in Panel (b) with the vertical lines. The red, blue and green circles indicate the p10, p50, and p90 values of OIP.

There are several publications in the literature proposing more robust strategies for selecting representative models from ensembles; see, e.g., [18,

394, 286] and references therein. The basic idea behind these methods is to select realizations considering multiple quantities of interest simultaneously.

However, the natural procedure in ensemble-based data assimilation is to utilize the entire ensemble of models to evaluate the uncertainty range for any quantity of interest. Nevertheless, fully adopting this ensemble-based approach can be more challenging than it appears. One major hurdle is the lack of methods and tools for analyzing and validating ensembles of models. Another challenge is the additional computational complexity involved in evaluating different operational conditions with ensembles. For example, optimization processes involving ensembles of models are exceptionally demanding in terms of computational resources [152, 397].



Elements of Linear Algebra

Abstract: *This appendix summarizes key linear algebra concepts relevant to reservoir data assimilation.*

A.1 Preliminaries

This book adopts a standard notation where lowercase x represents a scalar, lowercase bold $\mathbf{x}$ represents a vector, and uppercase bold $\mathbf{X}$ represents a matrix. The vectors are always column vectors:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}. \quad (\text{A.1})$$

Often, matrices are written in terms of their columns as

$$\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_N], \quad (\text{A.2})$$

where the j th column is the vector $\mathbf{x}_j$ and the entry (i, j) is the scalar x_{ij} .

The product between two matrices $\mathbf{C} = \mathbf{AB}$ is computed as

$$c_{ij} = \sum_k a_{ik} b_{k,j}. \quad (\text{A.3})$$

If $\mathbf{A}$ is $N_1 \times N_2$ and $\mathbf{B}$ is $N_2 \times N_3$, then $\mathbf{C}$ is $N_1 \times N_3$. Note that $\mathbf{AB} \neq \mathbf{BA}$, i.e., matrix multiplication is, in general, not commutative. Matrix multiplication is distributive, $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$, and associative, $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.

$\mathbf{A}^\top$ denotes the transpose of the matrix $\mathbf{A}$. If $\mathbf{A} = \mathbf{A}^\top$ we say that $\mathbf{A}$ is symmetric. The transpose of a product is

$$(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top, \quad (\text{A.4})$$

and the same is valid for vectors:

$$(\mathbf{x}^\top \mathbf{y})^\top = \mathbf{y}^\top \mathbf{x}. \quad (\text{A.5})$$

The inner (dot) product between two vectors is

$$\mathbf{x}^\top \mathbf{y} = [x_1 \cdots x_N] \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \sum_{i=1}^N x_i y_i. \quad (\text{A.6})$$

We prefer to use $\mathbf{x}^\top \mathbf{y}$ to denote the inner product instead of notations such as $\mathbf{x} \cdot \mathbf{y}$ and $\langle \mathbf{x}, \mathbf{y} \rangle$.

The matrix-vector product appears several times in this book

$$\mathbf{C}\mathbf{x} = \mathbf{y}. \quad (\text{A.7})$$

If the size of $\mathbf{x}$ is N_x and the size of $\mathbf{y}$ is N_y , then $\mathbf{C}$ is a $N_y \times N_x$ matrix. This product can be used to represent a system with N_y linear equations and N_x unknowns.

The inverse of a square matrix $\mathbf{C}$ is denoted by $\mathbf{C}^{-1}$ and it is defined as the matrix such that

$$\mathbf{C}^{-1}\mathbf{C} = \mathbf{C}\mathbf{C}^{-1} = \mathbf{I}. \quad (\text{A.8})$$

$\mathbf{I}$ is the identity matrix and has the same dimension of $\mathbf{C}$ and $\mathbf{C}^{-1}$.

Morrison-Woodbury formula: Let $\mathbf{C}$ be a $N_x \times N_x$ matrix, $\mathbf{U}$ and $\mathbf{V}$ be $N_x \times N_y$ matrices. Then, the following identity is true

$$(\mathbf{C} + \mathbf{UV}^\top)^{-1} = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{U}(\mathbf{I} + \mathbf{V}^\top \mathbf{C}^{-1}\mathbf{U})^{-1} \mathbf{V}^\top \mathbf{C}^{-1}. \quad (\text{A.9})$$

A.2 Vector Spaces

Definition 1: A set $\mathbb{V}$ over a field $\mathbb{F}$ is said to be a vector or a linear space if

1. For $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{V}$, $\mathbf{x} + \mathbf{y}$ is an element of $\mathbb{V}$ and $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
2. For any $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ in $\mathbb{V}$, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
3. There exists a unique vector $\mathbf{0}$ in $\mathbb{V}$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{V}$.

4. For each $\mathbf{x}$ in $\mathbb{V}$, there exists a unique element $-\mathbf{x}$ in $\mathbb{V}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
5. For every pair of scalars α and β , $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$ and $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ for every $\mathbf{x}$ in $\mathbb{V}$.
6. For every scalar α and pair of elements $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{V}$, $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$.

Elements of $\mathbb{V}$ are referred to as vectors. If $\mathbb{F}$ is the set of real numbers, $\mathbb{V}$ is said to be real vector space. The vector space we are interested in here is the set of all N_x -dimensional columns vectors in $\mathbb{R}^{N_x}$, i.e., $\mathbf{x} \in \mathbb{R}^{N_x}$.

Definition 2: A real valued function $\|\cdot\|$ defined on a vector space $\mathbb{V}$ is said to be a norm on $\mathbb{V}$ if

1. $\|\mathbf{x}\| \geq 0$ for every $\mathbf{x}$ in $\mathbb{V}$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
2. For every $\mathbf{x}$ in $\mathbb{V}$ and every scalar α , $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$.
3. For every $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{V}$, $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Some useful norms on $\mathbb{R}^{N_x}$:

- Euclidian norm (ℓ_2):

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{N_x} x_i^2 \right)^{1/2} = \sqrt{\mathbf{x}^\top \mathbf{x}}. \quad (\text{A.10})$$

- Maximum norm (ℓ_∞):

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq N_x} |x_i|. \quad (\text{A.11})$$

- p -norm (ℓ_p):

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{N_x} |x_i|^p \right)^{1/p}. \quad (\text{A.12})$$

- Mahalanobis distance [285]:

$$\|\mathbf{x}\|_{\mathbf{C}} = \sqrt{\mathbf{x}^\top \mathbf{C} \mathbf{x}}. \quad (\text{A.13})$$

with $\mathbf{C} \in \mathbb{R}^{N_x \times N_x}$.

Definition 3: The determinant of a square matrix $\mathbf{C}$, denoted $\det(\mathbf{C})$, is a function that maps $\mathbb{R}^{N_x \times N_x} \rightarrow \mathbb{R}$. The absolute value of $\det(\mathbf{C})$ quantifies the extent to which the matrix $\mathbf{C}$ scales the volume of space. A $\det(\mathbf{C}) = 0$ indicates that the matrix $\mathbf{C}$ collapses the space along at least one dimension, resulting in a loss of volume. Consequently, if $\det(\mathbf{C}) = 0$, the matrix $\mathbf{C}$ is singular. Conversely, if $\det(\mathbf{C}) = 1$, the matrix preserves the volume during transformation.

Definition 4: An eigenvalue and eigenvector of a square matrix $\mathbf{C} \in$

$\mathbb{R}^{N_x \times N_x}$ are a scalar λ and a non-zero vector $\mathbf{x} \in \mathbb{R}^{N_x}$ such that

$$\mathbf{C}\mathbf{x} = \lambda\mathbf{x}. \quad (\text{A.14})$$

This equation can be written as

$$(\mathbf{C} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}, \quad \mathbf{x} \neq \mathbf{0}. \quad (\text{A.15})$$

This means that the matrix $(\mathbf{C} - \lambda\mathbf{I})$ is singular, hence

$$\det(\mathbf{C} - \lambda\mathbf{I}) = 0. \quad (\text{A.16})$$

The last equation forms a polynomial equation with degree N_x . This means that $\mathbf{C}$ has N_x eigenvalues (roots of the polynomial), counting multiplicities.

Let $\mathbf{\Lambda}$ be the $N_x \times N_x$ diagonal matrix containing the eigenvalues of $\mathbf{C}$ and $\mathbf{X}$ be the $N_x \times N_x$ matrix with the corresponding eigenvectors in its columns. We can write

$$\mathbf{C}\mathbf{X} = \mathbf{X}\mathbf{\Lambda}. \quad (\text{A.17})$$

If all the eigenvectors are linearly independent, then $\mathbf{X}^{-1}$ exists and

$$\mathbf{C} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}, \quad (\text{A.18})$$

which is called eigenvalue decomposition of $\mathbf{C}$, also known as spectral decomposition.

In this book, we are particularly interested in covariance matrices. These matrices are:

- Real: $\mathbf{C} \in \mathbb{R}^{N_x \times N_x}$.
- Symmetric: $\mathbf{C} = \mathbf{C}^\top$.
- Non-singular: $\lambda_j > 0$, for $j = 1, \dots, N_x$.

The eigenvalue decomposition of a non-singular covariance matrices have a nice property: the eigenvectors are orthogonal, which mean that

$$\mathbf{C} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^\top, \quad (\text{A.19})$$

i.e., $\mathbf{X}^\top = \mathbf{X}^{-1}$. Note that orthogonality is equivalent to write

$$\mathbf{x}_i^\top \mathbf{x}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \quad (\text{A.20})$$

Definition 5: A real symmetric matrix $\mathbf{C}$ is said to be positive definite if for every $\mathbf{x} \in \mathbb{R}^{N_x}$, $\mathbf{x}^\top \mathbf{C} \mathbf{x} \geq 0$ and $\mathbf{x}^\top \mathbf{C} \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$. We use the notation

$$\mathbf{C} \succ \mathbf{O}, \quad (\text{A.21})$$

where $\mathbf{O}$ is the null matrix. $\mathbf{C}$ is said positive semidefinite if $\mathbf{x}^\top \mathbf{C} \mathbf{x} \geq 0$ with $\mathbf{x}^\top \mathbf{C} \mathbf{x} = 0$ for some $\mathbf{x} \neq \mathbf{0}$. We write

$$\mathbf{C} \succeq \mathbf{O}. \quad (\text{A.22})$$

Lemma 1: Let $\mathbf{C} \in \mathbb{R}^{N_x \times N_x}$:

- If $\mathbf{C}$ is real symmetric, then the eigenvalues $\{\lambda_j\}_{j=1}^{N_x}$ of $\mathbf{C}$ are all real and $\mathbf{C}$ has a set of orthonormal eigenvectors, $\{\mathbf{x}_j\}_{j=1}^{N_x}$.
- If $\mathbf{C}$ is positive definite, then all eigenvalues of $\mathbf{C}$ are positive and hence $\mathbf{C}$ is nonsingular.
- If $\mathbf{C}$ is real symmetric, then $\mathbf{C}$ is positive definite if and only if all eigenvalues of $\mathbf{C}$ are positive.
- $\det(\mathbf{C})$ is equal to the products of the eigenvalues of $\mathbf{C}$.

Definition 6: The null (or kernel) space of $\mathbf{C} \in \mathbb{R}^{N_x \times N_x}$ is the set of all vectors $\mathbf{x} \in \mathbb{R}^{N_x}$ that satisfy $\mathbf{C}\mathbf{x} = \mathbf{0}$. The null space of $\mathbf{C}$ is denoted by $\mathfrak{N}(\mathbf{C})$. The dimension of the null space is denoted by $\dim \mathfrak{N}(\mathbf{C})$ and it is referred to as the nullity of $\mathbf{C}$.

Definition 7: The range (or column space) of $\mathbf{C} \in \mathbb{R}^{N_y \times N_x}$ is the set of all vector $\mathbf{y} \in \mathbb{R}^{N_y}$ that there exists at least one $\mathbf{x} \in \mathbb{R}^{N_x}$ which satisfies $\mathbf{C}\mathbf{x} = \mathbf{y}$. The range is denoted by $\mathfrak{R}(\mathbf{C})$. The dimension of the range is denoted by $\dim \mathfrak{R}(\mathbf{C})$ and it is called the rank of $\mathbf{C}$. Two important results:

$$\dim \mathfrak{R}(\mathbf{C}) = \dim \mathfrak{R}(\mathbf{C}^\top) \quad (\text{A.23})$$

and

$$\dim \mathfrak{R}(\mathbf{C}) + \dim \mathfrak{N}(\mathbf{C}) = N_x. \quad (\text{A.24})$$

Note that N_x is the number of columns of $\mathbf{C}$.

For $j = 1, \dots, N_x$, let $\mathbf{c}_j$ denote the j th column of the $N_y \times N_x$ matrix $\mathbf{C}$, i.e.,

$$\mathbf{C} = [\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_{N_x}]. \quad (\text{A.25})$$

For $\mathbf{x} \in \mathbb{R}^{N_x}$

$$\mathbf{C}\mathbf{x} = [\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_{N_x}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_x} \end{bmatrix} = \sum_{j=1}^{N_x} x_j \mathbf{c}_j. \quad (\text{A.26})$$

Therefore $\mathbf{y} = \mathbf{C}\mathbf{x}$ means that $\mathbf{y}$ is a linear combination of the columns of $\mathbf{C}$.

Lemma 2: Let $\mathbf{C} \in \mathbb{R}^{N_y \times N_x}$ and $\mathbb{X} \subseteq \mathbb{R}^{N_x}$ be the vector space spanned by the columns of $\mathbf{C}$. Then,

- Every vector in $\mathfrak{R}(\mathbf{C})$ is a linear combination of the columns of $\mathbf{C}$ (vectors $\mathbf{c}_j$'s). In other words, the range of $\mathbf{C}$ is identical to the span of $\{\mathbf{c}_j\}_{j=1}^{N_x}$.

- If n denotes the number of linearly independent columns of $\mathbf{C}$, then $\dim \mathfrak{R}(\mathbf{C}) = \dim \mathbb{X} = n$ and $n \leq N_x$.
- In all cases, $\dim \mathfrak{R}(\mathbf{C}) \leq \min\{N_y, N_x\}$.
- The following three statements are equivalent:
 - (a) The set $\{\mathbf{c}_j\}_{j=1}^{N_x}$ is linearly independent.
 - (b) $\dim \mathfrak{R}(\mathbf{C}) = 0$.
 - (c) $\dim \mathfrak{R}(\mathbf{C}) = N_x$.

Lemma 3: Let $\mathbf{C} \in \mathbb{R}^{N_y \times N_x}$ be the matrix mapping $\mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_y}$:

- The number of linearly independent columns of $\mathbf{C}$ is equal to N_y , if and only if for every $\mathbf{y} \in \mathbb{R}^{N_y}$, there exists at least one $\mathbf{x} \in \mathbb{R}^{N_x}$ which satisfy $\mathbf{C}\mathbf{x} = \mathbf{y}$.
- If $N_x < N_y$, then there exists vectors $\mathbf{y} \in \mathbb{R}^{N_y}$ such that $\mathbf{C}\mathbf{x} = \mathbf{y}$ has no solution.

Lemma 4: The following statements are equivalent:

- $\mathbf{C}\mathbf{x} = \mathbf{y}$ has a unique solution for every $\mathbf{y} \in \mathbb{R}^{N_y}$.
- $\dim \mathfrak{R}(\mathbf{C}) = N_y$ and $\dim \mathfrak{N}(\mathbf{C}) = 0$.
- $\dim \mathfrak{R}(\mathbf{C}) = N_y$ and $N_y = N_x$ ($\mathbf{C}$ is square).

Lemma 5: $\mathbf{C}\mathbf{x} = \mathbf{y}$ has more than one solution for every $\mathbf{y} \in \mathbb{R}^{N_y}$ if and only if $\dim \mathfrak{R}(\mathbf{C}) = N_y$ and $\dim \mathfrak{N}(\mathbf{C}) > 0$.

A.3 Vector and Matrix Calculus

Let $\mathbf{x} \in \mathbb{R}^{N_x}$ be a column vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_x} \end{bmatrix} = [x_1 \ x_2 \ \cdots \ x_{N_x}]^\top \quad (\text{A.27})$$

The gradient operator is also a column vector

$$\nabla = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_{N_x}} \end{bmatrix}. \quad (\text{A.28})$$

The gradient ∇ acts as N_x -dimensional vector operator that can be applied to any $1 \times N_x$ matrices. Therefore, it does not make sense to write $\nabla \mathbf{x}$, but we can write

$$\begin{aligned}
 \nabla \mathbf{x}^\top &= \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_{N_x}} \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_{N_x} \end{bmatrix} \\
 &= \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_2}{\partial x_1} & \cdots & \frac{\partial x_{N_x}}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_2}{\partial x_2} & \cdots & \frac{\partial x_{N_x}}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial x_{N_x}} & \frac{\partial x_2}{\partial x_{N_x}} & \cdots & \frac{\partial x_{N_x}}{\partial x_{N_x}} \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \mathbf{I}.
 \end{aligned} \tag{A.29}$$

The gradient of a real-valued function $g(\mathbf{x})$ is a vector

$$\nabla g(\mathbf{x}) = \begin{bmatrix} \frac{\partial g(\mathbf{x})}{\partial x_1} \\ \frac{\partial g(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial g(\mathbf{x})}{\partial x_{N_x}} \end{bmatrix}, \tag{A.30}$$

and the Hessian matrix has the form

$$\mathbf{B} = \nabla \left[(\nabla g(\mathbf{x}))^\top \right]. \tag{A.31}$$

Let $\mathbf{y} = \mathbf{g}(\mathbf{x})$ be a real-valued vector function. Then

$$\begin{aligned}
 \nabla \mathbf{g}(\mathbf{x})^\top &= \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_{N_x}} \end{bmatrix} \begin{bmatrix} g_1(\mathbf{x}) & g_2(\mathbf{x}) & \cdots & g_{N_y}(\mathbf{x}) \end{bmatrix} \\
 &= \begin{bmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \frac{\partial g_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial g_{N_y}(\mathbf{x})}{\partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1(\mathbf{x})}{\partial x_{N_x}} & \frac{\partial g_2(\mathbf{x})}{\partial x_{N_x}} & \cdots & \frac{\partial g_{N_y}(\mathbf{x})}{\partial x_{N_x}} \end{bmatrix} \\
 &= \left[\nabla g_1(\mathbf{x}) \quad \nabla g_2(\mathbf{x}) \quad \cdots \quad \nabla g_{N_y}(\mathbf{x}) \right] \equiv \mathbf{G}^\top.
 \end{aligned} \tag{A.32}$$

Recall that it does not make sense to compute $\nabla \mathbf{g}(\mathbf{x})$.

We call the transpose of $\mathbf{G}^\top$ as the Jacobian or sensitivity matrix:

$$\mathbf{G} = \begin{bmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \frac{\partial g_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial g_1(\mathbf{x})}{\partial x_{N_x}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_{N_y}(\mathbf{x})}{\partial x_1} & \frac{\partial g_{N_y}(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial g_{N_y}(\mathbf{x})}{\partial x_{N_x}} \end{bmatrix}. \quad (\text{A.33})$$

Let $\mathbf{f} = \mathbf{f}(\mathbf{x})$ and $\mathbf{g} = \mathbf{g}(\mathbf{x})$ be N_x -dimensional vectors, both vectors are function of $\mathbf{x}$. Then

$$\nabla (\mathbf{f}^\top \mathbf{g}) = \nabla (\mathbf{f}^\top) \mathbf{g} + \nabla (\mathbf{g}^\top) \mathbf{f}. \quad (\text{A.34})$$

For example, if $\mathbf{b}$ is a constant vector, then

$$\nabla (\mathbf{b}^\top \mathbf{x}) = (\nabla \mathbf{b}^\top) \mathbf{x} + (\nabla \mathbf{x}^\top) \mathbf{b} = \mathbf{b} \quad (\text{A.35})$$

and

$$\nabla (\mathbf{x}^\top \mathbf{b}) = (\nabla \mathbf{x}^\top) \mathbf{b} + (\nabla \mathbf{b}^\top) \mathbf{x} = \mathbf{b}. \quad (\text{A.36})$$

Let $\mathbf{C}$ be a $N_x \times N_x$ constant matrix. Then

$$\begin{aligned} \nabla (\mathbf{x}^\top \mathbf{C} \mathbf{x}) &= (\nabla \mathbf{x}^\top) \mathbf{C} \mathbf{x} + [\nabla (\mathbf{C} \mathbf{x})]^\top \mathbf{x} \\ &= \mathbf{C} \mathbf{x} + [\nabla (\mathbf{x}^\top \mathbf{C}^\top)] \mathbf{x} \\ &= \mathbf{C} \mathbf{x} + [\nabla (\mathbf{x}^\top) \mathbf{C}^\top + \nabla (\mathbf{C}) \mathbf{x}] \mathbf{x} \\ &= \mathbf{C} \mathbf{x} + \mathbf{C}^\top \mathbf{x}. \end{aligned} \quad (\text{A.37})$$

If $\mathbf{C}$ is symmetric, i.e. $\mathbf{C} = \mathbf{C}^\top$, then $\nabla (\mathbf{x}^\top \mathbf{C} \mathbf{x}) = 2\mathbf{C} \mathbf{x}$.

Derivative with Respect to a Matrix: Let $\mathbf{X}$ be a $N_1 \times N_2$ matrix and $f = f(\mathbf{X})$ be a real-valued function of $\mathbf{X}$. Then, we define

$$\frac{\partial f}{\partial \mathbf{X}} = \left[\frac{\partial f}{\partial x_{ij}} \right], \quad (\text{A.38})$$

which is a $N_1 \times N_2$ matrix.

For example, let $\mathbf{X}$ be a $N_x \times N_x$ (square) and define

$$f = \text{tr}(\mathbf{X}) = \sum_{i=1}^{N_x} x_{ii}. \quad (\text{A.39})$$

Then,

$$\frac{\partial f}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{X})}{\partial \mathbf{X}} = \left[\frac{\partial \sum_{i=1}^{N_x} x_{ii}}{\partial x_{ij}} \right] = \mathbf{I}. \quad (\text{A.40})$$

Let $\mathbf{X}$ be a $N_1 \times N_2$ matrix and, $\mathbf{C}$ be a $N_2 \times N_1$ constant matrix. Then

$$\frac{\partial \text{tr}(\mathbf{X}\mathbf{C})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{C}^\top \mathbf{X}^\top)}{\partial \mathbf{X}} = \mathbf{C}^\top. \quad (\text{A.41})$$

Also, if $\mathbf{C}$ is $N_2 \times N_2$ real-symmetric constant matrix. Then and

$$\frac{\partial \text{tr}(\mathbf{X}\mathbf{C}\mathbf{X}^\top)}{\partial \mathbf{X}} = 2\mathbf{X}\mathbf{C}. \quad (\text{A.42})$$

A.4 Singular Value Decomposition

Let $\mathbf{C} \in \mathbb{R}^{N_y \times N_x}$. The singular value decomposition (SVD) of $\mathbf{C}$ is a decomposition as product of three matrices:

$$\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top, \quad (\text{A.43})$$

where $\mathbf{U} \in \mathbb{R}^{N_y \times N_y}$ is the matrix with left singular vectors of $\mathbf{C}$, $\mathbf{V} \in \mathbb{R}^{N_x \times N_x}$ is the matrix with right singular vectors of $\mathbf{C}$, and $\mathbf{\Sigma} \in \mathbb{R}^{N_y \times N_x}$ is the matrix containing the singular values of $\mathbf{C}$ in its main diagonal. All singular values of $\mathbf{C}$ are non-negative.

- If $N_y < N_x$:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{N_y} & 0 & \cdots & 0 \end{bmatrix} \quad (\text{A.44})$$

- If $N_y = N_x$:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{N_y} \end{bmatrix} \quad (\text{A.45})$$

- If $N_y > N_x$:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{N_x} \\ 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad (\text{A.46})$$

The matrices $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, i.e.,

$$\mathbf{U}\mathbf{U}^\top = \mathbf{U}^\top \mathbf{U} = \mathbf{I} \quad (\text{A.47})$$

and

$$\mathbf{V}\mathbf{V}^\top = \mathbf{V}^\top \mathbf{V} = \mathbf{I}. \quad (\text{A.48})$$

Therefore,

$$\mathbf{U}^\top = \mathbf{U}^{-1} \quad (\text{A.49})$$

and

$$\mathbf{V}^\top = \mathbf{V}^{-1}. \quad (\text{A.50})$$

The product $\mathbf{C}\mathbf{C}^\top$ is a $N_y \times N_y$ symmetric matrix and

$$\begin{aligned} \mathbf{C}\mathbf{C}^\top &= (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top) (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)^\top \\ &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top \mathbf{V}\mathbf{\Sigma}^\top \mathbf{U}^\top \\ &= \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^\top \mathbf{U}^\top. \end{aligned} \quad (\text{A.51})$$

Note that $\mathbf{\Sigma}\mathbf{\Sigma}^\top$ is a $N_y \times N_y$ diagonal matrix. Hence, $\mathbf{U}$ are eigenvectors of $\mathbf{C}\mathbf{C}^\top$ and $\lambda_i = \sigma_i^2$ is an eigenvalue of $\mathbf{C}\mathbf{C}^\top$.

The product $\mathbf{C}^\top \mathbf{C}$ is a $N_x \times N_x$ symmetric matrix and

$$\begin{aligned} \mathbf{C}^\top \mathbf{C} &= (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)^\top (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top) \\ &= \mathbf{V}\mathbf{\Sigma}\mathbf{U}^\top \mathbf{U}\mathbf{\Sigma}^\top \mathbf{V}^\top \\ &= \mathbf{V}\mathbf{\Sigma}^\top \mathbf{\Sigma}\mathbf{V}^\top. \end{aligned} \quad (\text{A.52})$$

Note that $\mathbf{\Sigma}^\top \mathbf{\Sigma}$ is a $N_x \times N_x$ diagonal matrix. Hence, $\mathbf{V}$ are eigenvectors of $\mathbf{C}^\top \mathbf{C}$ and $\lambda_i = \sigma_i^2$ is an eigenvalue of $\mathbf{C}^\top \mathbf{C}$.

As opposed to the eigenvalue decomposition, the SVD exists for rectangular and/or rank-deficient matrices. However, SVD and eigenvalue decomposition coincide for real symmetric positive-definite matrices. Let $\mathbf{C}$ be a real symmetric positive definite, then

$$\mathbf{C} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^\top \quad (\text{eigen-decomposition}) \quad (\text{A.53})$$

$$\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top \quad (\text{SVD}) \quad (\text{A.54})$$

and $\mathbf{U} = \mathbf{V} = \mathbf{X}$ and $\mathbf{\Sigma} = \mathbf{\Lambda}$.

A.5 Pseudoinverse

Let $\mathbf{C} \in \mathbb{R}^{N_y \times N_x}$. The Moore-Penrose pseudoinverse of a matrix $\mathbf{C}$ is defined as

$$\mathbf{C}^+ = \lim_{\alpha \rightarrow 0} (\mathbf{C}^\top \mathbf{C} + \alpha \mathbf{I})^{-1} \mathbf{C}^\top. \quad (\text{A.55})$$

Practical algorithms to compute $\mathbf{C}^+$ are based on SVD, in which case the pseudoinverse of $\mathbf{C}$ is

$$\mathbf{C}^+ = \mathbf{V}\mathbf{\Sigma}^+ \mathbf{U}^\top, \quad (\text{A.56})$$

where Σ^+ is the pseudoinverse of Σ , which is formed by replacing every nonzero diagonal entry by its reciprocal and transposing the resulting matrix.

Note that because of the orthogonality properties of $\mathbf{U}$ and $\mathbf{V}^\top$, we have

$$\begin{aligned}\mathbf{C}\mathbf{C}^+ &= (\mathbf{U}\Sigma\mathbf{V}^\top)(\mathbf{V}\Sigma^+\mathbf{U}^\top) \\ &= \mathbf{U}\Sigma\Sigma^+\mathbf{U}^\top \\ &= \mathbf{U}\mathbf{U}^\top \\ &= \mathbf{I}\end{aligned}\tag{A.57}$$

and

$$\mathbf{C}^+\mathbf{C} = \mathbf{I}.\tag{A.58}$$

Let $\mathbf{C}$ be a $N_y \times N_x$ positive-definite matrix, then the inverse $\mathbf{C}^{-1}$ exists. Also

$$\mathbf{C}\mathbf{C}^{-1} = \mathbf{C}^{-1}\mathbf{C} = \mathbf{I}.\tag{A.59}$$

In this particular case, $\mathbf{C}^{-1} = \mathbf{C}^+$ (the inverse and pseudoinverse of $\mathbf{C}$ coincide).

In many cases of practical interest, we want to keep only the N_r nonzero singular values (and the corresponding singular vectors). Note that if $\mathbf{C}$ is $N_y \times N_x$, then $N_r \leq \min\{N_y, N_x\}$. For the case with $N_y < N_x$, $\mathbf{C}$ can be decomposed as

$$\begin{aligned}\mathbf{C} &= \mathbf{U}\Sigma\mathbf{V}^\top \\ &= \begin{bmatrix} \mathbf{U}_r & \vdots & \mathbf{U}_0 \end{bmatrix} \begin{bmatrix} \Sigma_r & \vdots & \mathbf{0} \\ \cdots & \cdots & \cdots \\ \mathbf{0} & \vdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_r^\top \\ \cdots \\ \mathbf{V}_0^\top \end{bmatrix} \\ \mathbf{C} &= \mathbf{U}_r \Sigma_r \mathbf{V}_r^\top.\end{aligned}\tag{A.60}$$

$\mathbf{U}_r$ is $N_y \times N_r$ and $\mathbf{U}_r^\top \mathbf{U}_r = \mathbf{I}$, but $\mathbf{U}_r \mathbf{U}_r^\top \neq \mathbf{I}$. $\mathbf{V}_r$ is $N_x \times N_r$ and $\mathbf{V}_r^\top \mathbf{V}_r = \mathbf{I}$, but $\mathbf{V}_r \mathbf{V}_r^\top \neq \mathbf{I}$. Σ_r is a $N_r \times N_r$ matrix.

A.5.1 Range, Null Space and Rank

Another application of the SVD is that it provides an explicit representation of the range and null space of a matrix $\mathbf{C}$:

- $\mathbf{U}_r$ forms a basis for $\mathfrak{R}(\mathbf{C})$.
- $\mathbf{V}_0$ forms a basis for $\mathfrak{N}(\mathbf{C})$.

As a consequence, the rank of $\mathbf{C}$ is equal to the number of nonzero singular values of $\mathbf{C}$.

A.5.2 Poor Conditioning

The conditioning indicates how rapidly a function value changes with respect to small changes in its arguments. Consider the function

$$f(\mathbf{x}) = \mathbf{C}^{-1}\mathbf{x}. \quad (\text{A.61})$$

We define the condition number of the matrix $\mathbf{C}$, denoted as $\kappa(\mathbf{C})$, as the ratio of the magnitude of the largest and the smallest eigenvalues

$$\kappa(\mathbf{C}) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}. \quad (\text{A.62})$$

When $\kappa(\mathbf{C})$ is large, the function $f(\mathbf{x})$ becomes particularly sensitive to errors in the argument.

A.6 Square Root of a Matrix

Let $\mathbf{C}$ be a real $N_x \times N_x$ real symmetric positive definite matrix. $\mathbf{C}$ has only positive eigenvalues $\{\lambda_j\}_{j=1}^{N_x}$ with eigenvectors $\{\mathbf{x}_j\}_{j=1}^{N_x}$. Hence, we can write

$$\mathbf{C}\mathbf{X} = \mathbf{X}\mathbf{\Lambda}, \quad (\text{A.63})$$

where $\mathbf{X}$ is the $N_x \times N_x$ matrix with eigenvectors of $\mathbf{C}$ and $\mathbf{\Lambda}$ is the $N_x \times N_x$ diagonal matrix with eigenvalues of $\mathbf{C}$. Recall that $\mathbf{X}$ is an orthogonal matrix, i.e., $\mathbf{X}\mathbf{X}^\top = \mathbf{X}^\top\mathbf{X} = \mathbf{I}$. Therefore, we can write $\mathbf{C}$ as

$$\mathbf{C} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^\top. \quad (\text{A.64})$$

$\mathbf{\Lambda}$ is a diagonal matrix, with elements greater than zero. Hence, we can define the square root of $\mathbf{\Lambda}$, denoted by $\mathbf{\Lambda}^{1/2}$, as

$$\mathbf{\Lambda}^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_{N_x}} \end{bmatrix} \quad (\text{A.65})$$

So we can write $\mathbf{C}$ as

$$\begin{aligned} \mathbf{C} &= \mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{X}^\top \\ &= \mathbf{X}\mathbf{\Lambda}^{1/2}(\mathbf{X}^\top\mathbf{X})\mathbf{\Lambda}^{1/2}\mathbf{X}^\top \\ &= (\mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{X}^\top)(\mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{X}^\top) \\ &= (\mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{X}^\top)^2. \end{aligned} \quad (\text{A.66})$$

Therefore, we can define the square root of $\mathbf{C}$ as

$$\mathbf{C}^{1/2} = \mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{X}^\top \quad (\text{A.67})$$

and its inverse as

$$\mathbf{C}^{-1/2} = \left(\mathbf{C}^{1/2}\right)^{-1} = \mathbf{X}\mathbf{\Lambda}^{-1/2}\mathbf{X}^\top. \quad (\text{A.68})$$

Note that $\mathbf{C}^{1/2}$ and $\mathbf{C}^{-1/2}$ are real symmetric and positive definite matrices. More generally, we can define the square root of $\mathbf{C}$ a matrix $\mathbf{L}$ such that

$$\mathbf{C} = \mathbf{L}\mathbf{L}^\top. \quad (\text{A.69})$$

In particular, very often, we choose $\mathbf{L}$ to be the lower triangular matrix obtained from the Cholesky decomposition of $\mathbf{C}$.

Think of this as a breather for your eyes.



Elements of Linear Inverse Problems

Abstract: *This appendix reviews fundamental concepts of linear inverse problem theory. While data assimilation is inherently nonlinear, many methods are based on principles from linear problems. In particular, this appendix revisits key topics such as least squares and regularization.*

B.1 Discrete Linear Problems

Although physical problems are typically continuous, we focus our attention on discrete inverse problems, where we aim to estimate a finite set of model parameters. These parameters are described by an N_m -dimensional vector

$$\mathbf{m} = [m_1 \ m_2 \ \cdots \ m_{N_m}]^\top, \quad (\text{B.1})$$

given a N_d -dimensional vector of observations

$$\mathbf{d}_{\text{obs}} = [d_{\text{obs},1} \ d_{\text{obs},2} \ \cdots \ d_{\text{obs},N_d}]^\top. \quad (\text{B.2})$$

We refer to linear problems as those in which the theoretical relationship between data and model parameters takes the form

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) = \mathbf{G}\mathbf{m}, \quad (\text{B.3})$$

where $\mathbf{G}$ is the $N_d \times N_m$ sensitivity matrix.

For convenience, we define the vector of model parameters as an element of the linear vector space $\mathbb{R}^{N_m}$, referred to as the model space. Similarly, the set of all possible vectors of observed data belongs to a data space, which is assumed to be $\mathbb{R}^{N_d}$.

The sensitivity matrix $\mathbf{G}$ maps the model space into the data space, i.e., $\mathbf{G} : \mathbb{R}^{N_m} \rightarrow \mathbb{R}^{N_d}$. Recall that the null space of $\mathbf{G}$, $\mathfrak{N}(\mathbf{G})$, is the set of all vectors in the model space such that $\mathbf{G}\mathbf{m} = \mathbf{0}$. In contrast, the range of $\mathbf{G}$, $\mathfrak{R}(\mathbf{G})$, is the set of all vectors $\mathbf{d}$ in the data space such that there is at least one $\mathbf{m}$ satisfying $\mathbf{G}\mathbf{m} = \mathbf{d}$.

Let $\mathbf{g}_i$ denote the i th column of the $N_d \times N_m$ sensitivity matrix $\mathbf{G}$

$$\mathbf{G} = [\mathbf{g}_1 \ \mathbf{g}_2 \ \cdots \ \mathbf{g}_{N_m}]. \quad (\text{B.4})$$

Then, for a given N_m -dimensional vector $\mathbf{m}$, we have

$$\mathbf{d} = \mathbf{G}\mathbf{m} = [\mathbf{g}_1 \ \mathbf{g}_2 \ \cdots \ \mathbf{g}_{N_m}] \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{N_m} \end{bmatrix} = \sum_{i=1}^{N_m} m_i \mathbf{g}_i. \quad (\text{B.5})$$

This result shows that every vector $\mathbf{d}$ in the range of $\mathbf{G}$ is a linear combination of the columns of $\mathbf{G}$.

Recall that if the columns of $\mathbf{G}$ are linearly independent, then there are no nontrivial solutions to $\mathbf{G}\mathbf{m} = \mathbf{0}$ and vice versa. (This also means that $N_m \leq N_d$). If the number of linearly independent columns of $\mathbf{G}$ is equal to N_d , then there is at least one vector $\mathbf{m}$ which satisfies $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$. If $N_m < N_d$, then there exist data vectors $\mathbf{d}_{\text{obs}}$ such that the equation $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ has no solution. $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ has a unique solution for every $\mathbf{d}_{\text{obs}}$ in the data space if all of the rows of $\mathbf{G}$ are independent and $N_m = N_d$ (i.e., $\mathbf{G}$ is a nonsingular square matrix). In this case, $\mathbf{m} = \mathbf{G}^{-1}\mathbf{d}_{\text{obs}}$.

B.2 Ill-Posed Problems

Practical inverse problems are typically ill-posed. To define an ill-posed problem, it is convenient to first define a well-posed problem [182]. A problem is considered well-posed if it satisfies the following conditions:

1. There exists a solution to the problem (existence).
2. There is at most one solution to the problem (uniqueness).
3. The solution depends continuously on the data (stability).

A problem that does not satisfy these criteria is termed ill-posed.

B.3 Classification of Linear Inverse Problems

1. **Purely underdetermined problems:** The inverse problem $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ is considered purely underdetermined if, for every $\mathbf{d}_{\text{obs}}$ in the data space,

there exists more than one vector $\mathbf{m}$ in the model space that satisfies the equation. This condition holds true if and only if the following two conditions are satisfied:

- (i) There are more model variables than data ($N_m > N_d$, i.e., $\dim \mathfrak{N}(\mathbf{G}) > 0$).
 - (ii) All of the rows of $\mathbf{G}$ are linearly independent (the data are independent, i.e., $\dim \mathfrak{R}(\mathbf{G}) = N_d$).
2. **Purely overdetermined problems:** The inverse problem $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ is considered purely overdetermined if the dimension of the null space of $\mathbf{G}$ is 0 and there exist $\mathbf{d}_{\text{obs}}$ for which the equation has no solution. In simpler terms, there is no vector $\mathbf{m}$ that satisfies $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$. However, it is still desirable to find "solutions" $\mathbf{m}$ to the inverse problem, meaning vectors $\mathbf{m}$ that approximately satisfy $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ in some sense. If the dimension of the range of $\mathbf{G}$ is less than N_d , then there exists at least one $\mathbf{d}_{\text{obs}}$ for which $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ has no solution. It can be demonstrated that the following three conditions are equivalent:
- (i) The $N_m \times N_m$ real-symmetric matrix $\mathbf{G}^\top \mathbf{G}$ is a nonsingular positive-definite matrix.
 - (ii) $\dim \mathfrak{R}(\mathbf{G}) = N_m$.
 - (iii) $\dim \mathfrak{N}(\mathbf{G}) = 0$.
3. **Mixed determined problems:** The inverse problem of $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ is termed mixed-determined if it exhibits characteristics of both underdetermined and overdetermined problems. This implies that some parameters are overdetermined while others are underdetermined. Establishing a mixed-determined problem involves demonstrating that it does not fall strictly into the category of either purely underdetermined or purely overdetermined problems. It is noteworthy that the majority of practical inverse problems encountered in various fields are mixed determined.

B.4 Solutions for Discrete Linear Inverse Problems

B.4.1 Least Squares

Consider the inverse linear problem $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$, where $N_d > N_m$, indicating more data points than unknown parameters. In such cases, exact solutions are typically non-existent. One approach is to seek the least-squares solution ($\mathbf{m}_{\text{ls}}$):

$$\begin{aligned}
\mathbf{m}_{\text{ls}} &= \arg \min_{\mathbf{m}} \|\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}\|_2^2 \\
&= \arg \min_{\mathbf{m}} \left[(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) \right] \\
&= \arg \min_{\mathbf{m}} \mathcal{O}(\mathbf{m}).
\end{aligned} \tag{B.6}$$

We refer to $\mathcal{O}(\mathbf{m})$ as the objective function. To find $\mathbf{m}_{\text{ls}}$ we need to find the minimum of $\mathcal{O}(\mathbf{m})$. First, we compute the gradient of $\mathcal{O}(\mathbf{m})$

$$\begin{aligned}
\nabla \mathcal{O}(\mathbf{m}) &= \nabla \left[(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) \right] \\
&= -2\mathbf{G}^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}).
\end{aligned} \tag{B.7}$$

Then, to compute $\mathbf{m}_{\text{ls}}$, we set $\nabla \mathcal{O}(\mathbf{m}) = \mathbf{0}$ and solve for $\mathbf{m}$

$$-2\mathbf{G}^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) = \mathbf{0}, \tag{B.8}$$

or

$$\mathbf{G}^\top \mathbf{G}\mathbf{m} = \mathbf{G}^\top \mathbf{d}_{\text{obs}}. \tag{B.9}$$

If the $N_m \times N_m$ matrix $\mathbf{G}^\top \mathbf{G}$ is nonsingular, implying $\dim \mathfrak{R}(\mathbf{G}) = N_m$, then we encounter an overdetermined problem, and the least-squares solution becomes:

$$\mathbf{m}_{\text{ls}} = (\mathbf{G}^\top \mathbf{G})^{-1} \mathbf{G}^\top \mathbf{d}_{\text{obs}}. \tag{B.10}$$

B.4.2 Weighted Least Squares

Consider the same inverse linear problem $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ and assume that $N_d > N_m$, which typically has no exact solution. We may give a different weight for each data point when computing the least-squares estimate. For example, we may want to give higher weight to more reliable data points. In this case, we define the weighted least-squares solution:

$$\begin{aligned}
\mathbf{m}_{\text{wls}} &= \arg \min_{\mathbf{m}} \|\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}\|_{\mathbf{W}}^2 \\
&= \arg \min_{\mathbf{m}} \left[(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top \mathbf{W} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) \right] \\
&= \arg \min_{\mathbf{m}} \mathcal{O}(\mathbf{m}),
\end{aligned} \tag{B.11}$$

where $\mathbf{W}$ is a diagonal matrix containing the data weights.

To find $\mathbf{m}_{\text{wls}}$ we use the same procedure:

$$\begin{aligned}\nabla \mathcal{O}(\mathbf{m}) &= \nabla \left[(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top \mathbf{W} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) \right] \\ &= -2\mathbf{G}^\top \mathbf{W} (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) = \mathbf{0}.\end{aligned}\tag{B.12}$$

Hence,

$$\mathbf{G}^\top \mathbf{W} \mathbf{G} \mathbf{m} = \mathbf{G}^\top \mathbf{W} \mathbf{d}_{\text{obs}}.\tag{B.13}$$

If the $N_m \times N_m$ matrix $\mathbf{G}^\top \mathbf{W} \mathbf{G}$ is nonsingular, i.e., $\dim \mathfrak{R}(\mathbf{G}) = N_m$, then we have an overdetermined problem and

$$\mathbf{m}_{\text{wls}} = (\mathbf{G}^\top \mathbf{W} \mathbf{G})^{-1} \mathbf{G}^\top \mathbf{W} \mathbf{d}_{\text{obs}}.\tag{B.14}$$

B.4.3 Constrained Least Squares

Now, let's consider the scenario where $N_d < N_m$, i.e., more parameters than data, a common situation in reservoir data assimilation. Suppose $\dim R(\mathbf{G}) = N_d$. This implies that $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ has a solution for all $\mathbf{d}_{\text{obs}}$, actually presenting an infinite number of solutions (an underdetermined problem). Suppose that we have a prior estimate of $\mathbf{m}$, denoted as $\mathbf{m}_{\text{pr}}$. For instance, $\mathbf{m}_{\text{pr}}$ could represent the best guess provided by a geologist regarding our model. In this case, we might be interested in finding $\mathbf{m}$ that satisfies the problem $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$ while staying close to the prior estimate, $\mathbf{m}_{\text{pr}}$. This problem can be formulated as the following constrained least-squares problem:

$$\begin{aligned}\min \quad & \|\mathbf{m} - \mathbf{m}_{\text{pr}}\|_2^2 \\ \text{s.t.} \quad & \mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}.\end{aligned}\tag{B.15}$$

Constrained optimization problems can be solved using the method of Lagrange multipliers¹, in which case we define the Lagrangian function

¹ The method of Lagrange multipliers is designed to solve constrained optimization problems of the form

$$\begin{aligned}\min \quad & \mathcal{O}(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{f}(\mathbf{x}) = \mathbf{0}.\end{aligned}$$

To solve this problem, first we define the Lagrangian function

$$\mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathcal{O}(\mathbf{x}) - \boldsymbol{\lambda}^\top \mathbf{f}(\mathbf{x}),$$

set $\nabla_{\mathbf{x}} \mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$, $\nabla_{\boldsymbol{\lambda}} \mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$, and solve for $\mathbf{x}$ and $\boldsymbol{\lambda}$.

$$\mathcal{L}(\mathbf{m}, \boldsymbol{\lambda}) = (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top (\mathbf{m} - \mathbf{m}_{\text{pr}}) - \boldsymbol{\lambda}^\top (\mathbf{G}\mathbf{m} - \mathbf{d}_{\text{obs}}), \quad (\text{B.16})$$

and solve

$$\nabla_{\mathbf{m}} \mathcal{L}(\mathbf{m}, \boldsymbol{\lambda}) = 2(\mathbf{m} - \mathbf{m}_{\text{pr}}) - \mathbf{G}^\top \boldsymbol{\lambda} = \mathbf{0}, \quad (\text{B.17})$$

for $\mathbf{m}$, which leads to

$$\mathbf{m} = \mathbf{m}_{\text{pr}} + \frac{1}{2} \mathbf{G}^\top \boldsymbol{\lambda}. \quad (\text{B.18})$$

Repeating the process for the gradient with respect to $\boldsymbol{\lambda}$ results in

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{m}, \boldsymbol{\lambda}) = \mathbf{G}\mathbf{m} - \mathbf{d}_{\text{obs}} = \mathbf{0}, \quad (\text{B.19})$$

or $\mathbf{G}\mathbf{m} - \mathbf{d}_{\text{obs}} = \mathbf{0}$. Using (B.17) results in

$$\mathbf{G} \left(\mathbf{m}_{\text{pr}} + \frac{1}{2} \mathbf{G}^\top \boldsymbol{\lambda} \right) - \mathbf{d}_{\text{obs}} = \mathbf{0}, \quad (\text{B.20})$$

which leads to

$$\mathbf{G}\mathbf{G}^\top \boldsymbol{\lambda} = 2(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}}). \quad (\text{B.21})$$

If the rank of the $N_d \times N_d$ matrix $\mathbf{G}\mathbf{G}^\top$ is N_d (underdetermined problem), then $(\mathbf{G}\mathbf{G}^\top)^{-1}$ exists and

$$\boldsymbol{\lambda} = 2(\mathbf{G}\mathbf{G}^\top)^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}}). \quad (\text{B.22})$$

Using this result in (B.17) leads to the final estimate

$$\mathbf{m}_{\text{cls}} = \mathbf{m}_{\text{pr}} + \mathbf{G}^\top (\mathbf{G}\mathbf{G}^\top)^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}}). \quad (\text{B.23})$$

B.4.4 Regularized Least Squares

In real-world scenarios, $\mathbf{d}_{\text{obs}}$ is always corrupted with noise stemming from measurement errors. Consequently, there is no need to enforce an estimate $\mathbf{m}$ that precisely satisfies $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$. Additionally, we may have some prior knowledge about the solution, aiding in the regularization of the inverse problem. The most straightforward approach to integrate these two aspects is by defining the following objective function:

$$\begin{aligned} \mathcal{O}(\mathbf{m}) &= \alpha \|\mathbf{m} - \mathbf{m}_{\text{pr}}\|_2^2 + \|\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}\|_2^2 \\ &= \alpha (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top (\mathbf{m} - \mathbf{m}_{\text{pr}}) + (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}), \end{aligned} \quad (\text{B.24})$$

where $\alpha > 0$ is a scalar.

To minimize $\mathcal{O}(\mathbf{m})$, we set $\nabla \mathcal{O}(\mathbf{m}) = \mathbf{0}$ and solve for $\mathbf{m}$.

$$\begin{aligned}\nabla \mathcal{O}(\mathbf{m}) &= \alpha (\mathbf{m} - \mathbf{m}_{\text{pr}}) - \mathbf{G}^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) = \mathbf{0} \\ &= \alpha \mathbf{I} (\mathbf{m} - \mathbf{m}_{\text{pr}}) - \mathbf{G}^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m} - \mathbf{G}\mathbf{m}_{\text{pr}} + \mathbf{G}\mathbf{m}_{\text{pr}}) = \mathbf{0},\end{aligned}\quad (\text{B.25})$$

or

$$(\alpha \mathbf{I} + \mathbf{G}^\top \mathbf{G}) (\mathbf{m} - \mathbf{m}_{\text{pr}}) = \mathbf{G}^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}}). \quad (\text{B.26})$$

Note that $\alpha \mathbf{I} + \mathbf{G}^\top \mathbf{G} = (\alpha \mathbf{I} + \mathbf{G}^\top \mathbf{G})^\top$. Hence, $\alpha \mathbf{I} + \mathbf{G}^\top \mathbf{G}$ is real symmetric. Moreover, this is also positive definite because

$$\begin{aligned}\mathbf{x}^\top (\alpha \mathbf{I} + \mathbf{G}^\top \mathbf{G}) \mathbf{x} &= \alpha \mathbf{x}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{G}^\top \mathbf{G} \mathbf{x} \\ &= \alpha \|\mathbf{x}\|_2^2 + \|\mathbf{G}\mathbf{x}\|_2^2 \geq \alpha \|\mathbf{x}\|_2^2 > 0, \text{ for } \mathbf{x} \neq \mathbf{0},\end{aligned}\quad (\text{B.27})$$

which means that $\mathbf{x}^\top (\alpha \mathbf{I} + \mathbf{G}^\top \mathbf{G}) \mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$. Therefore, this matrix has a real inverse and the regularized least-squares estimate becomes

$$\mathbf{m}_{\text{rls}} = \mathbf{m}_{\text{pr}} + (\alpha \mathbf{I} + \mathbf{G}^\top \mathbf{G})^{-1} \mathbf{G}^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}_{\text{pr}}). \quad (\text{B.28})$$

By selecting $\alpha > 0$, we can compute the regularized least-squares solution. As α approaches 0, $\mathbf{m}_{\text{rls}}$ converges to $\mathbf{m}_{\text{ls}}$, the classical least-squares solution. The magnitude of α dictates the relative importance of $\|\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}\|_2^2$ and $\|\mathbf{m} - \mathbf{m}_{\text{pr}}\|_2^2$ in the objective function. With increasing α , the weight of $\|\mathbf{m} - \mathbf{m}_{\text{pr}}\|_2^2$ amplifies, while $\|\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}\|_2^2$ tends to rise, leading to a decrease in the fidelity of the data match.

B.4.5 Tikhonov Regularization

Named after the Russian mathematician Andrey Tikhonov, the Tikhonov regularization method combines the least-squares solution with a regularization term of the form

$$\begin{aligned}\mathbf{m}_{\text{tk}} &= \arg \min_{\mathbf{m}} \{ \|\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}\|_2^2 + \|\mathbf{L}(\mathbf{m} - \mathbf{m}_{\text{pr}})\|_2^2 \} \\ &= \arg \min_{\mathbf{m}} \left[(\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^\top (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) + \right. \\ &\quad \left. (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{L}^\top \mathbf{L} (\mathbf{m} - \mathbf{m}_{\text{pr}}) \right] \\ &= \arg \min_{\mathbf{m}} \mathcal{O}(\mathbf{m}),\end{aligned}\quad (\text{B.29})$$

where $\mathbf{L}$ is called the Tikhonov matrix. If we set $\mathbf{L} = \sqrt{\alpha} \mathbf{I}$, for $\alpha > 0$, we return to the standard regularized least squares. This procedure is closely related to the Levenberg-Marquardt optimization algorithm. To find $\mathbf{m}_{\text{tk}}$, we use the same procedure. We set $\nabla \mathcal{O}(\mathbf{m}) = \mathbf{0}$ and solve for $\mathbf{m}$.

$$\mathbf{m}_{\text{tk}} = \mathbf{m}_{\text{pr}} + (\mathbf{L}^\top \mathbf{L} + \mathbf{G}^\top \mathbf{G})^{-1} \mathbf{G}^\top (\mathbf{d}_{\text{obs}} - \mathbf{G} \mathbf{m}_{\text{pr}}). \quad (\text{B.30})$$

If we choose $\mathbf{L}$ such that $\mathbf{L}^\top \mathbf{L}$ is positive definite, then $(\mathbf{L}^\top \mathbf{L} + \mathbf{G}^\top \mathbf{G})^{-1}$ exists.

B.4.6 SVD Solution

SVD can be useful to construct approximate solutions of linear inverse problems. Recall that the $N_d \times N_m$ matrix $\mathbf{G}$ can be decomposed as the product of three matrices

$$\mathbf{G} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top,$$

where $\mathbf{U}$ is a $N_d \times N_d$ matrix with left singular vectors of $\mathbf{G}$, $\mathbf{V}$ is a $N_m \times N_m$ matrix with right singular vectors of $\mathbf{G}$, $\mathbf{\Sigma}$ is a $N_d \times N_m$ matrix with nonzero elements only in the diagonal.

If $\mathbf{G}$ is square and nonsingular, the solution of the inverse problem $\mathbf{G} \mathbf{m} = \mathbf{d}_{\text{obs}}$ can be obtained using

$$\begin{aligned} \mathbf{G} \mathbf{m} &= \mathbf{d}_{\text{obs}} \\ (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top) \mathbf{m} &= \mathbf{d}_{\text{obs}} \\ \mathbf{U}^\top \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top \mathbf{m} &= \mathbf{U}^\top \mathbf{d}_{\text{obs}} \\ \mathbf{\Sigma}^{-1} \mathbf{\Sigma} \mathbf{V}^\top \mathbf{m} &= \mathbf{\Sigma}^{-1} \mathbf{U}^\top \mathbf{d}_{\text{obs}} \\ \mathbf{V} \mathbf{V}^\top \mathbf{m} &= \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^\top \mathbf{d}_{\text{obs}} \\ \mathbf{m} &= \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^\top \mathbf{d}_{\text{obs}}, \end{aligned} \quad (\text{B.31})$$

where we use the fact that $\mathbf{U}^\top \mathbf{U} = \mathbf{I}$, $\mathbf{V} \mathbf{V}^\top = \mathbf{I}$ and $\mathbf{\Sigma}^{-1} \mathbf{\Sigma} = \mathbf{I}$. In this case, it is trivial to compute $\mathbf{\Sigma}^{-1}$ because this is a diagonal matrix. Actually, if $\mathbf{G}$ is square and nonsingular, $\mathbf{G}^{-1}$ exists and $\mathbf{G}^{-1} = \mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^\top$. (SVD gives the true inverse of $\mathbf{G}$).

In general, however, $\mathbf{G}$ is not square and some singular values are zero (or others may be close to zero). In this case, we consider the solution

$$\mathbf{m}_{\text{svd}} = \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^\top \mathbf{d}_{\text{obs}}, \quad (\text{B.32})$$

where $N_r \leq \min\{N_d, N_m\}$ is the number of nonzero singular values. $\mathbf{U}_r$ is $N_d \times N_r$ and $\mathbf{U}_r^\top \mathbf{U}_r = \mathbf{I}$, but $\mathbf{U}_r \mathbf{U}_r^\top \neq \mathbf{I}$. $\mathbf{V}_r$ is $N_m \times N_r$ and $\mathbf{V}_r^\top \mathbf{V}_r = \mathbf{I}$, but $\mathbf{V}_r \mathbf{V}_r^\top \neq \mathbf{I}$. $\mathbf{\Sigma}_r$ is a $N_r \times N_r$ matrix².

Note that (B.32) has the form

² The SVD solution has interesting properties. For example, if $\mathbf{G} \mathbf{m} = \mathbf{d}_{\text{obs}}$ has more than one solution, the SVD solution is the one with minimum length. In contrast, if $\mathbf{G} \mathbf{m} = \mathbf{d}_{\text{obs}}$ has no solution, the SVD solution is the least-squares solution.

$$\mathbf{m}_{\text{svd}} = \mathbf{V}_r \mathbf{b}, \quad (\text{B.33})$$

where $\mathbf{b} = \Sigma_r^{-1} \mathbf{U}_r^\top \mathbf{d}_{\text{obs}}$ is a N_r -dimensional vector. This implies that $\mathbf{m}_{\text{svd}}$ is a linear combination of the first N_r right-singular vectors of $\mathbf{G}$, i.e.,

$$\begin{aligned} \mathbf{m}_{\text{svd}} &= \mathbf{V}_r \mathbf{b} \\ &= [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_{N_r}] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{N_r} \end{bmatrix} \\ &= \sum_{j=1}^{N_r} \mathbf{v}_j b_j \\ &= \sum_{j=1}^{N_r} \mathbf{v}_j \left(\frac{\mathbf{u}_j^\top \mathbf{d}_{\text{obs}}}{\sigma_j} \right). \end{aligned} \quad (\text{B.34})$$

B.4.6.1 Removing Small Singular Values

In practical applications, we often remove the small singular values even if they are nonzero. The typical criterion is to keep only the $N_r \leq N_R = \{N_d, N_m\}$ largest singular values such that

$$\frac{\sum_{j=1}^{N_r} \sigma_j}{\sum_{j=1}^{N_R} \sigma_j} \geq \xi, \quad (\text{B.35})$$

where $\xi \leq 1$. Note that the condition (B.35) assumes that the singular values are sorted in a decreasing order.

To comprehend the rationale behind truncating the singular values, let's consider the SVD solution of the problem $\mathbf{G}\mathbf{m} = \mathbf{d}_{\text{obs}}$:

$$\mathbf{m}_{\text{svd}} = \mathbf{V}_r \Sigma_r^{-1} \mathbf{U}_r^\top \mathbf{d}_{\text{obs}}. \quad (\text{B.36})$$

Observed data always contain noise. So, even if we knew the ground truth values for $\mathbf{m}$, denoted as $\mathbf{m}_{\text{true}}$, $\mathbf{G}\mathbf{m}_{\text{true}}$ would not return $\mathbf{d}_{\text{obs}}$. Instead, we could express it as:

$$\mathbf{d}_{\text{obs}} = \mathbf{G}\mathbf{m}_{\text{true}} + \mathbf{e}. \quad (\text{B.37})$$

where $\mathbf{e}$ is the noise. Therefore,

$$\begin{aligned}
\mathbf{m}_{\text{svd}} &= \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^\top [\mathbf{G} \mathbf{m}_{\text{true}} + \mathbf{e}] \\
&= \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^\top [\mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^\top \mathbf{m}_{\text{true}} + \mathbf{e}] \\
&= \mathbf{V}_r \mathbf{V}_r^\top \mathbf{m}_{\text{true}} + \mathbf{V}_r \begin{bmatrix} \frac{1}{\sigma_1} \mathbf{u}_1^\top \mathbf{e} \\ \vdots \\ \frac{1}{\sigma_r} \mathbf{u}_r^\top \mathbf{e} \end{bmatrix}. \tag{B.38}
\end{aligned}$$

Suppose that $\mathbf{m}_{\text{true}}$ lies in the span of $\mathbf{V}_r$, i.e., we can write $\mathbf{m}_{\text{true}} = \mathbf{V}_r \mathbf{b}$. Hence

$$\begin{aligned}
\mathbf{m}_{\text{svd}} &= \mathbf{V}_r \mathbf{V}_r^\top \mathbf{V}_r \mathbf{b} + \mathbf{V}_r \begin{bmatrix} \frac{1}{\sigma_1} \mathbf{u}_1^\top \mathbf{e} \\ \vdots \\ \frac{1}{\sigma_r} \mathbf{u}_r^\top \mathbf{e} \end{bmatrix} \\
&= \mathbf{V}_r \begin{bmatrix} b_1 + \frac{1}{\sigma_1} \mathbf{u}_1^\top \mathbf{e} \\ \vdots \\ b_r + \frac{1}{\sigma_r} \mathbf{u}_r^\top \mathbf{e} \end{bmatrix}. \tag{B.39}
\end{aligned}$$

This outcome suggests that the noise vector is scaled by the singular values. Consequently, small singular values may amplify errors in estimating $\mathbf{m}$. Conversely, decreasing the number of singular values, along with their corresponding singular vectors, diminishes the capability to replicate the true model, as the columns of $\mathbf{V}_r$ will no longer fully span the model space.

B.4.6.2 Optimal Hard Threshold of Singular Values

Gavish and Donoho [165] derived the optimal threshold to truncate singular values, τ_* , in the case of estimating a low-rank matrix, $\mathbf{G}$, from an observed matrix, $\tilde{\mathbf{G}}$, corrupted with white noise of level $\gamma > 0$, i.e.,

$$\tilde{\mathbf{G}} = \mathbf{G} + \gamma \mathbf{Z}, \tag{B.40}$$

where the noise matrix $\mathbf{Z}$ has normality distributed zero-mean and unity-variance entries. They found that for a square $N \times N$ matrix $\mathbf{G}$, the optimal threshold is

$$\tau_* = \frac{4}{\sqrt{3}} \sqrt{N} \gamma. \tag{B.41}$$

If the noise level γ is unknown, they proposed approximations for τ_* in the form

$$\tau_* \approx \omega \sigma_{\text{med}}, \tag{B.42}$$

where σ_{med} is the median singular value of $\tilde{\mathbf{G}}$ and $\omega = 2.858$ if $\mathbf{G}$ is square. They also derived a procedure to compute ω for rectangular matrices.

B.4.6.3 Relation Between Regularization and SVD

Consider we are looking for the following regularized least-squares estimate

$$\begin{aligned}\mathbf{m}_{\text{rls}} &= \arg \min_{\mathbf{m}} \{ \|\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}\|_2^2 + \alpha \|\mathbf{m}\|_2^2 \} \\ &= \arg \min_{\mathbf{m}} \mathcal{O}(\mathbf{m}).\end{aligned}\tag{B.43}$$

with $\alpha > 0$. Setting $\nabla \mathcal{O}(\mathbf{m}) = \mathbf{0}$, we obtain

$$(\alpha \mathbf{I} + \mathbf{G}^\top \mathbf{G}) \mathbf{m} = \mathbf{G}^\top \mathbf{d}_{\text{obs}}.\tag{B.44}$$

Now the SVD of $\mathbf{G}$ and write

$$\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top.\tag{B.45}$$

Using this result

$$\begin{aligned}(\alpha \mathbf{I} + \mathbf{G}^\top \mathbf{G}) \mathbf{m} &= \mathbf{G}^\top \mathbf{d}_{\text{obs}} \\ (\alpha \mathbf{V}\mathbf{V}^\top + \mathbf{V}\mathbf{\Sigma}^\top \mathbf{\Sigma} \mathbf{V}^\top) \mathbf{m} &= \mathbf{V}\mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{d}_{\text{obs}} \\ \mathbf{V} (\alpha \mathbf{I} + \mathbf{\Sigma}^\top \mathbf{\Sigma}) \mathbf{V}^\top \mathbf{m} &= \mathbf{V}\mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{d}_{\text{obs}} \\ \mathbf{V}^\top \mathbf{V} (\alpha \mathbf{I} + \mathbf{\Sigma}^\top \mathbf{\Sigma}) \mathbf{V}^\top \mathbf{m} &= \mathbf{V}^\top \mathbf{V} \mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{d}_{\text{obs}} \\ \underbrace{(\alpha \mathbf{I} + \mathbf{\Sigma}^\top \mathbf{\Sigma})}_{\text{diagonal matrix}} \mathbf{V}^\top \mathbf{m} &= \mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{d}_{\text{obs}}.\end{aligned}\tag{B.46}$$

Therefore,

$$\begin{aligned}\mathbf{m}_{\text{rls}} &= \mathbf{V} (\alpha \mathbf{I} + \mathbf{\Sigma}^\top \mathbf{\Sigma})^{-1} \mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{d}_{\text{obs}} \\ &= \sum_{j=1}^{N_r} \mathbf{v}_j \left(\frac{\sigma_j \mathbf{u}_j^\top \mathbf{d}_{\text{obs}}}{\sigma_j^2 + \alpha} \right).\end{aligned}\tag{B.47}$$

Note that

$$\lim_{\alpha \rightarrow 0} \sum_{j=1}^{N_r} \mathbf{v}_j \left(\frac{\sigma_j \mathbf{u}_j^\top \mathbf{d}_{\text{obs}}}{\sigma_j^2 + \alpha} \right) = \sum_{j=1}^{N_r} \mathbf{v}_j \left(\frac{\mathbf{u}_j^\top \mathbf{d}_{\text{obs}}}{\sigma_j} \right) = \mathbf{m}_{\text{ls}},\tag{B.48}$$

which is the standard least-squares estimate.

We also can interpret the regularization effect by noting that

$$\mathbf{v}_j \left(\frac{\sigma_j \mathbf{u}_j^\top \mathbf{d}_{\text{obs}}}{\sigma_j^2 + \alpha} \right) \approx \begin{cases} \mathbf{0}, & \text{if } 0 \approx \sigma_j \ll \alpha \\ \mathbf{v}_j \left(\frac{\mathbf{u}_j^\top \mathbf{d}_{\text{obs}}}{\sigma_j} \right), & \text{if } \sigma_j \gg \alpha \end{cases} \quad (\text{B.49})$$

Therefore, the regularization term $\alpha \|\mathbf{m}\|_2^2$ functions similarly to the elimination of small singular values. Contributions from large singular values ($\sigma_j \gg \alpha$) remain (almost) unaffected, while contributions from small singular values ($\sigma_j \ll \alpha$) are (almost) nullified.



Elements of Probability and Geostatistics

Abstract: *This appendix summarizes essential concepts of probability and geostatistics pertinent to reservoir data assimilation.*

C.1 Basic Concepts

C.1.1 Random Variables

A random variable¹ x is a quantity that takes on varying random values. Random variables can be either discrete, assuming a finite (countable) set of values, or continuous, associated with real numbers.

C.1.2 Probability Mass Function

The probability of a discrete random variable is characterized by a probability mass function (PMF), denoted as $P(x)$, which must satisfy the following conditions:

1. The domain of $P(x)$ is the set of all possible values of x .
2. $0 \leq P(x) \leq 1$. An impossible event has $P(x) = 0$ and a certain event has $P(x) = 1$.
3. $\sum_x P(x) = 1$.

¹ This book departs from the conventional notation in the probability literature, where capital letters represent random variables.

C.1.3 Probability Density Function

For continuous random variables, the probability density function (PDF), denoted by $p(x)$, is used and must satisfy the following conditions:

1. The domain of $p(x)$ is the set of all possible values of x .
2. $p(x) \geq 0$. (Note that we do not require $p(x) \leq 1$.)
3. $\int_{-\infty}^{\infty} p(x)dx = 1$.

The PDF does not directly provide the probability of a specific event. In fact, for a continuous random variable x , the probability of x taking any exact value a is zero, i.e., $P(x = a) = 0$. Instead, we compute the probability of x falling within the interval between a and b as:

$$P(a \leq x < b) = \int_a^b p(x)dx. \quad (\text{C.1})$$

C.1.4 Cumulative Density Function

The cumulative density function (CDF) for a random variable x with PDF $p(x)$ is defined as

$$P(x \leq a) = \int_{-\infty}^a p(x)dx. \quad (\text{C.2})$$

C.1.5 Joint Probability

Let $x_1, \dots, x_N$ be a set of random variables with density $p(\cdot)$ such that the joint probability that a realization of the set lies within the region $a_1 \leq x_1 < b_1, \dots, a_N \leq x_N < b_N$ is

$$P(a_1 \leq x_1 < b_1, \dots, a_N \leq x_N < b_N) = \int_{a_1}^{b_1} \cdots \int_{a_N}^{b_N} p(x_1, \dots, x_N) dx_1 \dots dx_N. \quad (\text{C.3})$$

We represent this set of random variables as a random vector, $\mathbf{x} = [x_1 \cdots x_N]^\top$. The joint probability of $\mathbf{x}$ be within the region Ω is

$$P(\mathbf{x} \in \Omega) = \int_{\Omega} p(\mathbf{x})d\mathbf{x}. \quad (\text{C.4})$$

C.1.6 Marginal PDF

The marginal PDF for a random variable x_i is found by integrating the joint PDF over the remaining $N - 1$ variables. For example, the marginal density for x_1 is

$$p(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, x_2, \dots, x_N) dx_2 \dots dx_N. \quad (\text{C.5})$$

C.1.7 Conditional PDF

The conditional PDF of x given y is defined by

$$p(x|y) = \frac{p(x, y)}{p(y)}. \quad (\text{C.6})$$

The same is valid for random vectors, i.e.,

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})}. \quad (\text{C.7})$$

C.1.8 The Chain Rule of Conditional Probabilities

The joint PDF can be decomposed as a product of conditionals:

$$\begin{aligned} p(x_1, x_2, \dots, x_N) &= p(x_1) p(x_2|x_1) p(x_3|x_1, x_2) \cdots p(x_N|x_1, \dots, x_{N-1}) \\ &= p(x_1) \prod_{i=2}^N p(x_i|x_1, \dots, x_{i-1}). \end{aligned} \quad (\text{C.8})$$

This result follows from the definition of conditional probabilities and the proof uses mathematical induction. First, we note that the chain rule holds for $N = 2$:

$$p(x_1, x_2) = p(x_1) p(x_2|x_1). \quad (\text{C.9})$$

Then, we assume it holds for $N - 1$. For convenience, call $\hat{\mathbf{x}} = [x_1 \cdots x_{N-1}]^T$

$$p(x_1, \dots, x_{N-1}) = p(\hat{\mathbf{x}}) = p(x_1) \prod_{i=2}^{N-1} p(x_i|x_1, \dots, x_{i-1}). \quad (\text{C.10})$$

Finally, we show it holds for N

$$\begin{aligned}
 p(x_1, \dots, x_N) &= p(\hat{\mathbf{x}}, x_N) = p(x_N | \hat{\mathbf{x}}) p(\hat{\mathbf{x}}) \\
 &= p(x_N | x_1, \dots, x_{N-1}) p(x_1) \prod_{i=2}^{N-1} p(x_i | x_1, \dots, x_{i-1}) \\
 &= p(x_1) \prod_{i=2}^N p(x_i | x_1, \dots, x_{i-1}). \tag{C.11}
 \end{aligned}$$

Hence, by mathematical induction the chain rule of conditional probabilities holds. The chain rule is a basic component to justify the sequential data assimilation used in the Kalman filter.

C.1.9 Expected Value

The expected value of a function $g(\cdot)$ of a random variable x whose PDF is $p(x)$ is given by

$$\mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(x) p(x) dx. \tag{C.12}$$

The expected value of a random variable x is also called the mean value and is given by

$$\mathbb{E}[x] = \mu_x = \int_{-\infty}^{\infty} xp(x) dx. \tag{C.13}$$

The expected value of the product of two random variables x is

$$\mathbb{E}[xy] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp(x, y) dx dy. \tag{C.14}$$

C.1.10 Independent Random Variables

Two random variables x and y are said to be independent if

$$p(x, y) = p(x) p(y). \tag{C.15}$$

If x and y are independent then

$$\begin{aligned}
 \mathbb{E}[xy] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp(x) p(y) dx dy \\
 &= \left(\int_{-\infty}^{\infty} xp(x) dx \right) \left(\int_{-\infty}^{\infty} yp(y) dy \right) \\
 &= \mathbb{E}[x] \mathbb{E}[y].
 \end{aligned} \tag{C.16}$$

C.1.11 Variance

The variance of a random variable x is

$$\begin{aligned}
 \mathbb{V}[x] &= \mathbb{E}[(x - \mathbb{E}[x])^2] \\
 &= \mathbb{E}[(x - \mu_x)^2] \\
 &= \mathbb{E}[x^2] - 2\mathbb{E}[x\mu_x] + \mathbb{E}[\mu_x^2] \\
 &= \mathbb{E}[x^2] - 2\mu_x \mathbb{E}[x] + \mu_x^2 \\
 &= \mathbb{E}[x^2] - \mu_x^2 \\
 &= \sigma_x^2,
 \end{aligned} \tag{C.17}$$

where σ_x is the standard deviation of x .

The variance of the sum of two random variables, x and y , can be calculated using the following formula:

$$\begin{aligned}
 \mathbb{V}[x + y] &= \mathbb{E}[(x + y - \mathbb{E}[x + y])^2] \\
 &= \mathbb{E}[(x + y - (\mu_x + \mu_y))^2] \\
 &= \mathbb{E}[(x - \mu_x + y - \mu_y)^2] \\
 &= \mathbb{E}[(x - \mu_x)^2 + (y - \mu_y)^2 + 2(x - \mu_x)(y - \mu_y)] \\
 &= \mathbb{V}[x] + \mathbb{V}[y] + 2\mathbb{E}[(x - \mu_x)(y - \mu_y)] \\
 &= \mathbb{V}[x] + \mathbb{V}[y] + 2\mathbb{C}[x, y],
 \end{aligned} \tag{C.18}$$

where $\mathbb{C}[x, y]$ is the covariance of x and y defined as

$$\mathbb{C}[x, y] = \mathbb{E}[(x - \mu_x)(y - \mu_y)]. \tag{C.19}$$

C.1.12 Covariance

The covariance of x and y can be written as

$$\begin{aligned}
 \mathbb{C}[x, y] &= \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])] \\
 &= \mathbb{E}[(x - \mu_x)(y - \mu_y)] \\
 &= \mathbb{E}[xy - x\mu_y - y\mu_x + \mu_x\mu_y] \\
 &= \mathbb{E}[xy] - \mu_x\mu_y - \mu_y\mu_x + \mu_x\mu_y \\
 &= \mathbb{E}[xy] - \mu_x\mu_y.
 \end{aligned} \tag{C.20}$$

If x and y are independent, meaning $\mathbb{E}[xy] = \mathbb{E}[x]\mathbb{E}[y]$, then $\mathbb{C}[x, y] = 0$. It is important to note that while $\mathbb{C}[x, y] = 0$ indicates that x and y are uncorrelated, this does not imply that they are independent. However, for multivariate Gaussian distributions, uncorrelated random variables are indeed independent.

C.1.13 Correlation Coefficient

We define the Pearson correlation coefficient $\rho_{xy} \in [-1, 1]$ as

$$\rho_{xy} = \frac{\mathbb{C}[x, y]}{\sigma_x \sigma_y}. \tag{C.21}$$

The correlation coefficient measures a linear correlation between two quantities.

C.1.14 Covariance Matrix

Let $\mathbf{x}$ be a random vector. The covariance of $\mathbf{x}$, denoted as $\mathbf{C}_x$, is defined as

$$\begin{aligned}
 \mathbf{C}_x &= \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top] \\
 &= \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^\top] \\
 &= \mathbb{E}[\mathbf{x}\mathbf{x}^\top - \boldsymbol{\mu}_x\mathbf{x}^\top - \mathbf{x}\boldsymbol{\mu}_x^\top + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top] \\
 &= \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \mathbb{E}[\boldsymbol{\mu}_x\mathbf{x}^\top] - \mathbb{E}[\mathbf{x}\boldsymbol{\mu}_x^\top] + \mathbb{E}[\boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top] \\
 &= \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top \\
 &= \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top.
 \end{aligned} \tag{C.22}$$

Therefore, the covariance of $\mathbf{x}$ is a matrix (covariance matrix) with the form

$$\mathbf{C}_{\mathbf{x}} = \begin{bmatrix} \mathbb{E}[x_1 x_1] - \mu_1 \mu_1 & \cdots & \mathbb{E}[x_1 x_N] - \mu_1 \mu_N \\ \mathbb{E}[x_2 x_1] - \mu_2 \mu_1 & \cdots & \mathbb{E}[x_2 x_N] - \mu_2 \mu_N \\ \vdots & \vdots & \vdots \\ \mathbb{E}[x_N x_1] - \mu_N \mu_1 & \cdots & \mathbb{E}[x_N x_N] - \mu_N \mu_N \end{bmatrix}. \quad (\text{C.23})$$

If the components x_i and x_j are uncorrelated, i.e., $\mathbb{E}[x_i x_j] = \mu_i \mu_j$, $\mathbf{C}_{\mathbf{x}}$ is a diagonal matrix. The diagonal elements of $\mathbf{C}_{\mathbf{x}}$ are $\mathbb{V}[x_i] = \sigma_{x_i}^2$.

Similarly, the (cross-)covariance of two random vectors, $\mathbf{x}$ and $\mathbf{y}$, as

$$\mathbf{C}_{\mathbf{xy}} = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top]. \quad (\text{C.24})$$

C.1.15 Ensemble Estimators

Let $x_1, \dots, x_{N_e}$ be N_e independent and identically distributed samples from a distribution with mean μ_x and finite variance σ_x^2 . We can then use the following ensemble estimates:

- Mean:

$$\mathbb{E}[x] = \mu_x \approx \bar{x} = \frac{1}{N_e} \sum_{j=1}^{N_e} x_j. \quad (\text{C.25})$$

- Variance:

$$\mathbb{V}[x] = \sigma_x^2 \approx \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (x_j - \bar{x})^2. \quad (\text{C.26})$$

- Pearson correlation coefficient:

$$\rho_{xy} \approx \frac{\sum_{j=1}^{N_e} (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^{N_e} (x_j - \bar{x})^2} \sqrt{\sum_{j=1}^{N_e} (y_j - \bar{y})^2}}. \quad (\text{C.27})$$

Similar expressions can be written for random vectors:

- Mean:

$$\mathbb{E}[\mathbf{x}] \approx \bar{\mathbf{x}} = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{x}_j. \quad (\text{C.28})$$

- Covariance:

$$\begin{aligned}
\mathbf{C}_{\mathbf{xy}} &= \mathbb{E} \left[(\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{y} - \mathbb{E}[\mathbf{y}])^\top \right] \\
&\approx \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{x}_j - \bar{\mathbf{x}}) (\mathbf{y}_j - \bar{\mathbf{y}})^\top \\
&= \tilde{\mathbf{C}}_{\mathbf{xy}}.
\end{aligned} \tag{C.29}$$

C.1.16 Bias

Suppose we are estimating the true value x of our model using an estimator $\hat{x}$. The bias of the estimator is defined as

$$\begin{aligned}
\mathbb{B}[\hat{x}] &= \mathbb{E}[\hat{x} - x] \\
&= \mathbb{E}[\hat{x}] - x.
\end{aligned} \tag{C.30}$$

An estimator $\hat{x}$ is considered unbiased if $\mathbb{B}[\hat{x}] = 0$.

C.1.17 Mean Square Error

The MSE of an estimator $\hat{x}$ is defined as

$$\begin{aligned}
\text{MSE}[\hat{x}] &= \mathbb{E} \left[(\hat{x} - x)^2 \right] \\
&= \mathbb{E} \left[\left((\hat{x} - \mathbb{E}[\hat{x}]) + (\mathbb{E}[\hat{x}] - x) \right)^2 \right] \\
&= \mathbb{E} \left[(\hat{x} - \mathbb{E}[\hat{x}])^2 + 2(\hat{x} - \mathbb{E}[\hat{x}])(\mathbb{E}[\hat{x}] - x) + (\mathbb{E}[\hat{x}] - x)^2 \right] \\
&= \underbrace{\mathbb{E} \left[(\hat{x} - \mathbb{E}[\hat{x}])^2 \right]}_{=\mathbb{V}[\hat{x}]} + 2(\mathbb{E}[\hat{x}] - x) \underbrace{\mathbb{E}[(\hat{x} - \mathbb{E}[\hat{x}])]}_{=\mathbb{E}[\hat{x}] - \mathbb{E}[\hat{x}] = 0} + \underbrace{(\mathbb{E}[\hat{x}] - x)^2}_{=\mathbb{B}[\hat{x}]^2} \\
&= \mathbb{V}[\hat{x}] + \mathbb{B}[\hat{x}]^2.
\end{aligned} \tag{C.31}$$

This result indicates that an estimator $\hat{x}$ with small MSE represent a compromise between bias and variance.

C.1.18 Typical Distributions

C.1.18.1 Normal (Gaussian)

We denote that x is sampled from a normal (Gaussian) distribution with mean μ_x and variance σ_x^2 using $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$ (Fig. C.1). The PDF and CDF of a univariate Gaussian distribution are given by:

$$\text{PDF: } p(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{(x - \mu_x)^2}{2\sigma_x^2}\right) \quad (\text{C.32})$$

and

$$\text{CDF: } P(x) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x - \mu_x}{\sigma_x \sqrt{2}}\right) \right]. \quad (\text{C.33})$$

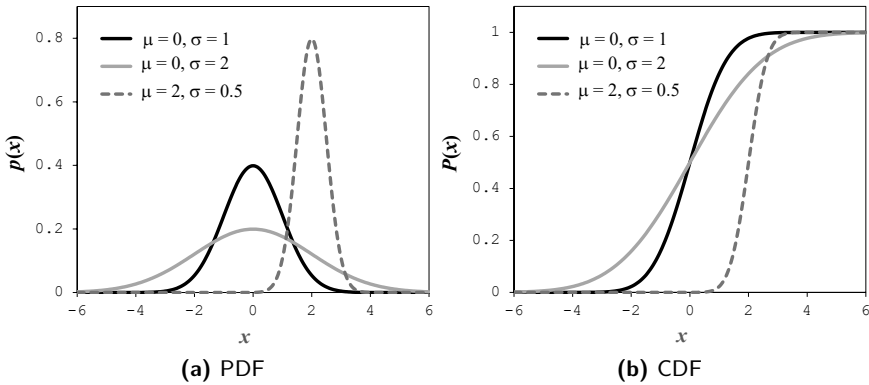


Fig. C.1: Normal (Gaussian) distribution.

Gaussian distributions are widely used and often considered a default choice for two key reasons:

1. Many real-world phenomena approximate a Gaussian distribution due to the Central Limit Theorem, which states that the average of independent, identically distributed random variables converges in distribution to a Gaussian.

2. Among all distributions with the same variance, the Gaussian has the highest level of uncertainty². In practice, choosing a Gaussian corresponds to adopting the least informative prior.

This book frequently assumes Gaussian distributions. A third reason for this assumption is that the theory for linear-Gaussian problems is well-established and forms the foundation for many of the methods discussed throughout the book.

C.1.18.2 Multivariate Gaussian

We use $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{C}_{\mathbf{x}})$ to denote that the random vector $\mathbf{x}$ is a sample from a multivariate Gaussian distribution with mean $\boldsymbol{\mu}_{\mathbf{x}}$ and covariance $\mathbf{C}_{\mathbf{x}}$. The PDF of $\mathbf{x}$ has the form:

$$p(\mathbf{x}) = \frac{1}{[(2\pi)^{N_x} \det(\mathbf{C}_{\mathbf{x}})]^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^\top \mathbf{C}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \right\}. \quad (\text{C.34})$$

Fig. C.2 displays the contour plot of a bivariate Gaussian distribution with a mean vector $\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{0}$ and covariance matrix

$$\mathbf{C}_{\mathbf{x}} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \quad (\text{C.35})$$

When $\rho = 0$, x_1 and x_2 are uncorrelated, which, for Gaussian distributions, also implies that they are independent.

C.1.18.3 Log-Normal

A random variable x is log-normally distributed if its logarithm, $z = \ln x$, is normally distributed, i.e., $\ln x \sim \mathcal{N}(\mu_x, \sigma_x^2)$ (Fig. C.3). The PDF and CDF of x have the forms:

$$\text{PDF: } p(x) = \frac{1}{x\sigma_x\sqrt{2\pi}} \exp \left(-\frac{(\ln x - \mu_x)^2}{2\sigma_x^2} \right) \quad (\text{C.36})$$

and

$$\text{CDF: } P(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\ln x - \mu_x}{\sigma_x\sqrt{2}} \right) \right]. \quad (\text{C.37})$$

² Specifically, the Gaussian distribution has the maximum entropy among distributions with the same variance and support on $(-\infty, \infty)$. In Information Theory, entropy is a common measure of uncertainty. The Maximum Entropy Principle suggests that, given the available information, the probability distribution with the highest entropy should be selected.

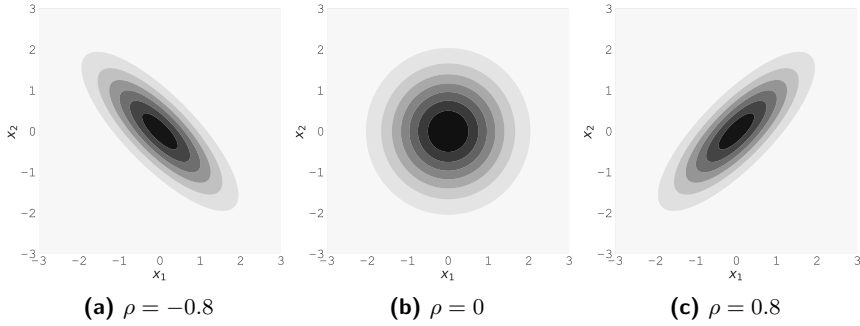


Fig. C.2: Bivariate Gaussian distributions.

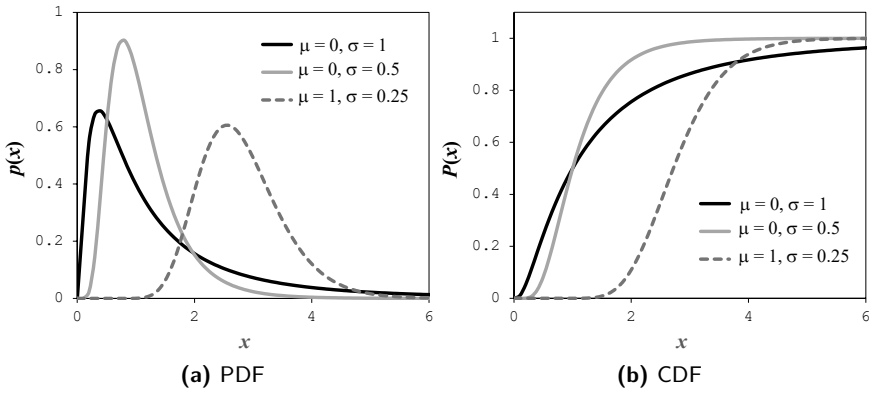


Fig. C.3: Log-normal distribution.

C.1.18.4 Chi-squared

A chi-squared distribution with k degrees of freedom is the distribution of a sum of the squares of k independent normally distributed random variables (Fig. C.4). This means that if $x = \sum_{i=1}^k z_i^2$ with $z_j \sim \mathcal{N}(0, 1)$, then $x \sim \chi^2(k)$. The PDF and CDF of x have the forms:

$$\text{PDF: } p(x) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} e^{-x/2} \quad (\text{C.38})$$

and

$$\text{CDF: } P(x) = \frac{1}{\Gamma(k/2)} \gamma\left(\frac{k}{2}, \frac{x}{2}\right), \quad (\text{C.39})$$

where $\Gamma(\cdot)$ represents the gamma function, and $\gamma(\cdot, \cdot)$ denotes the incomplete gamma function.

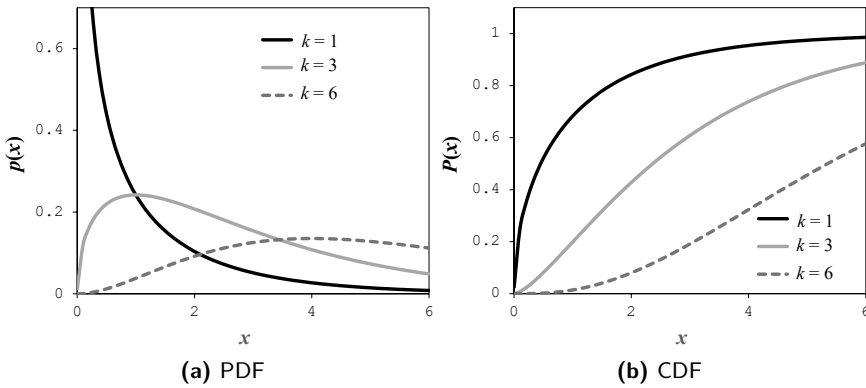


Fig. C.4: Chi-squared distribution.

C.1.18.5 Uniform

We use $x \sim \mathcal{U}(a, b)$ to denote that the x is a sample from a uniform distribution with boundary parameters a and b (Fig. C.5). The PDF and CDF of x have the forms:

$$\text{PDF: } p(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a \leq x \leq b \\ 0 & \text{for } x < a \text{ or } x > b \end{cases} \quad (\text{C.40})$$

and

$$\text{CDF: } P(x) = \begin{cases} 0, & \text{for } x < a \\ \frac{x-a}{b-a}, & \text{for } a \leq x \leq b \\ 1 & \text{for } x > b \end{cases} \quad (\text{C.41})$$

C.1.18.6 Triangular

The triangular distribution has three parameters: lower limit a , upper limit b , and mode c , where $a \leq c \leq b$ (Fig. C.6). The PDF and CDF of a triangular distribution have the forms:

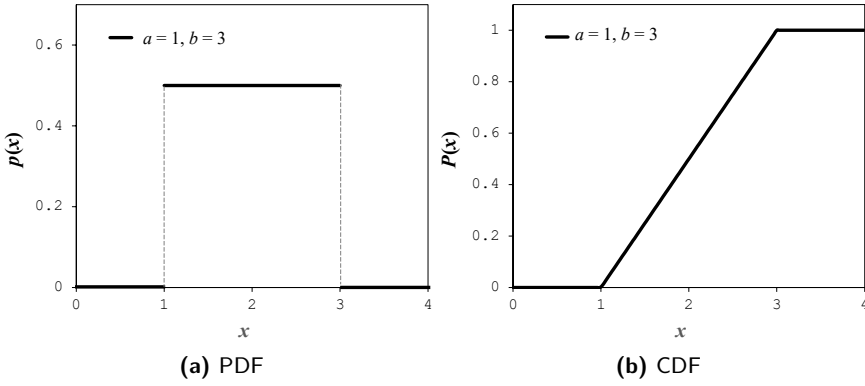


Fig. C.5: Uniform distribution.

$$\text{PDF: } p(x) = \begin{cases} 0, & \text{for } x < a \\ \frac{2(x-a)}{(b-a)(c-a)}, & \text{for } a \leq x < c \\ \frac{2(b-x)}{(b-a)(b-c)}, & \text{for } c < x \leq b \\ 0 & \text{for } x > b \end{cases} \quad (\text{C.42})$$

and

$$\text{CDF: } P(x) = \begin{cases} 0, & \text{for } x \leq a \\ \frac{(x-a)^2}{(b-a)(c-a)}, & \text{for } a < x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)}, & \text{for } c < x \leq b \\ 1 & \text{for } x > b \end{cases} \quad (\text{C.43})$$

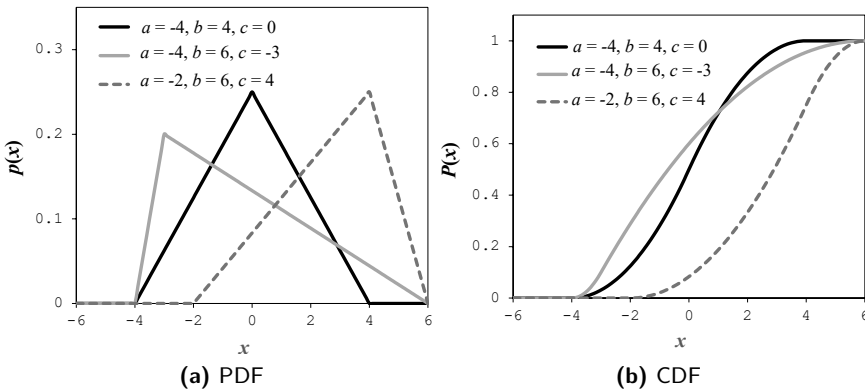


Fig. C.6: Triangular distribution.

C.1.18.7 PERT

The PERT distribution is similar to the triangular distribution but with a smoother shape. The distribution has three parameters: lower limit a , upper limit b , and mode c , where $a \leq c \leq b$ (Fig. C.7). The PDF and CDF of a PERT distribution have the forms:

$$\text{PDF: } p(x) = \frac{(x-a)^{\alpha_1-1} (b-x)^{\alpha_2-1}}{\mathcal{B}(\alpha_1, \alpha_2) (b-a)^{\alpha_1+\alpha_2-1}} \quad (\text{C.44})$$

and

$$\text{CDF: } P(x) = \mathcal{I}_z(\alpha_1, \alpha_2), \quad (\text{C.45})$$

where $\mathcal{B}(\cdot, \cdot)$ is the beta function and $\mathcal{I}_z(\cdot, \cdot)$ is the incomplete beta function with

$$\alpha_1 = \frac{4c + b - 5a}{b - a}, \quad (\text{C.46})$$

$$\alpha_2 = \frac{5b - a - 4c}{b - a}, \quad (\text{C.47})$$

and

$$z = \frac{x - a}{b - a}. \quad (\text{C.48})$$

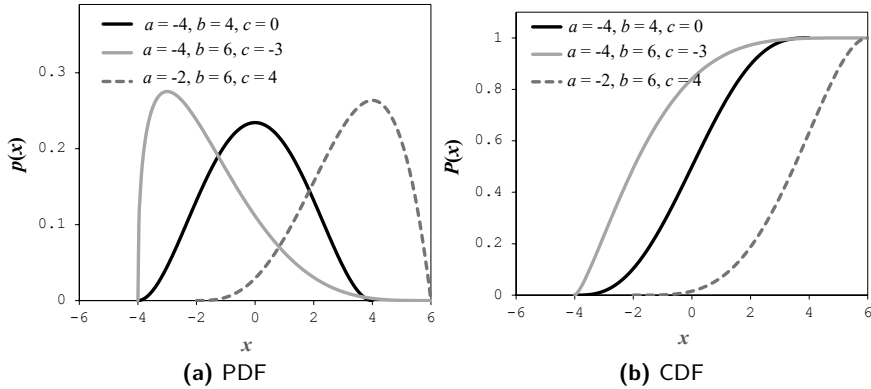


Fig. C.7: PERT distribution.

C.2 Likelihood

In statistics, likelihood is a concept used to evaluate how well a statistical model explains data. While probability asks about the chance of observing specific data given a particular model, likelihood reverses this perspective: it assesses the probability of different models given the observed data. Formally, the likelihood of a set of parameter values given some observed outcomes is defined as the probability of those outcomes occurring given the parameter values. This can be expressed as:

$$\mathcal{L}(\mathbf{m}|\mathbf{d}) = \mathcal{L}(\mathbf{m}) \equiv p(\mathbf{d}|\mathbf{m}). \quad (\text{C.49})$$

Here, $\mathcal{L}(\mathbf{m})$ represents the likelihood function, which measures how well the model $\mathbf{m}$ fits the data $\mathbf{d}$. Essentially, the likelihood function provides a quantitative measure of how plausible a particular model is in explaining the observed data [428].

C.2.1 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is a method used to identify the best parameters for a statistical model based on given data. The goal of MLE is to find the parameter values that maximize the likelihood function, which measures how well the model explains the observed data. In practice, this is often achieved by maximizing the log-likelihood function. The maximum likelihood estimate is obtained by solving the following optimization problem:

$$\mathbf{m}_{\text{ml}} = \arg \max_{\mathbf{m}} [\ln \mathcal{L}(\mathbf{m})]. \quad (\text{C.50})$$

C.3 Bayes' Rule

Perhaps the simplest way to introduce the Bayes' rule is using Venn diagrams. Fig. C.8 shows a Venn diagram with two possible events, A and B that can occur in a sample space Ω .

From the definitions of conditional probabilities

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (\text{C.51})$$

and

$$P(B|A) = \frac{P(B \cap A)}{P(A)}, \quad (\text{C.52})$$

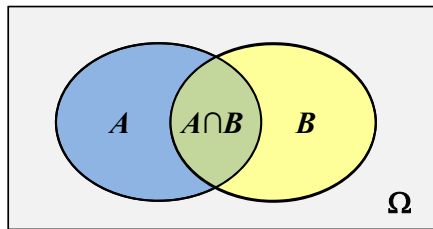


Fig. C.8: Venn diagram indicating the probability of occurrence of two events, A and B in sample space Ω .

we can write

$$P(A|B)P(B) = P(B|A)P(A). \quad (\text{C.53})$$

Hence,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\tilde{A})P(\tilde{A})}, \quad (\text{C.54})$$

where $P(A) + P(\tilde{A}) = P(\Omega) = 1$ and $\tilde{A}$ means “not A .” This last expression is Bayes’ rule, which allows us to evaluate, for example, the probability of event A given the occurrence of event B .

In this book, we use Bayes’ rule to obtain the conditional (posterior) distribution of a random vector of model parameters $\mathbf{m}$, given a set of observations $\mathbf{d}_{\text{obs}}$, expressed as

$$p(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{\text{obs}})}. \quad (\text{C.55})$$

Since $p(\mathbf{d}_{\text{obs}})$ is independent of $\mathbf{m}$, it acts as a constant term. This constant can be determined from the expression for the marginal distribution, leading to

$$p(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m})}{\int_{\mathbf{m}} p(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m})d\mathbf{m}}. \quad (\text{C.56})$$

Bayes’ rule is a direct application of the definitions of joint and conditional probabilities. However, the interpretation given to the resulting equation forms the foundation of Bayesian statistics. Specifically, we write

$$\underbrace{p(\mathbf{m}|\mathbf{d}_{\text{obs}})}_{\text{posterior}} = \text{const} \times \underbrace{\mathcal{L}(\mathbf{m}|\mathbf{d}_{\text{obs}})}_{\text{likelihood}} \times \underbrace{p(\mathbf{m})}_{\text{prior}}, \quad (\text{C.57})$$

which means that the posterior distribution is proportional to the product of the likelihood and the prior.

C.4 Geostatistics

Geostatistics is a specialized branch of statistics that focuses on analyzing and interpreting spatial or spatiotemporal datasets. Early geostatistical methods were closely linked to interpolation techniques, particularly with the development of kriging. However, contemporary geostatistics extends beyond mere interpolation, incorporating a range of sophisticated techniques for spatial analysis and modeling.

This section provides a brief review of fundamental geostatistical concepts that are essential for understanding the topics discussed in this book. For a more comprehensive exploration of geostatistics, several excellent resources are available, including Deutsch [104], Wackernagel [456], and Mariethoz and Caers [290]. Numerous commercial and open-source geostatistical software options are also available. Notable open-source examples include `GSLIB` [105], `gstat` [343], `SGeMS` [362], and `GeoStats.jl` [194].

C.4.1 Stationarity

Consider a random variable z . We want to model z as a function of its spatial position, i.e., $z(\mathbf{u})$. The covariance of z at two locations, $\mathbf{u}$ and $\mathbf{u} + \mathbf{h}$ is given by

$$\mathbb{C}[z(\mathbf{u}), z(\mathbf{u} + \mathbf{h})] = \mathbb{E}[(z(\mathbf{u}) - \mathbb{E}[z(\mathbf{u})])(z(\mathbf{u} + \mathbf{h}) - \mathbb{E}[z(\mathbf{u} + \mathbf{h})])]. \quad (\text{C.58})$$

- **First-order stationarity:** the expected value at any location in the specified region is constant.

$$\mathbb{E}[z(\mathbf{u})] = \mathbb{E}[z(\mathbf{u} + \mathbf{h})]. \quad (\text{C.59})$$

- **Second-order stationarity:** the covariance between the value of the random function at two points is only a function of the distance between the two points.

$$\mathbb{C}[z(\mathbf{u}), z(\mathbf{u} + \mathbf{h})] = \mathcal{C}(h). \quad (\text{C.60})$$

C.4.2 Transformation of Variables

Many estimation and simulation methods in geostatistics rely on the assumption that the PDF is Gaussian. Therefore, we often transform variables to approximately univariate Gaussian.

The Box-Cox transform is often used

$$z = \frac{x^\lambda - 1}{\lambda}, \quad \text{with } \lambda \neq 0. \quad (\text{C.61})$$

For $\lambda \rightarrow 0$, we have the log-transform

$$z = \ln(x). \quad (\text{C.62})$$

The log-transform is very useful for permeability and transmissibility values.

The normal score transform (Fig. C.9) is a nonparametric method that converts a given frequency distribution into a univariate Gaussian distribution. Also known as Gaussian anamorphosis, this transformation applies to each parameter's marginal distribution independently, meaning it does not account for the multivariate Gaussian distribution. Consequently, there is no guarantee that the joint distribution of all variables will be Gaussian. However, in practice, this method often improves the Gaussian properties of the joint PDF [456].

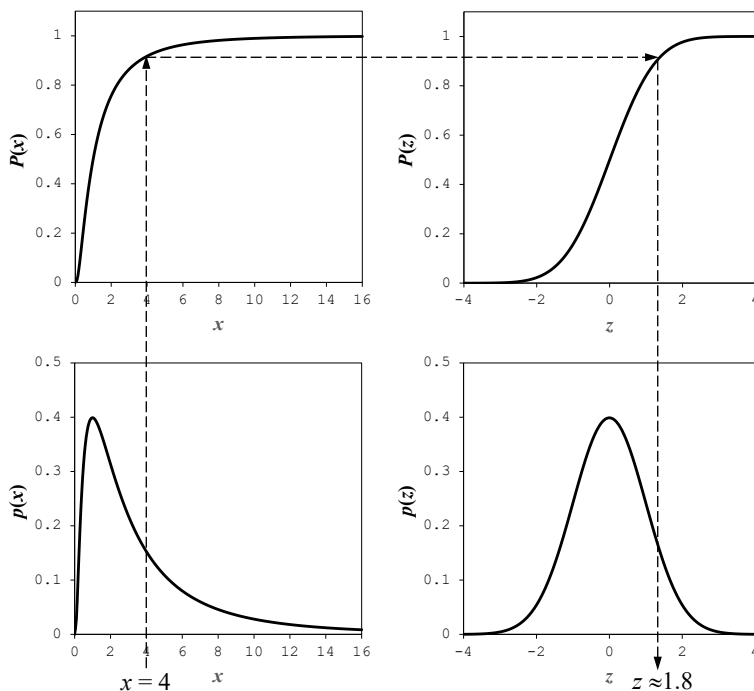


Fig. C.9: Normal score transform.

C.4.3 Variogram

In geostatistics, the concept of variogram is often preferred over the covariance, even though both contain equivalent information. Mathematically, the variogram is defined as half the variance of the difference between values at two locations separated by a distance h ; this is also referred to as the semivariogram. For stationary models, the variogram depends only on the distance h between locations, not their absolute positions. The variogram $\gamma(h)$ of a random variable z has the form

$$\begin{aligned}\gamma(h) &= \frac{1}{2} \mathbb{V}[z(\mathbf{u}) - z(\mathbf{u} + \mathbf{h})] \\ &= \mathcal{C}(0) - \mathcal{C}(h) \\ &= \mathcal{C}(0) \left[1 - \frac{\mathcal{C}(h)}{\mathcal{C}(0)} \right] \\ &= \mathcal{C}(0) [1 - \rho(h)],\end{aligned}\tag{C.63}$$

where $\gamma(h)$, $\mathcal{C}(h)$, and $\rho(h)$ represent the variogram, covariance, and correlation functions, respectively. Typical covariance functions include:

$$\text{Spherical: } \mathcal{C}(h) = \sigma^2 \begin{cases} 1 - \frac{3h}{2L} + \frac{h^3}{2L^3}, & 0 \leq h \leq L \\ 0, & h > L, \end{cases}\tag{C.64}$$

$$\text{Exponential: } \mathcal{C}(h) = \sigma^2 \exp\left(-\frac{3h}{L}\right),\tag{C.65}$$

and

$$\text{Gaussian: } \mathcal{C}(h) = \sigma^2 \exp\left(-\frac{3h^2}{L^2}\right),\tag{C.66}$$

where L is the correlation length.

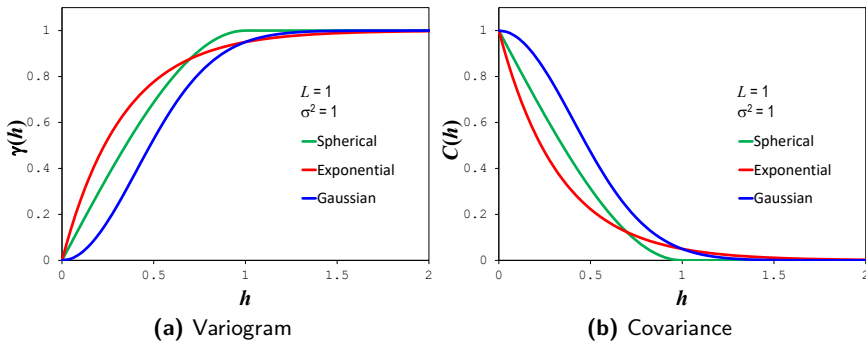


Fig. C.10: Variogram and covariance functions.

For stationary models, the covariance depends solely on the separation distance h between locations rather than their absolute positions. However, in practice, the covariance between two locations can also be a function of direction, reflecting anisotropy in the spatial relationships. For example, in channel deposits, we often observe longer correlations along the direction of the channel. In the 3D space, this anisotropy results in correlation lengths that form a 3D ellipsoid, capturing the directional variability in spatial continuity.

For the 3D case, we can transform an anisotropic covariance into isotropic by stretching and rotating.

- **Rotation:** the 3D ellipsoid is rotated from the x - y - z coordinate system to the x' - y' - z' using

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos \alpha & 0 & -\sin \alpha \\ 0 & 1 & 0 \\ \sin \alpha & 0 & \cos \alpha \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & \sin \beta \\ 0 & -\sin \beta & \cos \beta \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (\text{C.67})$$

The rotation angles are usually called θ = azimuth, β = dip and α = rake.

- **Stretching:** we compute the ratio h/L in the covariance function as

$$\frac{h}{L} = \sqrt{\left(\frac{\Delta x'}{L_1}\right)^2 + \left(\frac{\Delta y'}{L_2}\right)^2 + \left(\frac{\Delta z'}{L_3}\right)^2}, \quad (\text{C.68})$$

where $\Delta x'$, $\Delta y'$ and $\Delta z'$ are the distances in the x' , y' , and z' directions, respectively. L_1 , L_2 and L_3 are the corresponding correlation lengths.

C.4.4 Kriging

Kriging is one of the most fundamental methods in geostatistics, with its theoretical foundation developed by Matheron [295], building on the earlier work of Krige [242]. Initially conceived as an interpolation technique, kriging is mathematically related to regression analysis and can be derived as a best linear unbiased estimator (BLUE). It also has a Bayesian interpretation, where, under the assumption of a linear-Gaussian process, the kriging model corresponds to the posterior mean.

Several variants of kriging exist, including simple kriging, ordinary kriging (where the mean is locally estimated), kriging with trend (where the mean is modeled as a function of spatial position), cokriging (which incorporates correlations with secondary variables), among others.

C.4.5 Geostatistical Simulation

Kriging provides a smooth solution based on a local average of the available data. While this smoothing effect offers a reliable average value, it removes fine-scale details of reservoir heterogeneity that are critical for certain applications. Additionally, kriging is a deterministic method, providing a unique estimate of the property of interest. To address these limitations, geostatistical methods that generate multiple model realizations are often preferred, as they allow for a more detailed representation of heterogeneity and provide a means to assess uncertainty in model predictions.

Like kriging, geostatistical simulation can also be interpreted within a Bayesian framework. Under the assumption of a linear-Gaussian process, the simulated realizations represent samples from a multivariate Gaussian distribution. When averaged, multiple realizations converge to the kriging estimate, which corresponds to the posterior mean. In geostatistics, methods that rely on covariance functions or variograms are commonly referred to as “two-point geostatistics,” as they focus on spatial relationships between pairs of points.

There are various geostatistical (stochastic) simulation methods, with sequential Gaussian simulation (SGS) [171] being one of the most widely used [104]. The core concept of SGS is to simulate each gridblock sequentially along a random path. A sample from $p(\mathbf{m})$ is obtained using the chain rule of conditional probabilities:

1. Start at the first gridblock and sample $m_1 \sim p(m_1)$.
2. Proceed to the second gridblock and sample $m_2 \sim p(m_2|m_1)$.
3. Continue to the third gridblock and sample $m_3 \sim p(m_3|m_1, m_2)$.
- ...
4. Visit the last gridblock and sample $m_{N_m} \sim p(m_{N_m}|m_1, \dots, m_{N_m-1})$.

A model $\mathbf{m} = [m_1, \dots, m_{N_m}]^\top$ generated using this process represents a sample of the distribution $p(\mathbf{m})$. The model parameters can be visited in any sequence, as long as all parameters are eventually accounted for [171]. Each conditional distribution $p(m_i|m_1, \dots, m_{i-1})$ follows a one-dimensional Gaussian distribution. In SGS, the mean and variance of this Gaussian distribution are estimated using simple kriging. To maintain computational efficiency, SGS typically considers only a limited number of nearby parameters in the neighborhood when estimating each conditional distribution.

C.4.6 Geostatistical Simulation of Facies

The parameters of geostatistical models are derived from available sample data, such as well logs and seismic measurements. For a geostatistical model

to be reliable, the sample data must adequately capture the key characteristics of the spatial random variables. In cases where correlations are not expected across the entire domain, the domain is divided into regions (facies) where the geostatistical model is assumed to hold. For each facies, specific assumptions are made about the mean and covariance of the property of interest, allowing for a more accurate representation of spatial variability within each region.

One of the most widely used geostatistical methods for simulating facies is sequential indicator simulation (SIS) [7]. Similar to SGS, SIS visits each gridblock in a random order. At each step of the process, the following occurs:

1. Identify nearby data and previously simulated gridblocks.
2. Construct the conditional distribution by kriging, calculating the probability of each facies type occurring at the current location.
3. Draw a simulated facies type based on the computed probabilities.

Another commonly used method is truncated Gaussian simulation (TGS) [296]. The core concept behind TGS is to generate a realization from a continuous Gaussian field and then apply a series of thresholds to transform the continuous values into categorical facies. Since the facies are derived from an underlying continuous variable, the resulting facies tend to exhibit a natural order, meaning it is unlikely to observe non-adjacent facies (e.g., facies 1 and facies 3) in direct contact with each other. This method is particularly useful for modeling facies with a stratified or gradational structure.

C.4.7 Beyond Covariance-Based Models

Covariance is effective at capturing the similarity (correlation) between two spatial locations, providing insights into two-point correlations. However, this approach falls short when modeling more complex geological features, such as sinuous 3D channels. For these intricate patterns, methods that account for multiple-point correlations are required³.

There are geostatistical algorithms designed to reproduce complex facies distributions that cannot be accurately described with single covariance functions. Among these, the use of locally varying anisotropy (LVA) [43] is quite common. LVA applies standard two-point algorithms but calculates distances while considering changing anisotropy within the model, allowing

³ A more technical explanation for this limitation is that Gaussian models are constrained by the fact that the Gaussian distribution represents the probability distribution with maximum entropy for a given mean and covariance. Essentially, the multivariate Gaussian distribution maximizes spatial disorder and is thus unable to effectively represent structured features like channel systems [224]. This inherent property limits its ability to capture more complex geological patterns that extend beyond simple correlations.

for the creation of curvilinear patterns. Truncated plurigaussian simulation (TPG) [157, 17, 35] generates complex facies distributions by combining two or more Gaussian random fields with specific truncation rules.

Boolean or object-based simulation [107, 106] involves the random placement of geological structures (objects) within the model. The shapes and characteristics of these objects (e.g., length, width, and sinuosity) are also selected randomly. In this approach, data conditioning requires solving an inverse problem, where objects are adjusted and transformed until the data matches. Although object-based algorithms are capable of producing realistic geological structures, they are challenging to condition when dealing with large amounts of local data.

Another class of algorithms includes multiple-point statistics (MPS) [179, 417, 290], which use training images as a database of geological patterns. Originally, MPS was viewed as an alternative to object-based modeling, designed to generate complex geological structures while being easier to condition on large amounts of data [412]. Today, there are many algorithms in this category, employing techniques similar to the ones used in machine learning and computer graphics [289].

The previous methods are often called classical geostatistical methods, as they rely on probabilistic models to generate the distribution of geological structures. In contrast, there are methods specifically designed to describe the depositional processes that formed the reservoir. These approaches are generally classified as rule-based modeling and process-based modeling [352]. Rule-based methods [202, 424] use predefined geological rules, heuristics, and geometric constraints to construct models. While they can produce more realistic geological structures compared to classical geostatistical methods, they are less effective at conditioning to well and seismic data.

Process-based modeling [341, 89, 52] simulates the physical processes that govern geological formations, such as sediment transport, erosion, and deposition. These models use numerical solutions of physical laws to generate highly realistic geological structures. The main challenges of process-based modeling include poorly known initial and boundary conditions needed to constrain the numerical model, severe limitations in conditioning models to observations, and high computational costs, which hinder the generation of multiple realizations and reduce the ability to represent uncertainty. As a result, process-based approaches are not commonly used for full-field model generation. Instead, they are typically utilized to gain insights into the expected reservoir geometry and heterogeneity.

Nothing to see here. Move along.



Brief Literature Review

Abstract: *This appendix offers a brief literature review, highlighting influential papers and current applications of ensemble methods in reservoir modeling.*

D.1 Introduction

Ensemble-based data assimilation methods have gained widespread attention for their ability to manage high-dimensional problems and quantify uncertainty, making them particularly well-suited for complex geophysical systems. The literature on this topic is extensive and continues to grow. This section presents a concise review of ensemble-based data assimilation methods, focusing on their application to subsurface characterization and monitoring.

D.2 EnKF and Its Variants

Evensen [136] introduced the EnKF to overcome the limitations of the extended Kalman filter (EKF) in managing high-dimensional, nonlinear dynamical systems. The EKF relies on linearizing the model and observation equations around the estimated state mean, which requires derivative calculations. As noted by Evensen [139], these linearizations can result in unstable propagation of the state covariance. Additionally, updating covariance matrices in the EKF becomes impractical for large-scale problems. In contrast, the EnKF is a Monte Carlo-based method that uses an ensemble of state vectors to approximate the mean and covariance, updating them sequentially over time. Typically, the number of state vectors in the EnKF ensemble is much smaller than the number of unknowns, allowing the method to efficiently

update large state covariance using a low-rank approximation. Rather than directly updating a covariance matrix, the EnKF propagates uncertainty through the ensemble of states.

Although the EnKF was initially introduced by Evensen [136], its formulation was later clarified by Burgers et al. [51] and Houtekamer and Mitchell [197], who introduced the concept of updating each ensemble member with independently perturbed observations, which led to the modern implementation of the EnKF.

The initial development of the EnKF occurred primarily in the fields of numerical weather prediction (NWP) and oceanography. Recent review papers by Houtekamer and Zhang [200] and Carrassi et al. [60] summarize key developments and ongoing challenges in these areas. EnKF-like methods are currently operational at the TOPAZ Ocean System, developed by the Nansen Environmental and Remote Sensing Centre (NERSC) [375], and in the atmospheric model at the Canadian Meteorological Centre (CMC) [201].

Variants of the EnKF fall into two main categories: “stochastic” and “deterministic (or square root)” schemes. Stochastic versions of the EnKF use a perturbed-observation¹ scheme to derive the analysis equation. Advocates of square-root schemes argue that data perturbation introduces additional sampling errors into the process [60], which can be avoided with the square-root variants. Notable examples of deterministic schemes include the ensemble transform Kalman filter (ETKF) [39], the ensemble adjustment Kalman filter (EAKF) [13], the ensemble square root filter (EnSRF) [463], the local ensemble transform Kalman filter (LETKF) [207], and the deterministic ensemble Kalman filter (DEnKF) [374].

The adoption of the EnKF for reservoir data assimilation by Nævdal et al. [313] initiated a series of research efforts, particularly in petroleum engineering, leading to numerous subsequent studies. Comprehensive reviews of the main developments, challenges, and early applications of the EnKF in reservoir data assimilation from 2002 to early 2011 are provided by Aanonsen et al. [3] and Oliver and Chen [331]. The first reported field application of the EnKF for reservoir data assimilation was presented by Skjervheim et al. [399]. Subsequent field examples include studies by Bianco et al. [36], Evensen et al. [144], Haugen et al. [188], and Emerick and Reynolds [127].

¹ van Leeuwen [447] noted that while traditional interpretations of stochastic EnKF schemes focus on perturbing observations, it is more appropriate to perturb the predictions. This approach assumes that the true system is a realization from the same distribution as the ensemble members. Since the true system’s measurements are already affected by observation errors, the predictions of each ensemble member should be perturbed in a similar manner. In practice, the calculations are the same whether perturbations are applied to the observations or the predicted data. However, these interpretations may yield different results when extending the methods to address non-additive errors.

D.3 From EnKF to Iterative ES

Sequential data assimilation is a key factor in the success of the EnKF and its variants in NWP. The recursive updates in time help to maintain the model states close to the observations. However, data assimilation in NWP differs significantly from that in reservoir applications. In NWP, forecasts are made on a scale of hours to a few days, whereas in reservoir modeling, predictions are made over years. Additionally, NWP typically assimilates a much larger volume of observations. The short-term nature and the large number of data points in NWP tend to reduce the violation of the linearity during the update step.

In contrast, reservoir data assimilation traditionally focuses on parameter estimation, aimed at determining “static” model parameters, while NWP deals with state estimation, focusing on dynamic variables like pressure and temperature distributions. The main uncertainty in NWP lies in the initial condition, whereas in reservoir applications, it is the spatial distribution of rock properties. In addition, NWP involves unstable dynamics, where small changes in initial conditions can lead to divergent solutions, while reservoir models exhibit stable dynamics. Due to these differences, the sequential data assimilation approach of the EnKF is well-suited for NWP but less effective for reservoir modeling.

To make sequential data assimilation work for reservoirs, the process must be framed as a combined parameter-state estimation problem. This approach prevents the need to rerun reservoir simulations from the initial condition at each assimilation time step. This approach assumes that the updated reservoir states (i.e., the simulator’s primary variables) remain statistically consistent with those obtained from forward simulations initialized at time zero using the updated model parameters. However, this consistency is guaranteed only in the linear-Gaussian case [435]. In practice, consistency violation tends to degrade the data assimilation results [435, 459].

One way to ensure consistency between updated parameters and primary variables is to rerun the reservoir simulator from time zero using the latest model parameters after each assimilation step. While this ensures consistency, it significantly increases computational cost. Wang et al. [459] referred to this method as the half-iteration EnKF (HI-EnKF). To improve efficiency, Wang et al. [459] proposed rerunning the ensemble from time zero only if the average relative change in the ensemble mean exceeds a specified threshold.

Although sequential updates help to manage nonlinearity, applications with strong nonlinearities may require iterative methods to constrain ensemble members effectively, especially when high-quality observations are available [376, 42].

One of the first iterative EnKF methods was introduced by Gu and Oliver [178], who combined an ensemble-based sequential scheme with a Gauss-Newton (GN) RML formulation, calling it ensemble randomized maximum likelihood (EnRML). This method involves rerunning simulations from time

zero at every data assimilation time step, significantly increasing computational cost. To mitigate this, Chen and Oliver [76] applied a hybrid approach, combining EnKF and EnRML in the Brugge case [346]. In their approach, the standard EnKF update is used unless state variable changes are large, at which point EnRML is applied. A similar approach was proposed by Li and Reynolds [256], who switched from EnKF to an adjoint-based update when necessary. Sakov et al. [376] proposed an improved EnRML implementation by incorporating a square root scheme and rescaling predicted data anomalies to enhance stability. Bocquet and Sakov [42] modified this by replacing the GN approach with a Levenberg-Marquardt (LM) formulation.

Skjervheim et al. [400] proposed using the ES for reservoir applications, arguing that the batch update approach of ES might suffice given the diffusive and stable nature of reservoir models. ES offers significant computational advantages by eliminating the need for frequent simulation restarts and providing flexibility in integrating complex parametrizations, including structural parameters and facies. It also interfaces more easily with commercial simulators, facilitating its adoption in operational workflows.

However, subsequent studies revealed that a single ES update is insufficient for constraining reservoir models to multiple observations [78, 132], leading to the proposal of iterative versions of ES. Chen and Oliver [78] introduced batch-EnRML, which incorporated the EnRML scheme of Gu and Oliver [178] into the smoother framework, showing good results in the Brugge case, though it required several iterations to converge. Around the same time, Emerick and Reynolds [132] introduced the ES-MDA method, which exhibited good data assimilation results with a significantly lower computational cost compared to batch-EnRML. The same authors proposed in [131] a comparative study of ensemble methods in a simple but highly nonlinear reservoir problem. Among the methods investigated, ES-MDA showed remarkable performance, achieving good approximations of the posterior marginal distribution of log-permeability and water production, comparable to those obtained through rigorous MCMC sampling. In the same year, Chen and Oliver [79] presented an improved implementation of EnRML by replacing the GN update with an LM update formula and introducing a more stable Hessian approximation. The resulting LM-EnRML method demonstrated computational performance comparable to ES-MDA in both the two-phase flow case proposed in [131] and the Brugge case, with the added potential advantage of not requiring the pre-selection of the number of iterations.

Stordal [413] introduced the iterative adaptive Gaussian mixture smoother (IAGS), an extension of the adaptive Gaussian mixture filter proposed in Stordal et al. [416] and Stordal and Lorentzen [415]. This method integrates an EnKF-like update equation with weighting and resampling steps akin to those used in particle filters [114]. The authors draw parallels between IAGS, EnRML, and ES-MDA, particularly noting that the IAGS's bandwidth parameter corresponds to the LM parameter in EnRML or the inflation coefficient in ES-MDA. The key distinction lies in the inclusion of the

weighting/resampling step. The authors suggest that IAGS is expected to perform comparably to EnRML and ES-MDA in large-scale problems, with potential advantages in small-scale applications.

Luo et al. [279] introduced a variation of the EnRML within the framework of iterative ensemble smoothing, utilizing the regularized Levenberg-Marquardt (RLM) algorithm from Jin [220] to solve a minimum-average-cost (MAC) problem. This method, referred to as RLM-MAC, was compared with a version of LM-EnRML discussed in Chen and Oliver [79]. The results showed that RLM-MAC outperformed LM-EnRML in the Lorenz 96 model [276], while their performance was comparable in reservoir data assimilation scenarios.

The standard ES-MDA method requires specifying the number of assimilations and the sequence of inflation coefficients in advance. This requirement spurred research into adaptive ES-MDA schemes. Emerick [118] introduced an *ad hoc* adaptive scheme based on the evolution of the data-mismatch objective function. Building on this, Le et al. [253] incorporated an additional check to limit the maximum allowable change in model parameters. The same authors also introduced a scheme that employed the regularized LM approach from Hanke [184]. However, these early adaptive schemes were found to require an excessive number of iterations for operational use. To address this issue, Rafiee and Reynolds [356] proposed selecting the MDA inflation coefficients based on a geometrically decreasing sequence, an idea that was further explored in subsequent studies [122, 284, 396].

More recently, Raanes et al. [354] revisited the EnRML method and introduced a new derivation based on the fact that estimates from ensemble methods are confined to the subspace defined by the prior ensemble. The resulting method, the subspace iterative ensemble smoother (SIES), is simpler in both derivation and implementation compared to EnRML.

Currently, iterative forms of ES rank among the most effective and widely adopted methods for reservoir data assimilation. Their application in real-world field cases has expanded significantly, and these methods have become standard tools in applications for petroleum reservoirs.

D.4 Localization

Localization is a standard technique used to address the limitations imposed by small ensemble sizes. Houtekamer and Mitchell [197] first introduced localization by applying a distance cutoff to the Kalman gain, ensuring that only parameters and states within a specified distance from the observation were updated. Later, Houtekamer and Mitchell [198] improved this approach by incorporating the Schur product for localization. The authors observed that using the Schur product resulted in relatively smoother mod-

els compared to those produced with a simple distance cutoff or without localization.

An alternative implementation of localization is known as local analysis or domain localization. This method involves performing multiple localized updates, using only observations near each parameter location. Compared to standard Schur-product localization, domain localization does not require the explicit formation of covariance or Kalman gain matrices, making it applicable to square root schemes. Sakov and Bertino [372] compared both methods and concluded that they produce similar results. Chen and Oliver [81] also compared domain localization with Schur-product localization of the Kalman gain and found that LM-EnRML converges faster with domain localization.

Arroyo-Negrete et al. [19] and Devegowda et al. [108] proposed using streamlines to identify model gridblocks that are sensitive to data, which are then used to define the corresponding localization values. Chen and Oliver [77] and Emerick and Reynolds [128] noted that, while streamline-based localization is intuitive and appealing because it incorporates some of the fluid transport physics into the localization process, relying solely on these regions of influence risks confining model updates too narrowly around wells, particularly when the prior model covariance exhibits long correlation ranges.

Localization schemes typically assume that the Kalman gain or the covariances between parameters and data depend on distance. However, some localization schemes are designed for more general covariance structures. One of the earliest non-distance-based localization methods is the hierarchical filter proposed by Anderson [14]. In this approach, an ensemble is divided into sub-ensembles, which are then used to compute correction coefficients for the Kalman gain. Instead of dividing the ensembles, Zhang and Oliver [482] suggested using bootstrap resampling to generate new ensembles. Furrer and Bengtsson [156] derived a localization function for general covariance structures by minimizing the norm of the difference between the true covariance and the localized estimate on a term-by-term basis without enforcing positive definiteness. The final expression depends on the true covariance, but the authors suggest replacing the true covariance with the ensemble estimate to construct a non-distance-based localization function. Bishop and Hodyss [37] proposed generating the localization matrix by raising each element of a sample correlation matrix of model variables to a specified power. Luo et al. [280] suggested calculating localization values by using correlation coefficients estimated from the ensemble and applying a threshold to determine the final values. Lacerda et al. [245] reviewed several of these methods and compared their effectiveness in updating non-local parameters in a synthetic reservoir data assimilation problem. They concluded that most of these methods require careful tuning of internal parameters, which significantly affects their performance.

Luo and Bhakta [278] extended their correlation-based method by introducing a procedure to estimate the noise level in the correlation esti-

mates. They also proposed replacing the hard threshold scheme with a continuous tapering approach based on the Gaspari-Cohn correlation function [162]. This method was further refined by Ranazzi et al. [358]. Soares et al. [405] combined the correlation-based approach with domain localization and concluded that this method is better suited for handling large-scale problems compared to Schur-product localization of the Kalman gain. Recently, Vishny et al. [452] introduced a method called noise-informed covariance estimation (NICE), which combines concepts present in correlation-based localization and the Furrer-Bengtsson taper.

The premature loss of ensemble variance remains a significant challenge in ensemble methods. Localization is currently the most effective approach to mitigate this issue, but it requires careful tuning to achieve optimal results. Developing robust strategies to address sampling errors, including adaptive localization schemes, is an active area of research.

D.5 Model Errors

The “perfect model assumption” is prevalent in many reservoir data assimilation applications. This assumption unrealistically posits that the model would accurately represent the true behavior of the reservoir if the correct values of uncertain parameters were known. In reality, however, models inevitably simplify the actual system, leading to errors from these inherent approximations. Such errors are commonly referred to as model errors, but other terms appear in the literature, including model inadequacy, model discrepancy, model bias, and structural uncertainty [235, 49].

The formal incorporation of model errors into reservoir data assimilation has been a relatively recent development in the literature. Studies such as [450, 330, 141, 9] have shown that neglecting model errors can lead to overconfidence in the data and underestimation of uncertainty in model forecasts.

Model errors in data assimilation are commonly addressed by assuming that they are additive and Gaussian or by reducing the weight of observations, which is achieved by increasing the data-error variance [80, 124]. This approach is discussed in Sun et al. [422], where the authors introduce an *ad hoc* procedure to mitigate the excessive reduction in uncertainty that occurs when model errors are present.

Oliver and Alfonzo [330] introduce a method for incorporating model errors by adjusting the data-error covariance, $\mathbf{C}_e$, based on the residuals (i.e., the difference between observed and simulated data) after data assimilation. The approach updates $\mathbf{C}_e$ and then repeats the assimilation process, starting from the prior ensemble. This method has been further explored in the context of production data assimilation [277] and 4D seismic data assimilation [9]. Emerick and Neto [125] introduced an efficient variation of

this method by calculating the residuals based on the difference between observed and projected data within the ensemble subspace for 4D seismic.

Evensen [141] expanded upon the traditional approach of assuming additive model errors by introducing a method to estimate the model error term concurrently with the model parameters during data assimilation. While this approach broadens the treatment of model errors, it still relies on the assumption of Gaussian error distributions. Overall, the development of robust methods for handling non-Gaussian model errors in data assimilation remains an open problem.

D.6 Facies

Due to the inherent Gaussianity assumptions in the formulation of ensemble methods, they are better suited for covariance (variogram) based reservoir models. Consequently, these methods may face difficulties in preserving complex geological features, with categorical facies being a particularly challenging example. As a result, the development of robust parametrizations strategies for facies has become a highly active area of research, leading to a substantial number of publications in recent years.

One of the earliest works on facies updating using ensemble methods was presented by Liu and Oliver [267], which employed the truncated plurigaussian (TPG) method [157, 17, 35]. The TPG parametrization has since been explored and further refined in subsequent studies [267, 484, 387, 386]. Today, TPG remains one of the most widely used and effective approaches for parameterizing facies models within ensemble-based methods.

Another approach to updating facies models with ensemble methods involves the use of level-set functions [64, 272, 305, 304, 347]. As noted by Mannseth [287], there is a relationship between level-set and TPG methods, as any plurigaussian field can be represented by a corresponding level-set function.

However, TPG and level-set functions present significant challenges when representing facies with complex geometries, such as channels, particularly those produced by object-based modeling [107, 106], multiple-point statistics techniques [179, 417, 290], or rule-based modeling [352]. Consequently, numerous studies have proposed alternative strategies for models with channelized facies distributions. Notable examples include discrete cosine transform [213], kernel principal component analysis [378, 377], sparse dictionary learning with K-SVD [237], and methods based on thresholded principal component analysis [453, 71, 120].

Another class of methods involves updating “probability maps,” which are then used to recompute the facies distribution using geostatistical algorithms. Variations of this approach exist; for example, some studies employ “pseudo-wells” in areas with high facies probability as conditioning data

for generating new realizations [432, 65], while others use probability maps directly as a soft constraint in geostatistical simulations [212, 252, 388].

Recently, deep learning networks, particularly deep generative models, have emerged as a promising research direction for developing facies parametrizations. Various approaches have been explored in the literature, including deep belief networks [55], variational autoencoders [246, 56], generative adversarial networks [62, 247, 115, 63, 308], style transfer networks [268], and latent diffusion models [148]. Some of these deep learning strategies are compared in [57]. One attractive feature of these methods is that they allow us to represent facies in terms of latent Gaussian random fields, which is particularly appealing in the context of ensemble methods. The primary challenge of using these models lies in the substantial computational cost required to train deep learning networks, particularly when applied to multi-million-cell 3D reservoir models.

D.7 4D Seismic Data Assimilation

4D (time-lapse) seismic imaging involves conducting repeated seismic surveys over time to monitor fluid movements and pressure variations within reservoirs. The literature on 4D seismic data assimilation is quite extensive. Oliver et al. [336] provides a comprehensive review of this topic, detailing key methods, challenges, and recent applications.

The assimilation of 4D seismic data presents several challenges, particularly due to the presence of significant errors and biases in the seismic datasets, as well as limitations in the forward modeling process. Moreover, 4D seismic typically involves a large number of data points, which can exacerbate issues with variance underestimation in ensemble methods. There are also scale issues, as seismic datasets have lower vertical resolution than current reservoir simulation models. These aspects have been the subject of several investigations. For example, Oliver [329] analyzed the impact of spatially correlated errors in 4D seismic data and found that ignoring these correlations led to biased parameter estimates and inadequate uncertainty quantification. The effect of model errors on the assimilation of 4D seismic data has also been investigated in [9, 320, 282, 125].

4D seismic assimilation is commonly performed in one of three main domains: the seismic domain, which involves seismic attributes such as amplitudes and time shifts; the elastic domain, which includes pressure and shear-wave velocity or impedance data; and the simulation domain, which focuses on pressure and phase saturation. Among these, the elastic domain is often preferred because it provides a balanced approach to both data preparation and forward modeling [336]. Additionally, several publications have proposed alternative representations of 4D datasets to address the challenges posed by the large volume of data points and data errors. These include dis-

crete images [438, 327], fluid fronts [443, 254], and coefficients derived from data compression methods [280, 146].

The number of methods used in the context of 4D data assimilation is quite varied, including methods such as simulated annealing [206, 457], gradual deformation [303], particle swarm optimization [219, 441], evolutionary algorithms [66, 327], and discrete Latin hypercube sampling [100]. In addition, several works have used derivative-based optimization methods, with derivatives calculated using the forward method [174], adjoint method [111, 134, 446, 225], and streamline sensitivities [460].

The first application of ensemble methods for assimilating 4D seismic data was introduced by Skjervheim et al. [399], who applied EnKF to integrate production and 4D data in an oilfield. Iterative forms of ES have also been tested in several publications, including ES-MDA [133, 118, 323, 475, 320, 321, 370, 125], RLM-MAC [274, 282, 275], and SIES [321, 294].

D.8 Data-Space Inversion

Data-space inversion (DSI) offers an alternative to the traditional model-space inversion commonly used in reservoir data assimilation. The core concept of DSI involves directly updating the predictions derived from a prior ensemble of models to reflect observed production history without modifying the underlying models themselves. This approach enables the generation of an ensemble of forecasts without the need for the time-consuming model updating process. One limitation of DSI is that it only provides forecast estimates, not the corresponding models. In practice, however, reservoir models are useful for evaluating various operational scenarios. Thus, DSI and model-space inversion should be viewed as complementary approaches rather than competing methods.

The DSI method introduced by Sun and Durlofsky [419] employs PCA to reparameterize the predicted data from the prior ensemble into a lower-dimensional space and uses RML to sample from the posterior distribution of predicted data. The authors applied a data transformation prior to PCA to enhance linearity. Sun et al. [421] further extended the DSI method by introducing a more generalized data transformation procedure, and Sun and Durlofsky [420] applied the method to a geological carbon storage problem.

Earlier, Scheidt et al. [381] proposed a similar concept in a method called prediction-focused analysis (PFA), which involves projecting prior predictions into a low-dimensional space and using kernel smoothing to estimate the joint distribution of historical and forecasted data. However, PFA seems to be limited to problems with a small number of data points, as it requires projection into very few dimensions (two or three in the examples presented). Satija and Caers [379] improved the PFA method by incorporating canonical functional component analysis to enhance linearity in the

projected data, and Satija and Caers [380] applied this approach to a reservoir problem, finding that it provided uncertainty estimates for production forecasts in reasonable agreement with rejection sampling.

Lima et al. [263] combined ES-MDA and DSI, resulting in a more robust and efficient DSI implementation. This method has been further enhanced through the use of various deep learning techniques, leading to successful applications in naturally fractured reservoirs [218] and geological carbon storage [469, 217].

D.9 Ensemble Kalman Inversion

Ensemble methods have also been applied to general inverse problems. Iglesias et al. [210] introduced the term ensemble Kalman inversion (EKI), which has since gained significant attention in the inverse problem literature [240, 481, 208, 382, 109]. In practice, the EKI process closely resembles iterative ES. Notably, the regularization scheme used in EKI is essentially the same as that adopted in ES-MDA, with the primary difference being the termination criteria. EKI implementations typically use a regularization criterion based on the discrepancy principle [184], while ES-MDA terminates the data assimilation based on the sum of the reciprocals of the data-error inflation coefficients. Moreover, EKI appears to focus on generating the “best” model rather than quantifying uncertainty in the posterior, which might be one reason for the difference in stopping criteria.

D.10 Ensemble-Based Optimization

Ensemble methods have also been utilized as control optimization techniques, particularly in the petroleum literature. The first publication proposing the use of the EnKF equation for control optimization was presented by Lorentzen et al. [271]. However, the method was later refined by Chen et al. [83], who introduced regularization to the ensemble-gradient estimate and coined the term ensemble-based optimization (EnOpt). Another distinction between EnOpt and the earlier work by Lorentzen et al. [271] is that Chen et al. [83] designed EnOpt to optimize the expected value of an objective function calculated across the ensemble of model realizations.

Using EnOpt, Chen and Oliver [76] achieved the highest net present value in the Brugge benchmark problem [346], drawing significant attention to the method. The idea behind EnOpt is to sample an ensemble of controls around the current solution, use the ensemble to compute the covariance between controls and the optimization objective function, and use this covariance to define the search direction in the steepest ascent method.

The EnOpt method was investigated in later works; for example, Do and Reynolds [110] showed that EnOpt is closely related to previous stochastic optimization methods, such as simultaneous perturbation stochastic approximation (SPSA) [409, 410] and simplex gradient [44, 98]. Fonseca et al. [151] proposed improvements to the EnOpt method in the context of robust optimization and named the method stochastic simplex approximate gradient (StoSAG). These methods have been further investigated and tested in numerous subsequent publications [397, 215, 355, 483, 322].

D.11 Ensemble-Based Methods in Multimodal Distributions

Even for Gaussian priors with Gaussian data errors, the objective function discussed in Chapter 2 (Eq. (2.28)) is strictly quadratic only when the forward model is linear. In the nonlinear case, it is impossible to draw general conclusions about the expected shape of the objective function, particularly because the model space in reservoir data assimilation is extremely high-dimensional. It is conceivable—and even likely—that the objective function is far more complex than quadratic, potentially exhibiting multiple local and global minima. As a result, the posterior PDF may be multimodal.

A well-known example of reservoir data assimilation objective function with multiple minima is the Fault Problem introduced by Tavassoli et al. [433]. This problem involves a cross-sectional model of a reservoir with intercalated high- and low-permeability layers and a single fault. By varying the fault throw and the permeability values, the authors obtained a data-mismatch objective function with a very irregular shape and multiple minima. Zhang et al. [479] also examined the shape of the objective function for a problem involving the estimation of a single channel's position based on pressure data. They concluded that the resulting objective function has multiple minima and is non-differentiable at some locations. Additional examples of multimodal posterior PDFs for synthetic reservoir problems are provided by [293, 480].

A common aspect in these cases is that all examples with multimodal posteriors involve problems with relatively few model parameters. There is limited discussion on the shape of the objective function in cases where reservoir parameters vary on a gridblock-by-gridblock basis. One of the few attempts to explore this is presented in Oliver et al. [334, Chap. 8]. They analyzed a two-dimensional model on an 11×11 grid operating in a five-spot geometry and generated multiple posterior models by assimilating pressure data. By interpolating between distinct posterior models, they observed that the objective function exhibited two minima. Further analysis, however, revealed that these minima were connected by an elongated curved valley with relatively low objective function values.

The non-Gaussian nature of the posterior PDF, including the potential presence of multiple modes, clearly challenges the effectiveness of ensemble methods. This issue is readily demonstrated in low-dimensional toy problems; for example, Fig. 5.4 in Section 5.3.4 shows how applying ES-MDA fails to capture the two modes of the posterior distribution. This limitation stems from the method’s formulation, which is optimal only for Gaussian distributions. In low dimensions, however, this limitation can be partially addressed through strategies involving multiple ensembles. For example, Zhang et al. [480] proposed the iterative local updating ensemble smoother (ILUES), which updates each ensemble member with a sub-ensemble of the closest realizations, enabling it to capture multiple distribution modes effectively in low-dimensional cases. A similar approach is explored by Moyén and Gentilhomme [310]. In small dimensions, even simple reweighting steps, such as importance sampling, are often sufficient to correct the sampling. The main challenge lies in high dimensions, where these techniques yield only marginal improvements over standard iterative ES implementations. Nonetheless, it is unclear that the PDFs generated by real reservoir problems are sufficiently non-Gaussian to render ensemble-based results unreliable. On the contrary, there is ample evidence that ensemble methods provide valuable results in practical reservoir applications.

D.12 Reservoir Applications

D.12.1 Petroleum Reservoirs

Petroleum reservoirs are among the areas with the most extensive application of ensemble methods for data assimilation. Oil and gas companies have been utilizing reservoir simulation models for more than 50 years. As a result, data assimilation—commonly referred to as “history matching” in the petroleum industry—has long been a critical component of reservoir management and decision-making processes.

The literature on assisted history matching methods is extensive and diverse, but the introduction of the EnKF to the petroleum field by Nævdal et al. [313] sparked a surge of research activity. This led to a significant number of publications, primarily focused on evaluating the method in synthetic problems. The first documented field application of the EnKF for reservoir data assimilation was presented by Skjervheim et al. [399]. Additional field applications were later reported by Bianco et al. [36], Evensen et al. [144], Haugen et al. [188], and Emerick and Reynolds [127]. An influential review paper on the EnKF for petroleum reservoirs was published by Aanonsen et al. [3], which outlined the key challenges of the method. This review significantly shaped the research agenda in the years that followed, providing

a baseline for addressing the limitations and improving the application of EnKF in reservoir data assimilation.

The Brugge benchmark case, introduced during a workshop in Brugge in June 2008, marked a significant milestone in the advancement of ensemble methods as viable tools for practical history-matching applications. Notably, the top three participants, ranked by net present value, employed ensemble-based methods for their history matching approaches [346]. In 2011, Oliver and Chen [331] published a comprehensive review on recent advances in reservoir history matching. In their concluding remarks, the authors stated, “*Of all of the methods, EnKF currently comes closest to being a history matching solution for realistic problems.*”

Despite these promising results, the limitations of the EnKF formulation for practical reservoir applications became increasingly apparent. A key issue was the sequential data assimilation scheme, which often introduced inconsistencies between model parameters and states, undermining the method’s effectiveness and increasing the computational cost of the process [435, 459].

Skjervheim et al. [400] proposed using the ES instead of the EnKF for reservoir applications. Although initial attempts with standard ES proved ineffective in subsequent studies [78, 132], this work spurred the development of iterative ES variants [78, 79, 132, 413, 279, 354]. Today, iterative ES methods are widely used for data assimilation in petroleum reservoirs, with both in-house and commercial software integrating these techniques. In addition to proprietary solutions, several open-source tools have gained prominence, with the most notable being the ERT [1]. There are also some Python open-source codes, including PET [2], `resmda` [383], and `pyESMDA` [91].

The number of publications reporting field applications of iterative forms of ES in petroleum reservoirs has become quite large. Some examples include [133, 80, 118, 297, 344, 371, 46, 6, 143, 273, 475, 11, 124, 125].

D.12.2 Groundwater Hydrology

Another field with extensive use of ensemble methods for data assimilation is hydrology, particularly in applications related to groundwater flow and contaminant transport. The earliest applications of ensemble data assimilation in hydrology were introduced by Reichle et al. [360], who used the EnKF to estimate soil moisture profiles. In a subsequent study, Reichle et al. [361] compared the EKF with the EnKF, highlighting the advantages of the ensemble-based approach. The first documented application of the EnKF for groundwater flow calibration was likely presented by Chen and Zhang [82], who used it to calibrate conductivity fields with hydraulic head data. Subsequent studies further explored the potential of the EnKF across a range of groundwater flow and contaminant transport problems [264, 205, 153, 485, 314, 167, 470].

Similar to the early developments in petroleum reservoirs, the EnKF was initially the preferred method for hydrogeology data assimilation. However, recent applications have shifted toward iterative forms of ES, with a significant number of publications evaluating the performance of these methods in a variety of scenarios.

For instance, Chen et al. [74] applied EnKF, ES, and the smoother version of EnRML to estimate the hydraulic conductivity field in the Hanford 300 Area [154] using tracer data. Their findings suggested that EnRML might be a superior alternative to both EnKF and ES for such applications. Kumar et al. [244] utilized an ES implementation to assimilate satellite data from the Gravity Recovery and Climate Experiment (GRACE) to monitor changes in total terrestrial water storage—comprising soil moisture, groundwater, surface waters, snow, ice, and wet biomass—across the United States. Building on this, Li et al. [255] extended the approach to a global scale, aiming to generate time series of groundwater storage for improved drought monitoring. Li et al. [258] also compared EnKF, ES, and iterative ES in a synthetic groundwater data assimilation problem and concluded that iterative ES outperformed standard ES and achieved results comparable to EnKF at a lower computational cost.

Lam et al. [248] compared ES-MDA and LM-EnRML in a synthetic case study for estimating log-hydraulic conductivity fields using transient hydraulic data. Their results showed that LM-EnRML outperformed ES-MDA using an ensemble size of 200. However, they also observed that the accuracy of ES-MDA improved consistently as the ensemble size increased, whereas LM-EnRML did not show the same level of improvement with larger ensembles.

Several studies have applied iterative ES methods for data assimilation in contaminant source identification and transport problems [172]. For example, Tso et al. [444] employed ES-MDA to estimate leak parameters using time-lapse electrical resistivity tomography data. Similarly, Chen et al. [86] used ES-MDA to jointly identify contaminant sources and hydraulic conductivities. In another study, Kang et al. [232] combined ES-MDA with a variational autoencoder to estimate the saturation of dense nonaqueous phase liquid (DNAPL) contamination. Additionally, Chen et al. [85] used ES-MDA to estimate the release history of a contaminant source.

Strategies for updating hydrogeological models generated with multiple-point statistics using iterative ES methods have also been proposed [59, 249, 26, 434]. More recently, open-source data assimilation software for groundwater problems based on ES has emerged. Notable examples include PEST++ [465], DART-PFLOTRAN [216], and *genES*-MDA [440].

D.12.3 Geological Carbon Storage

Geological carbon storage (GCS) is the main element of carbon capture and storage (CCS) and carbon capture, utilization and storage (CCUS) projects, offering a significant mechanism for reducing greenhouse gas emissions and helping to meet climate goals [366]. GCS involves the injection of CO_2 from industrial emissions or captured directly from the air into subsurface formations, such as saline aquifers [61, 155], depleted oil and gas reservoirs [185, 191], or as part of enhanced oil recovery (EOR) processes [180, 326], where it can be stored securely while also potentially enhancing hydrocarbon recovery.

GCS processes present additional challenges compared to petroleum and groundwater modeling. A key distinction is the longer forecast periods, often extending beyond 100 years, and the large-scale models required, both of which significantly increase computational costs. Moreover, GCS simulations must account for more complex physical processes, including near-wellbore effects, intricate fluid interactions, and thermal and geomechanical behavior. Adding to the complexity, GCS projects typically involve fewer wells and less available data than petroleum fields, resulting in higher uncertainties in long-term forecasts.

The existing literature on ensemble data assimilation for GCS focuses primarily on two areas. A set of studies evaluates the performance of these methods in synthetic problems [431, 173, 283, 67, 222, 68]. The second group explores strategies to reduce the computational cost of full-field GCS simulations, including the use of machine learning proxy techniques [427, 426, 385] and DSI schemes [469, 217].

D.12.4 Geothermal Energy

Geothermal energy is a renewable resource that harnesses the Earth's internal heat to provide a continuous, low-emission alternative to conventional energy sources. By extracting heated water from geological reservoirs, geothermal systems can generate electricity or provide direct heating.

The literature on data assimilation for geothermal reservoirs is relatively sparse compared to the extensive studies available for petroleum and groundwater hydrology. Most publications focus on evaluating data assimilation methods with different types of data for conditioning geothermal models. For instance, Marquart et al. [291] tested the EnKF for the assimilation of chemical tracer concentrations and temperature data. Wu et al. [466] explored the assimilation of tracer data using ES-MDA, while Oudshoorn et al. [339] examined the use of ES-MDA to assimilate electromagnetic data. Shariatnik et al. [392] tested the EnKF with electrical resistivity tomography. Békési et al. [32] investigated the use of ES-MDA to assimilate data

in a subsurface temperature model of the onshore Netherlands. There are also recent studies on fractured reservoir characterization [262, 384, 72, 27], adjoint-based RML implementations [437] and a comparison between RML and iterative ES [41].

D.12.5 Seismic Inversion

Seismic inversion is typically defined as the process of estimating subsurface properties such as acoustic impedance, velocity, and density from seismic reflection data. This process is closely related to rock-physics (or petrophysical) inversion, where the goal is to estimate reservoir characteristics like lithology, porosity, fluid content, and facies distribution from seismic data. The terms seismic inversion and rock-physics inversion are often used interchangeably. However, in practice, these processes are usually performed sequentially, with the results of seismic inversion serving as input for rock-physics inversion. That said, there are also joint inversion strategies [176].

The body of literature and the range of methods proposed for seismic inversion are extensive. These methods differ in their representation of rock physics and the inversion techniques employed, spanning from deterministic and stochastic optimization approaches to Monte Carlo methods. Grana et al. [176] presents a recent review of models and methodologies in this area.

Recently, ensemble methods have been investigated for seismic inversion problems. For example, Liu and Grana [265] applied ES-MDA to generate multiple realizations of reservoir properties, including seismic velocities, density, porosity, mineralogy, and saturation from seismic amplitude data. Moya and Gentilhomme [310] used a modified EnRML implementation with sensitivities estimated from sub-ensembles to invert acoustic impedance data into porosity, shale fraction, and fluid saturation. Cao et al. [58] employed ES-MDA with correlation-based localization for joint seismic inversion, while Exterkoetter et al. [147] combined ES-MDA with a variational autoencoder to invert seismic amplitude data into facies. Spremić et al. [411] used the LETKF [207] to invert seismic AVO (amplitude-variation-with-offset) data to clay content and fluid saturation for a petroleum field in the North Sea.

Ensemble methods have also been considered for full waveform inversion (FWI). Unlike traditional seismic inversion, which uses only parts of the seismic data, such as amplitudes or traveltimes, FWI leverages the entire seismic waveform to produce more accurate and detailed subsurface representations [451]. FWI typically employs a forward model based on the finite difference method and involves iteratively adjusting the subsurface model until the predicted seismic waveforms closely match the recorded seismic signals, making the inversion process computationally intensive. As a result,

the overall workflow is quite similar to reservoir data assimilation, making iterative ensemble methods attractive alternatives [436, 169, 168].

References

- [1] Ensemble based reservoir tool (ERT). Software. URL <https://ert.readthedocs.io/en/latest/>.
- [2] Python ensemble toolbox (PET). Software. URL <https://github.com/Python-Ensemble-Toolbox/PET>.
- [3] Aanonsen, S. I., Nævdal, G., Oliver, D. S., Reynolds, A. C., and Vallès, B. Review of ensemble Kalman filter in petroleum engineering. *SPE Journal*, 14(3):393–412, 2009. DOI: [10.2118/117274-PA](https://doi.org/10.2118/117274-PA).
- [4] Aanonsen, S. I., Tveit, S., and Alerini, M. Using Bayesian model probability for ranking different prior scenarios in reservoir history matching. *SPE Journal*, 24(4), 2019. DOI: [10.2118/194505-PA](https://doi.org/10.2118/194505-PA).
- [5] Aanonsen, S. I., Fossum, K., and Mannseth, T. Bayesian model evaluation for multiple scenarios. *Computational Geosciences*, 27, 2023. DOI: [10.1007/s10596-023-10241-2](https://doi.org/10.1007/s10596-023-10241-2).
- [6] Abadpour, A., Adejare, M., Chugunova, T., Mathieu, H., and Haller, N. Integrated geo-modeling and ensemble history matching of complex fractured carbonate and deep offshore turbidite fields, generation of several geologically coherent solutions using ensemble methods. In *Proceedings of the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 12–15 November*, number SPE-193028-MS, 2018. DOI: [10.2118/193028-MS](https://doi.org/10.2118/193028-MS).
- [7] Alabert, F. G. Stochastic imaging of spatial distributions using hard and soft information. Master’s thesis, Stanford University, 1987. URL https://pangea.stanford.edu/departments/ere/dropbox/scrf/documents/Theses/SCRF-Theses/1980-1989/1987_MS_Alabert.pdf.
- [8] Alcolea, A., Carrera, J., and Medina, A. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Advances in Water Resources*, 29:1678–1689, 2006. DOI: [10.1016/j.advwatres.2005.12.009](https://doi.org/10.1016/j.advwatres.2005.12.009).
- [9] Alfonzo, M. A. and Oliver, D. S. Seismic data assimilation with an imperfect model. *Computational Geosciences*, 24:889–905, 2020. DOI: [10.1007/s10596-019-09849-0](https://doi.org/10.1007/s10596-019-09849-0).
- [10] Allam, F. A., El-Banbi, A. H., Bustami, S. S., Saada, T. H., and Fahmy, I. I. History match tuning through different upscaling algorithms. In *Proceedings of the SPE Annual Technical Confer-*

- ence and Exhibition, Houston, Texas, number SPE-90292-MS, 2004. DOI: [10.2118/90292-MS](#).
- [11] Alqallabi, S., Gacem, M. T., Al-Jenaibi, F., Tahir, S., Ameri, S. M., Ouzzane, D. E., Adli, M., Malla, L., Cornejo, V. O. S., Almarzooqi, A., Sunagatullin, R., Hidayati, S. R., Voleti, D. K., Messabi, F. A., Montani, H., Shibasaki, T., Tani, K., Mohamed, F., Sankararaj, R., Ducroux, S., Wojnar, K., and Gourc, L. Key learnings from application of ensemble-based modelling on three giant carbonate oilfields for optimal economic decisions under static and dynamic uncertainties. In *Proceedings of the ADIPEC, Abu Dhabi, UAE, October*, number SPE-211409-MS, 2022. DOI: [10.2118/211409-MS](#).
- [12] Amaefule, J. O., Altunbay, M., Tiab, D., Kersey, D. G., and Keelan, D. K. Enhanced reservoir description: Using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/wells. In *Proceeding of the SPE Annual Technical Conference and Exhibition, Houston, Texas, October*, number SPE-26436-MS, 1993. DOI: [10.2118/26436-MS](#).
- [13] Anderson, J. L. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129(12):2884–2903, 2001. DOI: [10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](#).
- [14] Anderson, J. L. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, 230(1–2):99–111, 2007. DOI: [10.1016/j.physd.2006.02.011](#).
- [15] Anderson, J. L. and Anderson, S. L. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741–2758, 1999. DOI: [10.1175/1520-0493\(1999\)127<2741:AMCIOT>2.0.CO;2](#).
- [16] Anterion, F., Karcher, B., and Eymard, R. Use of parameter gradients for reservoir history matching. In *Proceedings of the 10th SPE Reservoir Simulation Symposium, Houston, Texas, 6–8 February*, number SPE-18433-MS, 1989. DOI: [10.2118/18433-MS](#).
- [17] Armstrong, M., Galli, A., Beucher, H., Loc’h, G. L., Renard, D., Doligez, B., Eschard, R., and Geffroy, F. *Plurigaussian simulations in geosciences*. Springer-Verlag Berlin Heidelberg, 2nd edition, 2011. DOI: [10.1007/978-3-642-19607-2](#).
- [18] Armstrong, M., Vincent, A., Galli, A. G., and Méheut, C. Genetic algorithms and scenario reduction. *Journal of The Southern African Institute of Mining and Metallurgy*, 114(3):237–244, 2014.
- [19] Arroyo-Negrete, E., Devegowda, D., Datta-Gupta, A., and Choe, J. Streamline-assisted ensemble Kalman filter for rapid and continuous reservoir model updating. *SPE Reservoir Evaluation & Engineering*, 11(6):1046–1060, 2008. DOI: [10.2118/104255-PA](#).
- [20] Avansi, G. D. and Schiozer, D. J. UNISIM-I: synthetic model for reservoir development and management applications. *International Journal of Modeling and Simulation for the Petroleum Industry*,

- 9(1):21–30, 2015. URL <http://www.ijmspi.org/ojs/index.php/ijmspi/article/view/152>.
- [21] Aziz, K. and Settari, A. *Petroleum reservoir simulation*. Elsevier Applied Science Publishers, London, 1979.
 - [22] Babin, V., Iaubatyrov, R., Ushmaev, O., Garcia, D. K., Golitsyna, M., Semenikhin, A., and Ciaurri, D. E. A variant of particle swarm optimization for uncertainty quantification. In *Proceedings of the SPE Reservoir Characterisation and Simulation Conference and Exhibition, Abu Dhabi, UAE, 8–10 May*, number SPE-186024-MS, 2017. DOI: [10.2118/186024-MS](https://doi.org/10.2118/186024-MS).
 - [23] Bahrami, P., Moghaddam, F. S., and James, L. A. A review of proxy modeling highlighting applications for reservoir engineering. *Energies*, 15(14), 2022. DOI: [10.3390/en15145247](https://doi.org/10.3390/en15145247).
 - [24] Bai, X., Wang, Y., and Zhang, W. Applying physics informed neural network for flow data assimilation. *Journal of Hydrodynamics*, 32, 2020. DOI: [10.1007/s42241-020-0077-2](https://doi.org/10.1007/s42241-020-0077-2).
 - [25] Bai, Z. D. and Yin, Y. Q. Convergence to the semicircle law. *The Annals of Probability*, 16(2), 1988. URL <http://www.jstor.org/stable/2243844>.
 - [26] Bao, J., Li, L., and Redoloza, F. Coupling ensemble smoother and deep learning with generative adversarial networks to deal with non-Gaussianity in flow and transport data assimilation. *Journal of Hydrology*, 590, 2020. DOI: [10.1016/j.jhydrol.2020.125443](https://doi.org/10.1016/j.jhydrol.2020.125443).
 - [27] Bao, J., Lee, J., and Yoon, H. Enhanced geothermal site characterization using generative adversarial network and ensemble method. In ARMA-2024-0579, editor, *Proceedings of the 58th U.S. Rock Mechanics/Geomechanics Symposium, Golden, Colorado, USA*, 2024. DOI: [10.56952/ARMA-2024-0579](https://doi.org/10.56952/ARMA-2024-0579).
 - [28] Bardsley, J. M., Solonen, A., Haario, H., and Laine, M. Randomize-then-optimize: a method for sampling from posterior distributions in nonlinear inverse problems. *SIAM Journal on Scientific Computing*, 36(4):A1895–A1910, 2014. DOI: [10.1137/140964023](https://doi.org/10.1137/140964023).
 - [29] Barker, J. W. and Dupouy, P. An analysis of dynamic pseudo-relative permeability methods for oil-water flows. *Petroleum Geoscience*, 5: 385–394, 1999. DOI: [10.1144/petgeo.5.4.38](https://doi.org/10.1144/petgeo.5.4.38).
 - [30] Bastian, P., Kraus, J., Scheich, R., and Wheeler, M., editors. *Simulation of Flow in Porous Media: Applications in Energy and Environment*. Series on Computational and Applied Mathematics Book 12. De Gruyter, 2013. ISBN 978-3110282245.
 - [31] Batzle, M. and Wang, Z. Seismic properties of pore fluids. *Geophysics*, 57(11):1396–1408, 1992. DOI: [10.1190/1.1443207](https://doi.org/10.1190/1.1443207).
 - [32] Békési, E., Struijk, M., Bonté, D., Veldkamp, H., Limberger, J., Fokker, P. A., Vrijlandt, M., and van Wees, J.-D. An updated geothermal model of the dutch subsurface based on inversion of temperature data. *Geothermics*, 88, 2020. DOI: [10.1016/j.geothermics.2020.101880](https://doi.org/10.1016/j.geothermics.2020.101880).

- [33] Bellman, R. E. *Dynamic Programming*. Princeton University Press, 1957.
- [34] Bertino, L., Evensen, G., and Wackernagel, H. Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71(2), 2003. DOI: [10.1111/j.1751-5823.2003.tb00194.x](https://doi.org/10.1111/j.1751-5823.2003.tb00194.x).
- [35] Beucher, H. and Renard, D. Truncated Gaussian and derived methods. *Mathematical Geology*, 348(7), 2016. DOI: [10.1016/j.crte.2015.10.004](https://doi.org/10.1016/j.crte.2015.10.004).
- [36] Bianco, A., Cominelli, A., Dovera, L., Nævdal, G., and Vallès, B. History matching and production forecast uncertainty by means of the ensemble Kalman filter: A real field application. In *Proceedings of the EAGE/EUROPEC Conference and Exhibition, London, U.K., 11–14 June*, number SPE-107161-MS, 2007. DOI: [10.2118/107161-MS](https://doi.org/10.2118/107161-MS).
- [37] Bishop, C. H. and Hodyss, D. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 133(629): 2029–2044, 2007. DOI: [10.1002/qj.169](https://doi.org/10.1002/qj.169).
- [38] Bishop, C. H. and Hodyss, D. Ensemble covariances adaptively localized with ECO-RAP. part 1: tests on simple error models. *Tellus*, 61: 84–96, 2009. DOI: [10.1111/j.1600-0870.2008.00371.x](https://doi.org/10.1111/j.1600-0870.2008.00371.x).
- [39] Bishop, C. H., Etherton, B. J., and Majumdar, S. J. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129:420–436, 2001. DOI: [10.1175/1520-0493\(2001\)129<0420:ASWTET>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2).
- [40] Bissell, R. Calculating optimal parameters for history matching. In *Proceedings of the 4th European Conference on the Mathematics of Oil Recovery (ECMOR IV), 07 June*, 1994. DOI: [10.3997/2214-4609.201411181](https://doi.org/10.3997/2214-4609.201411181).
- [41] Bjarkason, E. K., Maclaren, O. J., Nicholson, R., Yeh, A., and O’Sullivan, M. J. Uncertainty quantification of highly-parameterized geothermal reservoir models using ensemble-based methods. In *Proceedings World Geothermal Congress 2020+1*, 2021. URL <https://hdl.handle.net/2292/65029>.
- [42] Bocquet, M. and Sakov, P. Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems. *Nonlinear Processes in Geophysics*, 19(3), 2012. DOI: [10.5194/npg-19-383-2012](https://doi.org/10.5194/npg-19-383-2012).
- [43] Boisvert, J. B. *Geostatistics with locally varying anisotropy*. PhD thesis, University of Alberta, 2010.
- [44] Bortz, D. M. and Kelley, C. T. The simplex gradient and noisy optimization problems. In Borggaard, T., J. Burns, E. C., and Schreck, S., editors, *Computational Methods for Optimal Design and Control*. Birkhäuser Boston, 1998. ISBN 978-1461217800. DOI: [10.1007/978-1-4612-1780-0_5](https://doi.org/10.1007/978-1-4612-1780-0_5).
- [45] Boser, B. E., Guyon, I. M., and Vapnik, V. N. A training algorithm for optimal margin classifiers. In *COLT ’92 Proceedings of the fifth annual workshop on Computational learning theory Pittsburgh, Pennsylvania, USA, July 27–29*, pages 144–152, 1992. DOI: [10.1145/130385.130401](https://doi.org/10.1145/130385.130401).

- [46] Breslavich, I. D., Sarkisov, G. G., and Makarova, E. S. Experience of MDA ensemble smoother practice for Volga-Ural oilfield. In *Proceedings of the SPE Russian Petroleum Technology Conference held in Moscow, Russia, 16–18 October*, number SPE-187800-MS, 2017. DOI: [10.2118/187800-MS](https://doi.org/10.2118/187800-MS).
- [47] Bruhn, C. H. L., Gomes, J. A. T., Luchesse, C. D., and Johann, P. R. S. Campos Basin: Reservoir characterization and management – historical overview and future challenges. In *Proceedings of the Offshore Technology Conference, Houston, Texas*, number OTC 15220, 2003. DOI: [10.4043/15220-MS](https://doi.org/10.4043/15220-MS).
- [48] Bruyelle, J. and Guérillot, D. Neural networks and their derivatives for history matching and reservoir optimization problems. *Computational Geosciences*, 18(3–4):549–561, 2014. DOI: [10.1007/s10596-013-9390-y](https://doi.org/10.1007/s10596-013-9390-y).
- [49] Brynjarsdóttir, J. and O’Hagan, A. Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30, 2014. DOI: [10.1088/0266-5611/30/11/114007](https://doi.org/10.1088/0266-5611/30/11/114007).
- [50] Buckley, S. E. and Leverett, M. C. Mechanism of fluid displacement in sands. *Transactions of the AIME*, 146(1):107–116, 1942. DOI: [10.2118/942107-G](https://doi.org/10.2118/942107-G).
- [51] Burgers, G., van Leeuwen, P., and Evensen, G. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6):1719–1724, 1998. DOI: [10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2).
- [52] Burgess, P. M. *Regional Geology and Tectonics: Principles of Geologic Analysis*, chapter A brief review of developments in stratigraphic forward modelling, 2000–2009, pages 378–404. Elsevier, 2012. ISBN 978-0444530424. DOI: <https://doi.org/10.1016/B978-0-444-53042-4.00014-5>.
- [53] Caers, J. *Bayesianism in the Geosciences. In Handbook of Mathematical Geosciences: Fifty Years of IAMG*, pages 527–566. Springer International Publishing, 2018. DOI: [10.1007/978-3-319-78999-6_27](https://doi.org/10.1007/978-3-319-78999-6_27).
- [54] Camacho, A., Talavera, A., Emerick, A. A., Pacheco, M. A., and Zannia, J. Uncertainty quantification in reservoir simulation models with polynomial chaos expansions: Smolyak quadrature and regression method approach. *Journal of Petroleum Science and Engineering*, 153: 203–211, 2017. DOI: [10.1016/j.petrol.2017.03.046](https://doi.org/10.1016/j.petrol.2017.03.046).
- [55] Canchumuni, S. W. A., Emerick, A. A., and Pacheco, M. A. C. History matching geological facies models based on ensemble smoother and deep generative models. *Journal of Petroleum Science and Engineering*, 177:941–958, 2019. DOI: [10.1016/j.petrol.2019.02.037](https://doi.org/10.1016/j.petrol.2019.02.037).
- [56] Canchumuni, S. W. A., Emerick, A. A., and Pacheco, M. A. C. Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers & Geosciences*, 128:87–102, 2019. DOI: [10.1016/j.cageo.2019.04.006](https://doi.org/10.1016/j.cageo.2019.04.006).
- [57] Canchumuni, S. W. A., Castro, J. D. B., Potratz, J., Emerick, A. A., and Pacheco, M. A. C. Recent developments combining ensemble smoother and deep generative networks for facies history matching.

- Computational Geosciences*, 25(433–466), 2021. DOI: [10.1007/s10596-020-10015-0](https://doi.org/10.1007/s10596-020-10015-0).
- [58] Cao, Y., Zhou, H., Yu, B., Wei, S., Chen, H., and Tian, Y. The estimation of petrophysical parameters based on ensemble smoother with correlation localization. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 2024. DOI: [10.1109/TGRS.2024.3403663](https://doi.org/10.1109/TGRS.2024.3403663).
- [59] Cao, Z., Li, L., and Chen, K. Bridging iterative ensemble smoother and multiple-point geostatistics for better flow and transport modeling. *Journal of Hydrology*, 565, 2018. DOI: [10.1016/j.jhydrol.2018.08.023](https://doi.org/10.1016/j.jhydrol.2018.08.023).
- [60] Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Climate Change*, 9(8), 2018. DOI: [10.1002/wcc.535](https://doi.org/10.1002/wcc.535).
- [61] Celia, M. A., Bachu, S., Nordbotten, J. M., and Bandilla, K. W. Status of CO₂ storage in deep saline aquifers with emphasis on modeling approaches and practical simulations. *Water Resources Research*, 51(9), 2015. DOI: [10.1002/2015WR017609](https://doi.org/10.1002/2015WR017609).
- [62] Chan, S. and Elsheikh, A. H. Parametrization and generation of geological models with generative adversarial networks. *arXiv:1708.01810v1 [stat.ML]*, 2017. URL <https://arxiv.org/abs/1708.01810>.
- [63] Chan, S. and Elsheikh, A. H. Parametric generation of conditional geological realizations using generative neural networks. *Computational Geosciences*, 23:925–952, 2019. DOI: [10.1007/s10596-019-09850-7](https://doi.org/10.1007/s10596-019-09850-7).
- [64] Chang, H., Zhang, D., and Lu, Z. History matching of facies distributions with the EnKF and level set parameterization. *Journal of Computational Physics*, 229:8011–8030, 2010. DOI: [10.1016/j.jcp.2010.07.005](https://doi.org/10.1016/j.jcp.2010.07.005).
- [65] Chang, Y., Stordal, A. S., and Valestran, R. Facies parameterization and estimation for complex reservoirs – the Brugge field. In *Proceedings of the SPE Bergen One Day Seminar, Bergen, Norway, 22 April, number SPE-173872-MS*, 2015. DOI: [10.2118/173872-MS](https://doi.org/10.2118/173872-MS).
- [66] Chassagne, R., Obidegwu, D., Dambrine, J., and MacBeth, C. Binary 4D seismic history matching, a metric study. *Computers & Geosciences*, 96, 2016. DOI: [10.1016/j.cageo.2016.08.013](https://doi.org/10.1016/j.cageo.2016.08.013).
- [67] Chen, B., Harp, D. R., Lu, Z., and Pawar, R. J. Reducing uncertainty in geologic CO₂ sequestration risk assessment by assimilating monitoring data. *International Journal of Greenhouse Gas Control*, 94, 2020. DOI: [10.1016/j.ijggc.2019.102926](https://doi.org/10.1016/j.ijggc.2019.102926).
- [68] Chen, B., Morales, M. M., Ma, Z., Kang, Q., and Pawar, R. J. Assimilation of geophysics-derived spatial data for model calibration in geologic CO₂ sequestration. *SPE Journal*, 29(7), 2024. DOI: [10.2118/212975-PA](https://doi.org/10.2118/212975-PA).
- [69] Chen, C., Gao, G., Honorio, J., Gelderblom, P., Jimenez, E., and Jaakkola, T. Integration of principal-component-analysis and streamline information for the history matching of channelized reservoirs. In *Proceedings of the SPE Annual Technical Conference and Exhibition*,

- Amsterdam, The Netherlands, 27–29 October*, number SPE-170636-MS, 2014. DOI: [10.2118/170636-MS](https://doi.org/10.2118/170636-MS).
- [70] Chen, C., Gao, G., Ramirez, B. A., Vink, J. C., and Girardi, A. M. Assisted history matching of channelized models using pluri-principal component analysis. In *Proceedings of the SPE Reservoir Simulation Symposium, Houston, Texas, USA, 23–25 February*, number SPE-173192-MS, 2015. DOI: [10.2118/173192-MS](https://doi.org/10.2118/173192-MS).
 - [71] Chen, C., Gao, G., Gelderblom, P., and Jimenez, E. Integration of cumulative-distribution-function mapping with principal-component analysis for the history matching of channelized reservoirs. *SPE Reservoir Evaluation & Engineering*, 19(2):278–293, 2016. DOI: [10.2118/170636-PA](https://doi.org/10.2118/170636-PA).
 - [72] Chen, C., Deng, Y., Ma, H., Kang, X., Ma, L., and Qian, J. Deep learning-based inversion framework by assimilating hydrogeological and geophysical data for an enhanced geothermal system characterization and thermal performance prediction. *Energy*, 302, 2024. DOI: [10.1016/j.energy.2024.131713](https://doi.org/10.1016/j.energy.2024.131713).
 - [73] Chen, W. H., Gavalas, G. R., Seinfeld, J. H., and Wasserman, M. L. A new algorithm for automatic history matching. *SPE Journal*, 14(6): 593–608, 1974. DOI: [10.2118/4545-PA](https://doi.org/10.2118/4545-PA).
 - [74] Chen, X., Hammond, G. E., Murray, C. J., Rockhold, M. L., Vermeul, V. R., and Zachara, J. M. Application of ensemble-based data assimilation techniques for aquifer characterization using tracer data at hanford 300 area. *Water Resources Research*, 49(10), 2013. DOI: [10.1002/2012WR013285](https://doi.org/10.1002/2012WR013285).
 - [75] Chen, Y. Keynote - geologically consistent history matching using the ensemble based methods. In *Proceedings of the Petroleum Geostatistics, Biarritz, France, 7–11 September*, 2015. DOI: [10.3997/2214-4609.201413627](https://doi.org/10.3997/2214-4609.201413627).
 - [76] Chen, Y. and Oliver, D. S. Ensemble-based closed-loop optimization applied to Brugge field. *SPE Reservoir Evaluation & Engineering*, 13(1):56–71, 2010. DOI: [10.2118/118926-PA](https://doi.org/10.2118/118926-PA).
 - [77] Chen, Y. and Oliver, D. S. Cross-covariance and localization for EnKF in multiphase flow data assimilation. *Computational Geosciences*, 14(4):579–601, 2010. DOI: [10.1007/s10596-009-9174-6](https://doi.org/10.1007/s10596-009-9174-6).
 - [78] Chen, Y. and Oliver, D. S. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, 44(1):1–26, 2012. DOI: [10.1007/s11004-011-9376-z](https://doi.org/10.1007/s11004-011-9376-z).
 - [79] Chen, Y. and Oliver, D. S. Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computational Geosciences*, 17:689–703, 2013. DOI: [10.1007/s10596-013-9351-5](https://doi.org/10.1007/s10596-013-9351-5).
 - [80] Chen, Y. and Oliver, D. S. History matching of the Norne full-field model with an iterative ensemble smoother. *SPE Reservoir Evaluation & Engineering*, 17(2), 2014. DOI: [10.2118/164902-PA](https://doi.org/10.2118/164902-PA).

- [81] Chen, Y. and Oliver, D. S. Localization and regularization for iterative ensemble smoothers. *Computational Geosciences*, 21(1):13–30, 2017. DOI: [10.1007/s10596-016-9599-7](https://doi.org/10.1007/s10596-016-9599-7).
- [82] Chen, Y. and Zhang, D. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources*, 29(8):1107–1122, 2006. DOI: [10.1016/j.advwatres.2005.09.007](https://doi.org/10.1016/j.advwatres.2005.09.007).
- [83] Chen, Y., Oliver, D. S., and Zhang, D. Efficient ensemble-based closed-loop production optimization. *SPE Journal*, 14(4):634–645, 2009. DOI: [10.2118/112873-PA](https://doi.org/10.2118/112873-PA).
- [84] Chen, Z., Huan, G., and Ma, Y. *Computational Methods for Multiphase Flows in Porous Media*. SIAM, 2006. ISBN 978-0898718942. DOI: [10.1137/1.9780898718942.fm](https://doi.org/10.1137/1.9780898718942.fm).
- [85] Chen, Z., Xu, T., Gómez-Hernández, J. J., Zanini, A., and Zhou, Q. Reconstructing the release history of a contaminant source with different precision via the ensemble smoother with multiple data assimilation. *Journal of Contaminant Hydrology*, 252, 2023. DOI: [10.1016/j.jconhyd.2022.104115](https://doi.org/10.1016/j.jconhyd.2022.104115).
- [86] Chen, Z., Zong, L., Gómez-Hernández, J. J., Xu, T., Jiang, Y., Zhou, Q., Yang, H., Jia, Z., and Mei, S. Contaminant source and aquifer characterization: An application of ES-MDA demonstrating the assimilation of geophysical data. *Advances in Water Resources*, 181, 2023. DOI: [10.1016/j.advwatres.2023.104555](https://doi.org/10.1016/j.advwatres.2023.104555).
- [87] Chrysikopoulos, C. Artificial tracers for geothermal reservoir studies. *Environmental Geology*, 22, 1993. DOI: [10.1007/BF00775286](https://doi.org/10.1007/BF00775286).
- [88] Cipra, B. A. The best of the 20th Century: editors name top 10 algorithms. *SIAM News*, 33(4), 2000.
- [89] Cojan, I., Olivier Fouché, S. L., and Rivoirard, J. *Geostatistics Banff 2004*, chapter Process-based Reservoir Modelling in the Example of Meandering Channel, pages 611–619. Springer, Dordrecht, 2004. ISBN 978-1402036101. DOI: [10.1007/978-1-4020-3610-1_62](https://doi.org/10.1007/978-1-4020-3610-1_62).
- [90] Coley, D. A. *An introduction to genetic algorithms for scientists and engineers*. World Scientific, 1999. ISBN 978-9810236021. DOI: [10.1142/3904](https://doi.org/10.1142/3904).
- [91] Collet, A. pyESMDA – Python ensemble smoother with multiple data assimilation. Software. URL <https://pypi.org/project/pyesmda/>.
- [92] Conn, A. R., Scheinberg, K., and Vicente, L. N. *Introduction to derivative free optimization*. SIAM, Philadelphia, PA, 2009. DOI: [10.1137/1.9780898718768](https://doi.org/10.1137/1.9780898718768).
- [93] Costa, L. A. N., Maschio, C., and Schiozer, D. J. Application of artificial neural networks in a history matching process. *Journal of Petroleum Science and Engineering*, 123, 2014. DOI: [10.1016/j.petrol.2014.06.004](https://doi.org/10.1016/j.petrol.2014.06.004).
- [94] Coutinho, E. J., Emerick, A. A., Li, G., and Reynolds, A. C. Conditioning multi-layered geologic models to well test and production logging data using the ensemble Kalman filter. In *Proceedings of the SPE*

- Annual Technical Conference and Exhibition, Florence, Italy, 19–22 September*, number SPE-134542-MS, 2010. DOI: [10.2118/134542-MS](https://doi.org/10.2118/134542-MS).
- [95] Coutinho, E. J. R., Dall’Aqua, and Gildin, E. Physics-aware deep-learning-based proxy reservoir simulation model equipped with state and well output prediction. *Frontiers in Applied Mathematics and Statistics*, 7, 2021. DOI: [10.3389/fams.2021.651178](https://doi.org/10.3389/fams.2021.651178).
- [96] Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. Pressure matching for hydrocarbon reservoirs: A case study in the use of bayes linear strategies for large computer experiments. In *Case Studies in Bayesian Statistics. Lecture Notes in Statistics*, volume 121. Springer, 1997. DOI: [10.1007/978-1-4612-2290-3_2](https://doi.org/10.1007/978-1-4612-2290-3_2).
- [97] Cruz, N. M., Cruz, J. M., Teixeira, L. M., da Costa, M. M., de Oliveira, L. B., Urasaki, E. N., Bispo, T. P., de Sá Jardim, M., Grochau, M. H., and Maul, A. Tupi nodes pilot: A successful 4D seismic case for Brazilian presalt reservoirs. *The Leading Edge*, 40(12):886–896, 2021. DOI: [10.1190/tle40120886.1](https://doi.org/10.1190/tle40120886.1).
- [98] Custódio, A. L. and Vicente, L. N. Using sampling and simplex derivatives in pattern search methods. *SIAM Journal on Optimization*, 18: 537–555, 2007. DOI: [10.1137/050646706](https://doi.org/10.1137/050646706).
- [99] Dasgupta, A., Hostache, R., Ramsankaran, R., Grimaldi, S., Matgen, P., Chini, M., Pauwels, V. R., and Walker, J. P. Earth observation and hydraulic data assimilation for improved flood inundation forecasting. In Schumann, G. J.-P., editor, *Earth Observation for Flood Applications*. Elsevier, 2021. DOI: [10.1016/B978-0-12-819412-6.00012-2](https://doi.org/10.1016/B978-0-12-819412-6.00012-2).
- [100] Davolio, A. and Schiozer, D. J. Probabilistic seismic history matching using binary images. *Journal of Geophysics and Engineering*, 15(1), 2018. DOI: [10.1088/1742-2140/aa99f4](https://doi.org/10.1088/1742-2140/aa99f4).
- [101] de Marsily, G., Lavedan, G., Boucher, M., and Fasanino, G. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In Verly, G., David, M., Journel, A. G., and Marechal, A., editors, *Geostatistics for Natural Resources Characterization, Proceedings of the NATO Advanced Study Institute*, pages 831–849. Dordrecht, Holland, Dordrecht, Holland, 1984. DOI: [10.1007/978-94-009-3701-7_16](https://doi.org/10.1007/978-94-009-3701-7_16).
- [102] de Moraes, R. J., Rodrigues, J. R. P., Hajibeygi, H., and Jansen, J.-D. Multiscale gradient computation for flow in heterogeneous porous media. *Journal of Computational Physics*, 336, 2017. DOI: [10.1016/j.jcp.2017.02.024](https://doi.org/10.1016/j.jcp.2017.02.024).
- [103] de Moraes, R. J., de Zeeuw, W., Rodrigues, J. R. P., Hajibeygi, H., and Jansen, J.-D. Iterative multiscale gradient computation for heterogeneous subsurface flow. *Advances in Water Resources*, 129, 2019. DOI: [10.1016/j.advwatres.2019.05.016](https://doi.org/10.1016/j.advwatres.2019.05.016).
- [104] Deutsch, C. V. *Geostatistical reservoir modeling*. Oxford University Press, 2002.

- [105] Deutsch, C. V. and Journel, A. G. *GSLIB: Geostatistical software library and user's guide*. Oxford University Press, New York, 1992. ISBN 978-0195073928.
- [106] Deutsch, C. V. and Journel, A. G. *GSLIB: Geostatistical software library and user's guide*. Oxford University Press, New York, 2nd edition, 1997. ISBN 978-0195100150.
- [107] Deutsch, C. V. and Wang, L. Hierarchical object-based stochastic modeling of fluvial reservoirs. *Mathematical Geology*, 28(7):857–880, 1996. DOI: [10.1007/BF02066005](https://doi.org/10.1007/BF02066005).
- [108] Devegowda, D., Arroyo-Negrete, E., Datta-Gupta, A., and Douma, S. G. Efficient and robust reservoir model updating using ensemble Kalman filter with sensitivity-based covariance localization. In *Proceedings of the SPE Reservoir Simulation Symposium, Houston, Texas, USA 26–28 February*, number SPE-106144-MS, 2007. DOI: [10.2118/106144-MS](https://doi.org/10.2118/106144-MS).
- [109] Ding, Z. and Li, Q. Ensemble Kalman inversion: mean – field limit and convergence analysis. *Statistics and Computing*, 31(9), 2021. DOI: [10.1007/s11222-020-09976-0](https://doi.org/10.1007/s11222-020-09976-0).
- [110] Do, S. T. and Reynolds, A. C. Theoretical connections between optimization algorithms based on an approximate gradient. *Computational Geosciences*, 17(6):959–973, 2013. DOI: [10.1007/s10596-013-9368-9](https://doi.org/10.1007/s10596-013-9368-9).
- [111] Dong, Y. and Oliver, D. S. Quantitative use of 4D seismic data for reservoir description. *SPE Journal*, 10(1):51–65, 2005.
- [112] Donoho, D. L. and Johnstone, I. M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994. DOI: [10.1093/biomet/81.3.425](https://doi.org/10.1093/biomet/81.3.425).
- [113] Donoho, D. L. and Johnstone, I. M. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1224, 1995. DOI: [10.1080/01621459.1995.10476626](https://doi.org/10.1080/01621459.1995.10476626).
- [114] Doucet, A., de Freitas, N., and Gordon, N., editors. *Sequential Monte Carlo methods in practice*. Springer-Verlag, 2000. ISBN 978-1475734379. DOI: [10.1007/978-1-4757-3437-9](https://doi.org/10.1007/978-1-4757-3437-9).
- [115] Dupont, E., Zhang, T., Tilke, P., Liang, L., and Bailey, W. Generating realistic geology conditioned on physical measurements with generative adversarial networks. *arXiv:1802.03065v3 [stat.ML]*, 2018. URL <https://arxiv.org/abs/1802.03065>.
- [116] Durlafsky, L. J. Use of higher moments for the description of up-scaled, process independent relative permeabilities. *SPE Journal*, 2(4):474–484, 1997. DOI: [10.2118/37987-PA](https://doi.org/10.2118/37987-PA).
- [117] Elsheikh, A. H., Hoteit, I., and Wheeler, M. F. Efficient bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates. *Computer Methods in Applied Mechanics and Engineering*, 269, 2014. DOI: [10.1016/j.cma.2013.11.001](https://doi.org/10.1016/j.cma.2013.11.001).
- [118] Emerick, A. A. Analysis of the performance of ensemble-based assimilation of production and seismic data. *Jour-*

- nal of Petroleum Science and Engineering*, 139:219–239, 2016. DOI: [10.1016/j.petrol.2016.01.029](https://doi.org/10.1016/j.petrol.2016.01.029).
- [119] Emerick, A. A. Estimating uncertainty bounds in field production using ensemble-based methods. *Journal of Petroleum Science and Engineering*, 145:648–656, 2016. DOI: [10.1016/j.petrol.2016.06.037](https://doi.org/10.1016/j.petrol.2016.06.037).
 - [120] Emerick, A. A. Investigation on principal component analysis parameterizations for history matching channelized facies models with ensemble-based data assimilation. *Mathematical Geosciences*, 49(1): 85–120, 2017. DOI: [10.1007/s11004-016-9659-5](https://doi.org/10.1007/s11004-016-9659-5).
 - [121] Emerick, A. A. Deterministic ensemble smoother with multiple data assimilation as an alternative for history matching seismic data. *Computational Geosciences*, 22(5):1175–1186, 2018. DOI: [10.1007/s10596-018-9745-5](https://doi.org/10.1007/s10596-018-9745-5).
 - [122] Emerick, A. A. Analysis of geometric selection of the data-error covariance inflation for ES-MDA. *Journal of Petroleum Science and Engineering*, 182, 2019. DOI: [10.1016/j.petrol.2019.06.032](https://doi.org/10.1016/j.petrol.2019.06.032).
 - [123] Emerick, A. A. Practical considerations in the application of ensemble smoother for assimilating production and 4D seismic data. In *Proceedings of the European Conference for Mathematics of Geological Reservoirs ECMOR, September 2-5, Oslo, Norway*, 2024. DOI: [10.3997/2214-4609.202437027](https://doi.org/10.3997/2214-4609.202437027).
 - [124] Emerick, A. A. and Neto, G. M. S. Investigation on the production data frequency for assimilation with ensemble smoother. *Geoenergy Science and Engineering*, 231, 2023. DOI: [10.1016/j.geoen.2023.212356](https://doi.org/10.1016/j.geoen.2023.212356).
 - [125] Emerick, A. A. and Neto, G. M. S. Projection of 4D seismic onto the ensemble observation subspace for data assimilation. *Geoenergy Science and Engineering*, 237, 2024. DOI: [10.1016/j.geoen.2024.212835](https://doi.org/10.1016/j.geoen.2024.212835).
 - [126] Emerick, A. A. and Reynolds, A. C. Combining the ensemble Kalman filter with Markov chain Monte Carlo for improved history matching and uncertainty characterization. In *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA, 21–23 February*, number SPE-141336-MS, 2011. DOI: [10.2118/141336-MS](https://doi.org/10.2118/141336-MS).
 - [127] Emerick, A. A. and Reynolds, A. C. History matching a field case using the ensemble Kalman filter with covariance localization. *SPE Reservoir Evaluation & Engineering*, 14(4):423–432, 2011. DOI: [10.2118/141216-PA](https://doi.org/10.2118/141216-PA).
 - [128] Emerick, A. A. and Reynolds, A. C. Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Computational Geosciences*, 15 (2):251–269, 2011. DOI: [10.1007/s10596-010-9198-y](https://doi.org/10.1007/s10596-010-9198-y).
 - [129] Emerick, A. A. and Reynolds, A. C. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 16(3):639–659, 2012. DOI: [10.1007/s10596-012-9275-5](https://doi.org/10.1007/s10596-012-9275-5).

- [130] Emerick, A. A. and Reynolds, A. C. Combining the ensemble Kalman filter with Markov chain Monte Carlo for improved history matching and uncertainty characterization. *SPE Journal*, 17(2):418–440, 2012. DOI: [10.2118/141336-PA](https://doi.org/10.2118/141336-PA).
- [131] Emerick, A. A. and Reynolds, A. C. Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. *Computational Geosciences*, 17(2):325–350, 2013. DOI: [10.1007/s10596-012-9333-z](https://doi.org/10.1007/s10596-012-9333-z).
- [132] Emerick, A. A. and Reynolds, A. C. Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, 55:3–15, 2013. DOI: [10.1016/j.cageo.2012.03.011](https://doi.org/10.1016/j.cageo.2012.03.011).
- [133] Emerick, A. A. and Reynolds, A. C. History matching of production and seismic data for a real field case using the ensemble smoother with multiple data assimilation. In *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA, 18–20 February*, number SPE-163675-MS, 2013. DOI: [10.2118/163675-MS](https://doi.org/10.2118/163675-MS).
- [134] Emerick, A. A., Moraes, R. J., and Rodrigues, J. R. P. History matching 4D seismic data with efficient gradient based methods. In *Proceedings of the EUROPEC/EAGE Conference and Exhibition, 11–14 June, London, U.K.*, number SPE-107179-MS, 2007. DOI: [10.2118/107179-MS](https://doi.org/10.2118/107179-MS).
- [135] Evans, M. and Swartz, T. Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science*, 10(3), 1995. DOI: [10.1214/ss/1177009938](https://doi.org/10.1214/ss/1177009938).
- [136] Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143–10162, 1994. DOI: [10.1029/94JC00572](https://doi.org/10.1029/94JC00572).
- [137] Evensen, G. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003. DOI: [10.1007/s10236-003-0036-9](https://doi.org/10.1007/s10236-003-0036-9).
- [138] Evensen, G. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54(6):539–560, 2004. DOI: [10.1007/s10236-004-0099-2](https://doi.org/10.1007/s10236-004-0099-2).
- [139] Evensen, G. *Data assimilation: the ensemble Kalman filter*. Springer, Berlin, 2007. ISBN 978-3540383000. DOI: [10.1007/978-3-642-03711-5](https://doi.org/10.1007/978-3-642-03711-5).
- [140] Evensen, G. Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciences*, 2018. DOI: [10.1007/s10596-018-9731-y](https://doi.org/10.1007/s10596-018-9731-y).
- [141] Evensen, G. Accounting for model errors in iterative ensemble smoothers. *Computational Geosciences*, 23, 2019. DOI: [10.1007/s10596-019-9819-z](https://doi.org/10.1007/s10596-019-9819-z).
- [142] Evensen, G. On the formulation of the ensemble history-matching problem. In *Proceedings of the SPE Reservoir Simulation Conference, Galveston, Texas, USA, March 2023*, number SPE-212232-MS, 2023. DOI: [10.2118/212232-MS](https://doi.org/10.2118/212232-MS).

- [143] Evensen, G. and Eikrem, K. S. Conditioning reservoir models on rate data using ensemble smoothers. *Computational Geosciences*, 22(1): 1–20, 2018. DOI: [10.1007/s10596-018-9750-8](https://doi.org/10.1007/s10596-018-9750-8).
- [144] Evensen, G., Hove, J., Meisingset, H. C., Reiso, E., Seim, K. S., and Espelid, Ø. Using the EnKF for assisted history matching of a North Sea reservoir model. In *Proceedings of the SPE Reservoir Simulation Symposium, Houston, Texas, 26–28 February*, number SPE-106184-MS, 2007. DOI: [10.2118/106184-MS](https://doi.org/10.2118/106184-MS).
- [145] Evensen, G., Raanes, P. N., Stordal, A. S., and Hove, J. Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. *Frontiers in Applied Mathematics and Statistics*, 3, 2019. DOI: [10.3389/fams.2019.00047](https://doi.org/10.3389/fams.2019.00047).
- [146] Exterkoetter, R., Dutra, G. R., de Figueiredo, L. P., Bordignon, F., Neto, G. M. S., and Emerick, A. A. Feature extraction in time-lapse seismic using deep learning for data assimilation. *SPE Journal*, 1, 2023. DOI: [10.2118/212196-PA](https://doi.org/10.2118/212196-PA).
- [147] Exterkoetter, R., de Figueiredo, L. P., Bordignon, F. L., Emerick, A. A., Roisenberg, M., and Rodrigues, B. B. Ensemble smoother with fully convolutional VAE for seismic facies inversion. *Computers & Geosciences*, 189, 2024. DOI: [10.1016/j.cageo.2024.105619](https://doi.org/10.1016/j.cageo.2024.105619).
- [148] Federico, G. D. and Durlofsky, L. J. Latent diffusion models for parameterization and data assimilation of facies-based geomodels. *arXiv:2406.14815v3 [cs.CV]*, 2024. DOI: [10.48550/arXiv.2406.14815](https://doi.org/10.48550/arXiv.2406.14815).
- [149] Fletcher, R. and Reeves, C. M. Function minimization by conjugate gradient. *Computer Journal*, 7:149–154, 1964. DOI: [10.1093/comjnl/7.2.149](https://doi.org/10.1093/comjnl/7.2.149).
- [150] Floris, F. J. T., Bush, M. D., Cuypers, M., Roggero, F., and Syversveen, A. R. Methods for quantifying the uncertainty of production forecasts: a comparative study. *Petroleum Geoscience*, 7(SUPP): 87–96, 2001. DOI: [10.1144/petgeo.7.S.S87](https://doi.org/10.1144/petgeo.7.S.S87).
- [151] Fonseca, R. M., Chen, B., Jansen, J. D., and Reynolds, A. C. A stochastic simplex approximate gradient (StoSAG) for optimization under uncertainty. *International Journal For Numerical Methods in Engineering*, 109(13):1756–1776, 2017. DOI: [10.1002/nme.5342](https://doi.org/10.1002/nme.5342).
- [152] Fonseca, R. M., Rossa, E. D., Emerick, A. A., Hanea, R. G., and Jansen, J. D. Introduction to the special issue: Overview of OLYMPUS optimization benchmark challenge. *Computational Geosciences*, 24: 1933–1941, 2020. DOI: [10.1007/s10596-020-10003-4](https://doi.org/10.1007/s10596-020-10003-4).
- [153] Franssen, H. J. H., Kaiser, H. P., Kuhlmann, U., Bauser, G., Stauffer, F., Müller, R., and Kinzelbach, W. Operational real-time modeling with ensemble kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. *Water Resources Research*, 47(2), 2011. DOI: [10.1029/2010WR009480](https://doi.org/10.1029/2010WR009480).
- [154] Freshley, M. D. 300 area integrated field-scale subsurface research challenge (IFRC) field site management plan. Technical report, Pacific

- Northwest National Laboratory, 2008. URL <https://www.pnnl.gov/publications>.
- [155] Furre, A.-K., Ola Eiken, H. A., Vevatne, J. N., and Kiær, A. F. 20 years of monitoring CO₂-injection at Sleipner. *Energy Procedia*, 114: 3916–3926, 2017. DOI: [10.1016/j.egypro.2017.03.1523](https://doi.org/10.1016/j.egypro.2017.03.1523).
- [156] Furrer, R. and Bengtsson, T. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, 98(2):227–255, 2007. DOI: [10.1016/j.jmva.2006.08.003](https://doi.org/10.1016/j.jmva.2006.08.003).
- [157] Galli, A., Beucher, H., Le Loc’h, G., Doligez, B., and Group, H. *The pros and cons of the truncated Gaussian method*, pages 217–233. Springer Netherlands, Dordrecht, 1994. DOI: [10.1007/978-94-015-8267-4_18](https://doi.org/10.1007/978-94-015-8267-4_18).
- [158] Gao, G. and Reynolds, A. C. An improved implementation of the LBFGS algorithm for automatic history matching. *SPE Journal*, 11(1):5–17, 2006. DOI: [10.2118/90058-PA](https://doi.org/10.2118/90058-PA).
- [159] Gao, G., Li, G., and Reynolds, A. C. A stochastic algorithm for automatic history matching. *SPE Journal*, 12(2):196–208, 2007. DOI: [10.2118/90065-PA](https://doi.org/10.2118/90065-PA).
- [160] Gao, G., Vink, J. C., Chen, C., Alpak, F. O., and Du, K. Enhanced reparameterization and data-integration algorithms for robust and efficient history matching of geologically complex reservoirs. In *Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, Texas, USA, 28–30 September*, number SPE-175039-MS, 2015. DOI: [10.2118/175039-MS](https://doi.org/10.2118/175039-MS).
- [161] Garud, S. S., Karimi, I. A., and Kraft, M. Design of computer experiments: A review. *Computers and Chemical Engineering*, 106, 2017. DOI: [10.1016/j.compchemeng.2017.05.010](https://doi.org/10.1016/j.compchemeng.2017.05.010).
- [162] Gaspari, G. and Cohn, S. E. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999. DOI: [10.1002/cj.49712555417](https://doi.org/10.1002/cj.49712555417).
- [163] Gassmann, F. Elastic waves through a packing of spheres. *Geophysics*, 16:673–685, 1951. DOI: [10.1190/1.1437718](https://doi.org/10.1190/1.1437718).
- [164] Gavalas, G. R., Shah, P. C., and Seinfeld, J. H. Reservoir history matching by Bayesian estimation. *SPE Journal*, 16(6):337–350, 1976. DOI: [10.2118/5740-PA](https://doi.org/10.2118/5740-PA).
- [165] Gavish, M. and Donoho, D. L. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8): 5040–5053, 2014. DOI: [10.1109/TIT.2014.2323359](https://doi.org/10.1109/TIT.2014.2323359).
- [166] Geyer, C. J. *Handbook of Markov Chain Monte Carlo*, chapter Introduction to Markov chain Monte Carlo. Chapman & Hall/CRC, 2011.
- [167] Gharamti, M. E., Hoteit, I., and Valstar, J. Dual states estimation of a subsurface flow-transport coupled model using ensemble Kalman filtering. *Advances in Water Resources*, 60, 2013. DOI: [10.1016/j.advwatres.2013.07.011](https://doi.org/10.1016/j.advwatres.2013.07.011).

- [168] Gineste, M. and Eidsvik. Batch seismic inversion using the iterative ensemble Kalman smoother. *Computational Geosciences*, 25, 2021. DOI: [10.1007/s10596-021-10043-4](https://doi.org/10.1007/s10596-021-10043-4).
- [169] Gineste, M., Eidsvik, J., and Zheng, Y. Ensemble-based seismic inversion for a stratified medium. *Geophysics*, 85, 2020. DOI: [10.1190/geo2019-0017.1](https://doi.org/10.1190/geo2019-0017.1).
- [170] Golub, G. H. and van Loan, C. F. *Matrix computations*. The Johns Hopkins University Press, Baltimore, 3rd edition, 1989. ISBN 978-0801854149.
- [171] Gómez-Hernández, J. J. and Journel, A. G. Joint sequential simulation of multigaussian fields. In Soares, A., editor, *Geostatistics Tróia '92*, pages 133–144. 1992. DOI: [10.1007/978-94-011-1739-5_8](https://doi.org/10.1007/978-94-011-1739-5_8).
- [172] Gómez-Hernández, J. J. and Xu, T. Contaminant source identification in aquifers: A critical view. *Mathematical Geosciences*, 54, 2021. DOI: [10.1007/s11004-021-09976-4](https://doi.org/10.1007/s11004-021-09976-4).
- [173] González-Nicolás, A., Baú, D., and Alzraiee, A. Detection of potential leakage pathways from geological carbon storage by fluid pressure data assimilation. *Advances in Water Resources*, 86, 2015. DOI: [10.1016/j.advwatres.2015.10.006](https://doi.org/10.1016/j.advwatres.2015.10.006).
- [174] Gosselin, O., Aanonsen, S., Aavatsmark, I., Cominelli, A., Gonard, R., Kolasinski, M., Ferdinandi, F., Kovacic, L., and Neylon, K. History matching using time-lapse seismic (HUTS). In *Proceedings of the SPE Annual Technical Conference and Exhibition, Denver, Colorado, 5–8 October*, number SPE-84464-MS, 2003. DOI: [10.2118/84464-MS](https://doi.org/10.2118/84464-MS).
- [175] Grana, D., Mukerji, T., and Doyen, P. *Seismic Reservoir Modeling: Theory, Examples, and Algorithms*. John Wiley & Sons, 2021. ISBN 978-1119086185.
- [176] Grana, D., Azevedo, L., de Figueiredo, L., Connolly, P., and Mukerji, T. Probabilistic inversion of seismic data for reservoir petrophysical characterization: Review and examples. *Geophysics*, 87(5), 2022. DOI: [10.1190/geo2021-0776.1](https://doi.org/10.1190/geo2021-0776.1).
- [177] Gross, H., Settgast, R. R., Aronson, R. M., Borio, A., Bui, Q. M., Castelletto, N., Corbett, B. C., Cordier, P., Cremon, M. A., Cusini, M., Frambati, S., Franc, J., Gazzola, T., Hamon, F., Huang, J., Kachuma, D., Karimi-Fard, M., Klevtsov, S., Lapene, A., Magri, V. A. P., Mazuyer, A., N'diaye, M., Osei-Kuffuor, D., Sherman, C. S., Tchelepi, H. A., Tobin, W. R., Tomin, P., Wen, X., and White, J. A. GEOS: Exascale, multiphysics, open-source simulation for CCS. In *Proceedings of the European Conference for Mathematics of Geological Reservoirs ECMOR, September 2-5, Oslo, Norway*, 2024. DOI: [10.3997/2214-4609.202437054](https://doi.org/10.3997/2214-4609.202437054).
- [178] Gu, Y. and Oliver, D. S. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal*, 12(4):438–446, 2007. DOI: [10.2118/108438-PA](https://doi.org/10.2118/108438-PA).
- [179] Guardiano, F. B. and Srivastava, R. M. Multivariate geostatistics: Beyond bivariate moments. In Soares, A., editor, *Geostatistics Tróia '92*,

- volume 5 of *Quantitative Geology and Geostatistics*, pages 133–144. Springer Netherlands, 1993.
- [180] Guo, J.-X., Huang, C., Wang, J.-L., and Meng, X.-Y. Integrated operation for the planning of CO₂ capture path in CCS-EOR project. *Journal of Petroleum Science and Engineering*, 186, 2020. DOI: [10.1016/j.petrol.2019.106720](https://doi.org/10.1016/j.petrol.2019.106720).
 - [181] Guo, Z., Chen, C., Gao, G., Cao, R., Li, R., and Liu, C. EUR assessment of unconventional assets using machine learning and distributed computing techniques. In *Proceedings of the SPE/AAPG/SEG Unconventional Resources Technology Conference, Austin, Texas, USA, 24–26 July*, number URTEC-2659996-MS, 2017. DOI: [10.15530/URTEC-2017-2659996](https://doi.org/10.15530/URTEC-2017-2659996).
 - [182] Hadamard, J. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902.
 - [183] Hamill, T. M., Whitaker, J. S., and Snyder, C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11):2776–2790, 2001. DOI: [10.1175/1520-0493\(2001\)129<2776:DDFOBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2).
 - [184] Hanke, M. A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Problems*, 13(1):79–95, 1997. DOI: [10.1088/0266-5611/13/1/007](https://doi.org/10.1088/0266-5611/13/1/007).
 - [185] Hannis, S., Lu, J., Chadwick, A., Hovorka, S., Kirk, K., Romanak, K., and Pearce, J. CO₂ storage in depleted or depleting oil and gas fields: What can we learn from existing projects? *Energy Procedia*, 114, 2017. DOI: [10.1016/j.egypro.2017.03.1707](https://doi.org/10.1016/j.egypro.2017.03.1707).
 - [186] Hansen, N. and Ostermeier, A. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2), 2001. DOI: [10.1162/106365601750190398](https://doi.org/10.1162/106365601750190398).
 - [187] Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. DOI: [10.2307/2334940](https://doi.org/10.2307/2334940).
 - [188] Haugen, V., Nævdal, G., rgen Natvik, L.-J., Evensen, G., Berg, A. M., and Flornes, K. M. History matching using the ensemble Kalman filter on a North Sea field case. *SPE Journal*, 13(4):382–391, 2008. DOI: [10.2118/102430-PA](https://doi.org/10.2118/102430-PA).
 - [189] He, J., Sarma, P., and Durlofsky, L. J. Reduced-order flow modeling and geological parameterization for ensemble-based data assimilation. *Computers & Geosciences*, 55:54–69, 2013. DOI: [10.1016/j.cageo.2012.03.027](https://doi.org/10.1016/j.cageo.2012.03.027).
 - [190] He, Q., Barajas-Solano, D., Tartakovsky, G., and Tartakovsky, A. M. Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Advances in Water Resources*, 141, 2020. DOI: [10.1016/j.advwatres.2020.103610](https://doi.org/10.1016/j.advwatres.2020.103610).
 - [191] Heidarabad, R. G. and Shin, K. Carbon capture and storage in depleted oil and gas reservoirs: the viewpoint of wellbore injectivity. *Energies*, 17(5), 2024. DOI: [10.3390/en17051201](https://doi.org/10.3390/en17051201).

- [192] Hill, A. D. *Production Logging: Theoretical and Interpretive Elements*. Society of Petroleum Engineers, second edition, 2021. ISBN 978-1613999073. DOI: [10.2118/9781613998243](https://doi.org/10.2118/9781613998243).
- [193] Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. Bayesian model averaging: a tutorial. *Statistical Science*, 14(4), 1999. URL <http://www.jstor.org/stable/2676803>.
- [194] Hoffmann, J. GeoStats.jl – high-performance geostatistics in Julia. *The Journal of Open Source Software*, 2018. DOI: [10.21105/joss.00692](https://doi.org/10.21105/joss.00692).
- [195] Holland, J. H. *Adaptation in Natural and Artificial System*. The MIT Press, 1992. ISBN 978-0262581110.
- [196] Horn, R. A. The Hadamard product. In *Matrix Theory and Applications*, pages 87–170. American Mathematical Society, 1990. DOI: [10.1090/psapm/040/1059485](https://doi.org/10.1090/psapm/040/1059485).
- [197] Houtekamer, P. L. and Mitchell, H. L. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3):796–811, 1998. DOI: [10.1175/1520-0493\(1998\)126<0796:DAUAEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2).
- [198] Houtekamer, P. L. and Mitchell, H. L. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137, 2001. DOI: [10.1175/1520-0493\(2001\)129<0123:ASEKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2).
- [199] Houtekamer, P. L. and Mitchell, H. L. Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131:3269–3289, 2005. DOI: [10.1256/qj.05.135](https://doi.org/10.1256/qj.05.135).
- [200] Houtekamer, P. L. and Zhang, F. Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144, 2016. DOI: [10.1175/MWR-D-15-0440.1](https://doi.org/10.1175/MWR-D-15-0440.1).
- [201] Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review*, 133(3):604–620, 2005. DOI: [10.1175/MWR-2864.1](https://doi.org/10.1175/MWR-2864.1).
- [202] Howard, A. D. and Knutson, T. R. Sufficient conditions for river meandering: A simulation approach. *Water Resources Research*, 20(11), 1984. DOI: [10.1029/WR020i011p01659](https://doi.org/10.1029/WR020i011p01659).
- [203] Hu, L. Y. Gradual deformation and iterative calibration of Gaussian-related stochastic models. *Mathematical Geology*, 32(1):87–108, 2000. DOI: [10.1023/A:1007506918588](https://doi.org/10.1023/A:1007506918588).
- [204] Hu, L. Y., Le Ravalec, M., Blanc, G., Roggero, F., Noetinger, B., Haas, A., and Corre, B. Reducing uncertainties in production forecasts by constraining geological modeling to dynamic data. In *Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, Texas, USA, 3–6 October*, number SPE-56703-MS, 1999. DOI: [10.2118/56703-MS](https://doi.org/10.2118/56703-MS).
- [205] Huang, C., Hu, B. X., Li, X., and Ye, M. Using data assimilation method to calibrate a heterogeneous conductivity field and improve solute transport prediction with an unknown contamination source.

- Stochastic Environmental Research and Risk Assessment*, 23, 2009. DOI: [10.1007/s00477-008-0289-4](https://doi.org/10.1007/s00477-008-0289-4).
- [206] Huang, X., Will, R., Khan, M., and Stanley, L. Reservoir characterization by integration of time-lapse seismic and production data. In *Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 5–8 October*, number SPE-38695-MS, 1997.
- [207] Hunt, B. R., Kostelich, E. J., and Szunyogh, I. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, 230:112–26, 2007. DOI: [10.1016/j.physd.2006.11.008](https://doi.org/10.1016/j.physd.2006.11.008).
- [208] Iglesias, M. and Yang, Y. Adaptive regularisation for ensemble Kalman inversion. *Inverse Problems*, 37(2), 2021. DOI: [10.1088/1361-6420/abd29b](https://doi.org/10.1088/1361-6420/abd29b).
- [209] Iglesias, M. A. and Dawson, C. The regularizing Levenberg-Marquardt scheme for history matching of petroleum reservoirs. *Computational Geosciences*, 17(6):1033–1053, 2013. DOI: [10.1007/s10596-013-9373-z](https://doi.org/10.1007/s10596-013-9373-z).
- [210] Iglesias, M. A., Law, K. J. H., and Stuart, A. M. Ensemble Kalman methods for inverse problems. *Inverse Problems*, 29(4), 2013. DOI: [10.1088/0266-5611/29/4/045001](https://doi.org/10.1088/0266-5611/29/4/045001).
- [211] Jaber, A. K., Al-Jawad, S. N., and Alhuraishawy, A. K. A review of proxy modeling applications in numerical reservoir simulation. *Arabian Journal of Geosciences*, 12, 2019. DOI: [0.1007/s12517-019-4891-1](https://doi.org/0.1007/s12517-019-4891-1).
- [212] Jafarpour, B. and Khodabakhshi, M. A probability conditioning method (PCM) for nonlinear flow data integration into multipoint statistical facies simulation. *Mathematical Geosciences*, 43(2):133–164, 2011. DOI: [10.1007/s11004-011-9316-y](https://doi.org/10.1007/s11004-011-9316-y).
- [213] Jafarpour, B. and McLaughlin, D. B. History matching with an ensemble Kalman filter and discrete cosine parameterization. *Computational Geosciences*, 12(2):227–244, 2008. DOI: [10.1007/s10596-008-9080-3](https://doi.org/10.1007/s10596-008-9080-3).
- [214] Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society A*, 186(1007):453–461, 1946. DOI: [10.1098/rspa.1946.0056](https://doi.org/10.1098/rspa.1946.0056).
- [215] Jeong, H., Sun, A. Y., Jeon, J., Min, B., and Jeong, D. Efficient ensemble-based stochastic gradient methods for optimization under geological uncertainty. *Frontiers in Earth Science*, 8, 2020. DOI: [10.3389/feart.2020.00108](https://doi.org/10.3389/feart.2020.00108).
- [216] Jiang, P., Chen, X., Chen, K., Anderson, J., Collins, N., and Gharamti, M. E. DART-PFLOTTRAN: an ensemble-based data assimilation system for estimating subsurface flow and transport model parameters. *Environmental Modelling and Software*, 142, 142. DOI: [10.1016/j.envsoft.2021.105074](https://doi.org/10.1016/j.envsoft.2021.105074).
- [217] Jiang, S. and Durlofsky, L. J. History matching for geological carbon storage using data-space inversion with spatio-temporal data parameterization. *International Journal of Greenhouse Gas Control*, 134, 2024. DOI: [10.1016/j.ijggc.2024.104124](https://doi.org/10.1016/j.ijggc.2024.104124).
- [218] Jiang, S., Hui, M.-H., and Durlofsky, L. J. Data-space inversion with a recurrent autoencoder for naturally fractured sys-

- tems. *Frontiers in Applied Mathematics and Statistics*, 7, 2021. DOI: [10.3389/fams.2021.686754](https://doi.org/10.3389/fams.2021.686754).
- [219] Jin, L., Alpak, F., van den Hoek, P., Pirmez, C., Fehintola, T., Tendo, F., and Olaniyan, E. A comparison of stochastic data-integration algorithms for the joint history matching of production and time-lapse-seismic data. *SPE Reservoir Evaluation & Engineering*, 15(4), 2012. DOI: [10.2118/146418-PA](https://doi.org/10.2118/146418-PA).
- [220] Jin, Q. On a regularized Levenberg-Marquardt method for solving nonlinear inverse problems. *Numerische Mathematik*, 115, 2010. DOI: [10.1007/s00211-009-0275-x](https://doi.org/10.1007/s00211-009-0275-x).
- [221] Johann, P., Sansonowski, R., Oliveira, R., and Bampi, D. 4D seismic in heavy-oil, turbidite reservoir offshore Brazil. *The Leading Edge*, 28(6):718–729, June 2009. DOI: [10.1190/1.3148415](https://doi.org/10.1190/1.3148415).
- [222] Joon, S., Dawuda, I., Morgan, E., and Srinivasan, S. Rock physics-based data assimilation of integrated continuous active-source seismic and pressure monitoring data during geological carbon storage. *SPE Journal*, 27(4), 2022. DOI: [10.2118/209585-PA](https://doi.org/10.2118/209585-PA).
- [223] Joosten, G. J., Altintas, A., Essen, G. V., Doren, J. V., van den Hoek, P. G. P., and Foreste, K. Reservoir model maturation and assisted history matching based on production and 4D seismic data. In *Proceedings of the SPE Annual Technical Conference and Exhibition, Amsterdam, The Netherlands, 27–29 October*, number SPE-170604-MS, 2014. DOI: [10.2118/170604-MS](https://doi.org/10.2118/170604-MS).
- [224] Journel, A. G. and Deutsch, C. V. Entropy and spatial disorder. *Mathematical Geology*, 25(3):329–355, 1993. DOI: [10.1007/BF00901422](https://doi.org/10.1007/BF00901422).
- [225] Kahrobaei, S., van Essen, G. M., Doren, J. F. V., den Hof, P. M. V., and Jansen, J. D. Adjoint-based history matching of structural models using production and time-lapse seismic data. In *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas*, number SPE-163586-MS, 2013. DOI: [10.2118/163586-MS](https://doi.org/10.2118/163586-MS).
- [226] Kaipio, J. P. and Somersalo, E. *Statistical and Computational Inverse Problems*. Springer New York, 2005. DOI: [10.1007/b138659](https://doi.org/10.1007/b138659).
- [227] Kaleta, M. P., Hanea, R. G., Heemink, A. W., and Jansen, J.-D. Model-reduced gradient-based history matching. *Computational Geosciences*, 15, 2011. DOI: [10.1007/s10596-010-9203-5](https://doi.org/10.1007/s10596-010-9203-5).
- [228] Kalman, R. E. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering*, 82: 35–45, 1960. DOI: [10.1115/1.3662552](https://doi.org/10.1115/1.3662552).
- [229] Kalman, R. E. and Bucy, R. S. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(1), 1961. DOI: [10.1115/1.3658902](https://doi.org/10.1115/1.3658902).
- [230] Kam, D., Han, J., and Datta-Gupta, A. Streamline-based rapid history matching of bottomhole pressure and three-phase production data. In *Proceedings of the SPE Improved Oil Recovery Conference, Tulsa, Oklahoma, USA, 11–13 April*, number SPE-179549-MS, 2016. DOI: [10.2118/179549-MS](https://doi.org/10.2118/179549-MS).

- [231] Kam, D., Han, J., and Datta-Gupta, A. Streamline-based history matching of bottomhole pressure and three-phase production data using a multiscale approach. *Journal of Petroleum Science and Engineering*, 154, 2017. DOI: [10.1016/j.petrol.2017.04.022](https://doi.org/10.1016/j.petrol.2017.04.022).
- [232] Kang, X., Kokkinaki, A., Kitanidis, P. K., Shi, X., Lee, J., Mo, S., and Wu, J. Hydrogeophysical characterization of nonstationary DNAPL source zones by integrating a convolutional variational autoencoder and ensemble smoother. *Water Resources Research*, 57(2), 2021. DOI: [10.1029/2020WR028538](https://doi.org/10.1029/2020WR028538).
- [233] Karsli, H. and Dondurur, D. A procedure to reduce side lobes of reflection wavelets: A contribution to low frequency information. *Journal of Applied Geophysics*, 96, 2013. DOI: [10.1016/j.jappgeo.2013.07.002](https://doi.org/10.1016/j.jappgeo.2013.07.002).
- [234] Kennedy, J. and Eberhart, R. Particle swarm optimization. In *Proceedings IEEE international conference on neural networks, IV. IEEE Service Center, Piscataway, NJ*, pages 1942–1948, 1995. DOI: [10.1109/ICNN.1995.488968](https://doi.org/10.1109/ICNN.1995.488968).
- [235] Kennedy, M. C. and O’Hagan, A. Bayesian calibration of computer models. *Journal of The Royal Statistical Society: Series B*, 63:425–464, 2001. DOI: [10.1111/1467-9868.00294](https://doi.org/10.1111/1467-9868.00294).
- [236] Keppenne, C. L. and Rienecker, M. M. Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. *Journal of Marine Systems*, 40–41:363–380, 2003. DOI: [10.1016/S0924-7963\(03\)00025-3](https://doi.org/10.1016/S0924-7963(03)00025-3).
- [237] Khaninezhad, M. M., Jafarpour, B., and Li, L. Sparse geologic dictionaries for subsurface flow model calibration: Part i. inversion formulation. *Advances in Water Resources*, 39:106–121, 2012. DOI: [10.1016/j.advwatres.2011.09.002](https://doi.org/10.1016/j.advwatres.2011.09.002).
- [238] Kirkpatrick, S., Jr., C. D. G., and Vecchi, M. P. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. DOI: [10.1126/science.220.4598.671](https://doi.org/10.1126/science.220.4598.671).
- [239] Kitanidis, P. K. Quasi-linear geostatistical theory for inverting. *Water Resources Research*, 31(10):2411–2419, 1995. DOI: [10.1029/95WR01945](https://doi.org/10.1029/95WR01945).
- [240] Kovachki, N. B. and Stuart, A. M. Ensemble Kalman inversion: a derivative-free technique for machine learning tasks. *Inverse Problems*, 35, 2019. DOI: [10.1088/1361-6420/ab1c3a](https://doi.org/10.1088/1361-6420/ab1c3a).
- [241] Kragh, E. and Christie, P. Seismic repeatability, normalized RMS, and predictability. *The Leading Edge*, 21(7), 2002. DOI: [10.1190/1.1497316](https://doi.org/10.1190/1.1497316).
- [242] Krige, D. G. A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master’s thesis, University of the Witwatersrand, Faculty of Engineering, 1951.
- [243] Krogstad, S., Lie, K.-A., Moyner, O., Nilsen, H. M., Raynaud, X., and Skaffestad, B. MRST-AD – an open-source framework for rapid prototyping and evaluation of reservoir simulation problems. In *Proceedings*

- of the SPE Reservoir Simulation Symposium, Houston, Texas, USA, 2015. DOI: [10.2118/173317-MS](https://doi.org/10.2118/173317-MS).
- [244] Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., Jasinski, M., Mocko, D., Getirana, A., Lannoy, G. D., Cosh, M. H., Hain, C. R., Anderson, M., Arsenault, K. R., Xia, Y., and Ek, M. Assimilation of gridded GRACE terrestrial water storage estimates in the north american land data assimilation system. *Journal of Hydrometeorology*, 17(7), 2016. DOI: [10.1175/JHM-D-15-0157.1](https://doi.org/10.1175/JHM-D-15-0157.1).
 - [245] Lacerda, J. M., Emerick, A. A., and Pires, A. P. Methods to mitigate loss of variance due to sampling errors in ensemble data assimilation with non-local model parameters. *Journal of Petroleum Science and Engineering*, 172:690–706, 2019. DOI: [10.1016/j.petrol.2018.08.056](https://doi.org/10.1016/j.petrol.2018.08.056).
 - [246] Laloy, E., Hérault, R., Jacques, D., and Linde, N. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Advances in Water Resources*, 110:387–405, 2017. DOI: [10.1016/j.advwatres.2017.09.029](https://doi.org/10.1016/j.advwatres.2017.09.029).
 - [247] Laloy, E., Hérault, R., Jacques, D., and Linde, N. Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, 54(1):381–406, 2018. DOI: [10.1002/2017WR022148](https://doi.org/10.1002/2017WR022148).
 - [248] Lam, D.-T., Kerrou, J., Renard, P., Benabderrahmane, H., and Perrochet, P. Conditioning multi-gaussian groundwater flow parameters to transient hydraulic head and flowrate data with iterative ensemble smoothers: A synthetic case study. *Frontiers in Earth Science*, 2020. DOI: [10.3389/feart.2020.00202](https://doi.org/10.3389/feart.2020.00202).
 - [249] Lam, D.-T., Renard, P., Straubhaar, J., and Kerrou, J. Multiresolution approach to condition categorical multiple-point realizations to dynamic data with iterative ensemble smoothing. *Water Resources Research*, 56(2), 2020. DOI: [10.1029/2019WR025875](https://doi.org/10.1029/2019WR025875).
 - [250] LaVenue, A. M. and Pickens, J. F. Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model. *Water Resources Research*, 28(6):1543–1569, 1992. DOI: [10.1029/92WR00208](https://doi.org/10.1029/92WR00208).
 - [251] Lawson, W. G. and Hansen, J. A. Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Monthly Weather Review*, 132(8):1966–1981, 2004. DOI: [10.1175/1520-0493\(2004\)132<1966:IOSADF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1966:IOSADF>2.0.CO;2).
 - [252] Le, D. H., Younis, R., and Reynolds, A. C. A history matching procedure for non-Gaussian facies based on ES-MDA. In *Proceedings of the SPE Reservoir Simulation Symposium, Houston, Texas, USA, 23–25 February*, number SPE-173233-MS, 2015. DOI: [10.2118/173233-MS](https://doi.org/10.2118/173233-MS).
 - [253] Le, D. H., Emerick, A. A., and Reynolds, A. C. An adaptive ensemble smoother with multiple data assimilation for assisted history matching. *SPE Journal*, 21(6):2195–2207, 2016. DOI: [10.2118/173214-PA](https://doi.org/10.2118/173214-PA).
 - [254] Leeuwenburgh, O. and Arts, R. Distance parameterization for efficient seismic history matching with the ensemble Kalman filter. *Computa-*

- tional Geosciences*, 18(3–4):535–548, 2014. DOI: [10.1007/s10596-014-9434-y](https://doi.org/10.1007/s10596-014-9434-y).
- [255] Li, B., Rodell, M., Kumar, S., Beaudoin, H. K., Getirana, A., Zaitchik, B. F., Goncalves, L. G., Cossetin, C., Bhanja, S., Mukherjee, A., Tian, S., Tangdamrongsub, N., Long, D., Nanteza, J., Lee, J., Policelli, F., Goni, I. B., Daira, D., Bila, M., Lannoy, G., Mocko, D., Steele-Dunne, S. C., Save, H., and Bettadpur, S. Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research*, 55(9), 2019. DOI: [10.1029/2018WR024618](https://doi.org/10.1029/2018WR024618).
- [256] Li, G. and Reynolds, A. C. Iterative ensemble Kalman filters for data assimilation. *SPE Journal*, 14(3):496–505, 2009. DOI: [10.2118/109808-PA](https://doi.org/10.2118/109808-PA).
- [257] Li, G. and Reynolds, A. C. Uncertainty quantification of reservoir performance predictions using a stochastic optimization algorithm. *Computational Geosciences*, 15(3):451–462, 2011. DOI: [10.1007/s10596-010-9214-2](https://doi.org/10.1007/s10596-010-9214-2).
- [258] Li, L., Puzel, R., and Davis, A. Data assimilation in groundwater modelling: ensemble Kalman filter versus ensemble smoothers. *Hydrological Processes*, 32(13), 2018. DOI: [10.1002/hyp.13127](https://doi.org/10.1002/hyp.13127).
- [259] Li, R., Reynolds, A. C., and Oliver, D. S. History matching of three-phase flow production data. *SPE Journal*, 8(4):328–340, 2003. DOI: [10.2118/87336-PA](https://doi.org/10.2118/87336-PA).
- [260] Li, X. and Zhang, D. A backward automatic differentiation framework for reservoir simulation. *Computational Geosciences*, 18, 2014. DOI: [10.1007/s10596-014-9441-z](https://doi.org/10.1007/s10596-014-9441-z).
- [261] Lie, K.-A. *An Introduction to Reservoir Simulation Using MATLAB/GNU Octave – User Guide for the MATLAB Reservoir Simulation Toolbox (MRST)*. Cambridge University Press, 2019. DOI: [10.1017/9781108591416](https://doi.org/10.1017/9781108591416).
- [262] Liem, M., Matthai, S. K., and Jenny, P. Estimation of fracture aperture in naturally fractured reservoirs using an ensemble smoother with multiple data assimilation. In *Proceedings of the European Conference for Mathematics of Geological Reservoirs ECMOR, September 5-7, The Hague, The Netherlands*, 2022. DOI: [10.3997/2214-4609.202244068](https://doi.org/10.3997/2214-4609.202244068).
- [263] Lima, M. M., Emerick, A. A., and Ortiz, C. E. P. Data-space inversion with ensemble smoother. *Computational Geosciences*, 24(3): 1179–1200, 2020. DOI: [10.1007/s10596-020-09933-w](https://doi.org/10.1007/s10596-020-09933-w).
- [264] Liu, G., Chen, Y., and Zhang, D. Investigation of flow and transport processes at the MADE site using ensemble Kalman filter. *Advances in Water Resources*, 31(7):975–986, 2008. DOI: [10.1016/j.advwatres.2008.03.006](https://doi.org/10.1016/j.advwatres.2008.03.006).
- [265] Liu, M. and Grana, D. Stochastic nonlinear inversion of seismic data for the estimation of petroelastic properties using the ensemble smoother and data reparameterization. *Geophysics*, 83(3), 2018. DOI: [10.1190/geo2017-0713.1](https://doi.org/10.1190/geo2017-0713.1).

- [266] Liu, N. and Oliver, D. S. Evaluation of Monte Carlo methods for assessing uncertainty. *SPE Journal*, 8(2):188–195, 2003. DOI: [10.2118/84936-PA](#).
- [267] Liu, N. and Oliver, D. S. Critical evaluation of the ensemble Kalman filter on history matching of geologic facies. In *Proceedings of the SPE Reservoir Simulation Symposium*, number SPE-92867-MS, 2005. DOI: [10.2118/92867-MS](#).
- [268] Liu, Y., Sun, W., and Durlafsky, L. J. A deep-learning-based geological parameterization for history matching complex models. *Mathematical Geosciences*, 51:725–766, 2019. DOI: [10.1007/s11004-019-09794-9](#).
- [269] Lomeland, F., Ebeltoft, E., and Thomas, W. H. A new versatile relative permeability correlation. In *International Symposium of the Society of Core Analysts, Toronto, Canada, 21–25 August*, 2005. URL [jgmaas.com/SCA/2005/SCA2005-32.pdf](#).
- [270] Lorenc, A. C. The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3183–3203, 2003. DOI: [10.1256/qj.02.132](#).
- [271] Lorentzen, R. J., Berg, A. M., Nævdal, G., and Vefring, E. H. A new approach for dynamic optimization of waterflooding problems. In *Proceedings of the SPE Intelligent Energy Conference and Exhibition*, number SPE-99690-MS, 2006. DOI: [10.2118/99690-MS](#).
- [272] Lorentzen, R. J., Flornes, K., and Nævdal, G. History channelized reservoirs using the ensemble Kalman filter. *SPE Journal*, 17(1): 137–151, 2012. DOI: [10.2118/143188-PA](#).
- [273] Lorentzen, R. J., Luo, X., Bhakta, T., and Valestrand, R. History matching the full norne field model using seismic and production data. *SPE Journal*, 24(4):1452–1467., 2019. DOI: [10.2118/194205-PA](#).
- [274] Lorentzen, R. J., Bhakta, T., Grana, D., Luo, X., Valestrand, R., and Nævdal, G. Simultaneous assimilation of production and seismic data: application to the norne field. *Computational Geosciences*, 24:907–920, 2020. DOI: [10.1007/s10596-019-09900-0](#).
- [275] Lorentzen, R. J., Bhakta, T., Fossum, K., Haugen, J. A., Lie, E. O., Ndingwan, A. O., and Straith, K. R. Ensemble-based history matching of the Edvard Grieg field using 4D seismic data. *Computational Geosciences*, 28, 2024. DOI: [10.1007/s10596-024-10275-0](#).
- [276] Lorenz, E. N. and Emanuel, K. A. Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, 55(3), 1998. DOI: [10.1175/1520-0469\(1998\)055<0399:OSFSWO>2.0.CO;2](#).
- [277] Lu, M. and Chen, Y. Improved estimation and forecast through model error estimation – norne field example. In *Proceedings of the International Petroleum Technology Conference, Beijing, China, 26–28 March*, number IPTC-19142-MS, 2019. DOI: [10.2523/IPTC-19142-MS](#).

- [278] Luo, X. and Bhakta, T. Automatic and adaptive localization for ensemble-based history matching. *Journal of Petroleum Science and Engineering*, 184, 2020. DOI: [10.1016/j.petrol.2019.106559](https://doi.org/10.1016/j.petrol.2019.106559).
- [279] Luo, X., Stordal, A. S., Lorentzen, R. J., and Nævdal, G. Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: Theory and applications. *SPE Journal*, 20(5), 2015. DOI: [10.2118/176023-PA](https://doi.org/10.2118/176023-PA).
- [280] Luo, X., Bhakta, T., and Nævdal, G. Correlation-based adaptive localization with applications to ensemble-based 4D-seismic history matching. *SPE Journal*, 23(2):396–427, 2018. DOI: [10.2118/185936-PA](https://doi.org/10.2118/185936-PA).
- [281] Luo, X., Lorentzen, R. J., Valestrand, R., and Evensen, G. Correlation-based adaptive localization for ensemble-based history matching: Applied to the Norne field case study. *SPE Reservoir Engineering*, 22(3): 1084–1109, 2018. DOI: [10.2118/191305-PA](https://doi.org/10.2118/191305-PA).
- [282] Luo, X., Lorentzen, R. J., and Bhakta, T. Accounting for model errors of rock physics models in 4D seismic history matching problems: A perspective of machine learning. *Journal of Petroleum Science and Engineering*, 196, 2021. DOI: [10.1016/j.petrol.2020.107961](https://doi.org/10.1016/j.petrol.2020.107961).
- [283] Ma, W., Jafarpour, B., and Qin, J. Dynamic characterization of geologic CO₂ storage aquifers from monitoring data with ensemble Kalman filter. *International Journal of Greenhouse Gas Control*, 81, 2019. DOI: [10.1016/j.ijggc.2018.10.009](https://doi.org/10.1016/j.ijggc.2018.10.009).
- [284] Ma, X. and Bi, L. A robust adaptive iterative ensemble smoother scheme for practical history matching applications. *Computational Geosciences*, First Online, 2019. DOI: [10.1007/s10596-018-9786-9](https://doi.org/10.1007/s10596-018-9786-9).
- [285] Mahalanobis, P. C. On the generalised distance in statistics. In *Proceedings of the National Institute of Sciences of India*, volume 2, pages 49–55, 1936.
- [286] Mahjour, S. K., Santos, A. A. S., and Manuel G. Correia, D. J. S. Developing a workflow to select representative reservoir models combining distance-based clustering and data assimilation for decision making process. *Journal of Petroleum Science and Engineering*, 190, 2020. DOI: [10.1016/j.petrol.2020.107078](https://doi.org/10.1016/j.petrol.2020.107078).
- [287] Mannseth, T. Relation between level set and truncated pluri-Gaussian methodologies for facies representation. *Mathematical Geosciences*, 46 (6):711–731, 2014. DOI: [10.1007/s11004-013-9507-9](https://doi.org/10.1007/s11004-013-9507-9).
- [288] Mannseth, T. Comparison of five different ways to assimilate data for a simplistic weakly nonlinear parameter estimation problem. *Computational Geosciences*, 19(4):791–804, 2015. DOI: [10.1007/s10596-015-9490-y](https://doi.org/10.1007/s10596-015-9490-y).
- [289] Mariethoz, G. *Handbook of Mathematical Geosciences: Fifty Years of IAMG*, chapter When Should We Use Multiple-Point Geostatistics? Springer International Publishing, 2018. DOI: [10.1007/978-3-319-78999-6_31](https://doi.org/10.1007/978-3-319-78999-6_31).

- [290] Mariethoz, G. and Caers, J. *Multiple-point geostatistics – Stochastic modeling with training images*. John Wiley & Sons, Ltd., 2014. ISBN 978-1118662939.
- [291] Marquart, G., Vogt, C., Klein, C., and Widera, A. Estimation of geothermal reservoir properties using the ensemble Kalman filter. *Energy Procedia*, 40, 2013. DOI: [10.1016/j.egypro.2013.08.015](https://doi.org/10.1016/j.egypro.2013.08.015).
- [292] Maschio, C. and Schiozer, D. J. Bayesian history matching using artificial neural network and Markov chain Monte Carlo. *Journal of Petroleum Science and Engineering*, 123, 2014. DOI: [10.1016/j.petrol.2014.05.016](https://doi.org/10.1016/j.petrol.2014.05.016).
- [293] Maschio, C. and Schiozer, D. J. A new methodology for Bayesian history matching using parallel interacting Markov chain Monte Carlo. *Inverse Problems in Science and Engineering*, 26(4), 2018. DOI: [10.1080/17415977.2017.1322078](https://doi.org/10.1080/17415977.2017.1322078).
- [294] Maschio, C., Neto, G. M. S., Davolio, A., and Vinicius S. Rios, D. J. S. Performance assessment of an iterative ensemble smoother with local analysis to assimilate big 4D seismic datasets applied to a complex pre-salt-like benchmark case. *Journal of Geophysics and Engineering*, 21(1), 2024. DOI: [10.1093/jge/gxad099](https://doi.org/10.1093/jge/gxad099).
- [295] Matheron, G. The theory of regionalized variables and its applications. Technical Report No. 5, Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, 1971. URL http://cg.ensmp.fr/bibliotheque/public/MATHERON_Ouvrage_00167.pdf.
- [296] Matheron, G., Beucher, H., de Fouquet, C., Galli, A., Guerillot, D., and Ravenne, C. Conditional simulation of the geometry of fluvio-deltaic reservoirs. In *Proceedings of the SPE Annual Technical Conference and Exhibition*, number SPE-16753-MS, 1987. DOI: [10.2118/16753-MS](https://doi.org/10.2118/16753-MS).
- [297] Maucec, M., Ravanelli, F. M. D. M., Lyngra, S., Zhang, S. J., Alramadhan, A. A., Abdelhamid, O. A., and Al-Garni, S. A. Ensemble-based assisted history matching with rigorous uncertainty quantification applied to a naturally fractured carbonate reservoir. In *Proceedings of the SPE Annual Technical Conference and Exhibition, Dubai, UAE, 26–28 September*, number SPE-181325-MS, 2016. DOI: [10.2118/181325-MS](https://doi.org/10.2118/181325-MS).
- [298] Mavko, G., Mukerji, T., and Dvorkin, J. *The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media*. Cambridge University Press, 2nd edition, 2009. ISBN 978-0511626753. DOI: [10.1017/CBO9780511626753](https://doi.org/10.1017/CBO9780511626753).
- [299] McGrayne, S. B. *The Theory that Would Not Die: How Bayes’ Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press, 2011.
- [300] McHargue, T., Pyrcz, M. J., Sullivan, M. D., Clark, J. D., Fildani, A., Romans, B. W., Covault, J. A., Levy, M., Posamentier, H. W., and Drinkwater, N. J. Architecture of turbidite channel systems on the continental slope: Patterns and predictions. *Marine and Petroleum Geology*, 28:728–243, 2011. DOI: [10.1016/j.marpetgeo.2010.07.008](https://doi.org/10.1016/j.marpetgeo.2010.07.008).

- [301] McKay, M. D., Beckman, R. J., and Conover, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 1979. DOI: [10.2307/1268522](#).
- [302] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953. DOI: [10.1063/1.1699114](#).
- [303] Mezghani, M., Fornel, A., Langlais, V., and Lucet, N. History matching and quantitative use of 4D seismic data for an improved reservoir characterization. In *Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, Texas, 26–29 September*, number SPE-90420-MS, 2004. DOI: [10.2118/90420-MS](#).
- [304] Moreno, D. L. and Aanonsen, S. I. Continuous facies updating using the ensemble Kalman filter and the level set method. *Mathematical Geosciences*, 43(8):951–970, 2011. DOI: [10.1007/s11004-011-9347-4](#).
- [305] Moreno, D. L., Aanonsen, S. I., Evensen, G., and Skjervheim, J.-A. Channel facies estimation based on Gaussian perturbations in the EnKF. In *Proceedings of the 11th European Conference on the Mathematics of Oil Recovery (ECMOR XI)*, 2008. DOI: [10.3997/2214-4609.20146403](#).
- [306] Morozov, V. *Methods for Solving Incorrectly Posed Problems*. Springer-Verlag New York, 1984. DOI: [10.1007/978-1-4612-5280-1](#).
- [307] Morris, M. D. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2), 1991. DOI: [10.2307/1269043](#).
- [308] Mosser, L., Dubrule, O., and Blunt, M. J. DeepFlow: History matching in the space of deep generative models. *arXiv:1905.05749v1 [cs.LG]*, 2019. DOI: [10.48550/arXiv.1905.05749](#).
- [309] Motaei, E. and Ganat, T. Smart proxy models art and future directions in the oil and gas industry:a review. *Geoenergy Science and Engineering*, 227, 2023. DOI: [10.1016/j.geoen.2023.211918](#).
- [310] Moyen, R. and Gentilhomme, T. Adaptive ensemble-based optimisation for petrophysical inversion. *Mathematical Geosciences*, 53, 2021. DOI: [10.1007/s11004-020-09900-2](#).
- [311] Moyner, O. JutulDarcy.jl - a fully differentiable high-performance reservoir simulator based on automatic differentiation. In *Proceedings of the European Conference for Mathematics of Geological Reservoirs ECMOR, September 2-5, Oslo, Norway*, 2024. DOI: [10.3997/2214-4609.202437111](#).
- [312] Myers, M., Stalker, L., Pejicic, B., and Ross, A. Tracers – past, present and future applications in CO₂ geosequestration. *Applied Geochemistry*, 30, 2013. DOI: [10.1016/j.apgeochem.2012.06.001](#).
- [313] Nævdal, G., Mannseth, T., and Vefring, E. H. Near-well reservoir monitoring through ensemble Kalman filter. In *Proceedings of the SPE/DOE Improved Oil Recovery Symposium, 13–17 April*, number SPE-75235-MS, 2002. DOI: [10.2118/75235-MS](#).

- [314] Nan, T. and Wu, J. Groundwater parameter estimation using the ensemble Kalman filter with localization. *Hydrogeology Journal*, 19, 2011. DOI: [10.1007/s10040-010-0679-9](https://doi.org/10.1007/s10040-010-0679-9).
- [315] Neal, R. M. Sampling from multimodal distributions using tempered transitions. *Statistics and Computing*, 6(4):353–366, 1996. DOI: [10.1007/BF00143556](https://doi.org/10.1007/BF00143556).
- [316] Neal, R. M. Annealed importance sampling. *Statistics and Computing*, 11(2), 2001. DOI: [10.1023/A:1008923215028](https://doi.org/10.1023/A:1008923215028).
- [317] Nelder, J. A. and Mead, R. A simplex method for function minimization. *The Computer Journal*, 7(4), 1965. DOI: [10.1093/comjnl/8.1.27](https://doi.org/10.1093/comjnl/8.1.27).
- [318] Nelsen, R. B. *An Introduction to Copulas*. Springer New York, NY, second edition, 2007. DOI: [10.1007/0-387-28678-0](https://doi.org/10.1007/0-387-28678-0).
- [319] Neto, G. M. S., Rios, V. S., Davolio, A., and Schiozer, D. J. Improving fluid modeling representation for seismic data assimilation in compositional reservoir simulation. *Journal of Petroleum Science and Engineering*, 194, 2020. DOI: [10.1016/j.petrol.2020.107446](https://doi.org/10.1016/j.petrol.2020.107446).
- [320] Neto, G. M. S., Davolio, A., and Schiozer, D. J. Assimilating time-lapse seismic data in the presence of significant spatially correlated model errors. *Journal of Petroleum Science and Engineering*, 207, 2021. DOI: [10.1016/j.petrol.2021.109127](https://doi.org/10.1016/j.petrol.2021.109127).
- [321] Neto, G. M. S., Soares, R. V., Davolio, G. E. A., and Schiozer, D. J. Subspace ensemble randomized maximum likelihood with local analysis for time-lapse-seismic-data assimilation. *SPE Journal*, 26(2): 1011–1031, 2021. DOI: [10.2118/205029-PA](https://doi.org/10.2118/205029-PA).
- [322] Nilsen, M. M., Stordal, A. S., Raanes, P. N., Lorentzen, R. J., and Eikrem, K. S. Non-gaussian ensemble optimization. *Mathematical Geosciences*, 2024. DOI: [10.1007/s11004-024-10148-3](https://doi.org/10.1007/s11004-024-10148-3).
- [323] Nóbrega, D. V., de Moraes, F. S., and Emerick, A. A. Data assimilation of a legacy 4D seismic in a brown field. *Journal of Geophysics and Engineering*, 15(6), 2018. DOI: [10.1088/1742-2140/aadd68](https://doi.org/10.1088/1742-2140/aadd68).
- [324] Nocedal, J. Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, 35(151):773–782, 1980. DOI: [10.2307/2006193](https://doi.org/10.2307/2006193).
- [325] Nocedal, J. and Wright, S. J. *Numerical Optimization*. Springer, New York, 2006. ISBN 978-0387303031. DOI: [10.1007/978-0-387-40065-5](https://doi.org/10.1007/978-0-387-40065-5).
- [326] Nunes, J. P. P., Seabra, G. S., and de Sousa Junior, L. C. A review of CO₂-injection projects in the Brazilian pre-salt – storage capacity and geomechanical constraints. *International Journal of Greenhouse Gas Control*, 137, 2024. DOI: [10.1016/j.ijggc.2024.104232](https://doi.org/10.1016/j.ijggc.2024.104232).
- [327] Obidegwu, D., Chassagne, R., and MacBeth, C. Seismic assisted history matching using binary maps. *Journal of Natural Gas Science and Engineering*, 42, 2017. DOI: [10.1016/j.jngse.2017.03.001](https://doi.org/10.1016/j.jngse.2017.03.001).
- [328] Oliver, D. S. Minimization for conditional simulation: Relationship to optimal transport. *Journal of Computational Physics*, 265:1–15, 2014. DOI: [10.1016/j.jcp.2014.01.048](https://doi.org/10.1016/j.jcp.2014.01.048).

- [329] Oliver, D. S. Information content in 4D seismic data: Effect of correlated noise. *Journal of Petroleum Science and Engineering*, 208, 2022. DOI: [10.1016/j.petrol.2021.109728](https://doi.org/10.1016/j.petrol.2021.109728).
- [330] Oliver, D. S. and Alfonzo, M. Calibration of imperfect models to biased observations. *Computational Geosciences*, 22:145–161, 2018. DOI: [10.1007/s10596-017-9678-4](https://doi.org/10.1007/s10596-017-9678-4).
- [331] Oliver, D. S. and Chen, Y. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15(1):185–221, 2011. DOI: [10.1007/s10596-010-9194-2](https://doi.org/10.1007/s10596-010-9194-2).
- [332] Oliver, D. S., He, N., and Reynolds, A. C. Conditioning permeability fields to pressure data. In *Proceedings of the 5th European Conference on the Mathematics of Oil Recovery (ECMOR V)*, 03 September, 1996. DOI: [10.3997/2214-4609.201406884](https://doi.org/10.3997/2214-4609.201406884).
- [333] Oliver, D. S., Reynolds, A. C., Bi, Z., and Abacioglu, Y. Integration of production data into reservoir models. *Petroleum Geoscience*, 7 (SUPP):65–73, 2001. DOI: [10.1144/petgeo.7.S.S65](https://doi.org/10.1144/petgeo.7.S.S65).
- [334] Oliver, D. S., Reynolds, A. C., and Liu, N. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge University Press, Cambridge, UK, 2008. ISBN 978-0511535642. DOI: [10.1017/CBO9780511535642](https://doi.org/10.1017/CBO9780511535642).
- [335] Oliver, D. S., Chen, Y., and Nævdal, G. Updating Markov chain models using the ensemble Kalman filter. *Computational Geosciences*, 15:325–344, 2011. DOI: [10.1007/s10596-010-9220-4](https://doi.org/10.1007/s10596-010-9220-4).
- [336] Oliver, D. S., Fossum, K., Bhakta, T., Sandø, I., Nævdal, G., and Lorentzen, R. J. 4D seismic history matching. *Journal of Petroleum Science and Engineering*, 207, 2021. DOI: [10.1016/j.petrol.2021.109119](https://doi.org/10.1016/j.petrol.2021.109119).
- [337] OpenGoSim. PFLOTRAN-OGS. Software, 2024. URL <https://docs.opengosim.com/manual/introduction/>.
- [338] Otsuka, S., Kotsuki, S., and Miyoshi, T. Nowcasting with data assimilation: A case of global satellite mapping of precipitation. *Weather and Forecasting*, 31:1409–1416, 2016. DOI: [10.1175/WAF-D-16-0039.1](https://doi.org/10.1175/WAF-D-16-0039.1).
- [339] Oudshoorn, C., Werthmüller, D., Slob, E., and Voskov, D. Numerical experiment on data assimilation for geothermal doublets using production data and electromagnetic observations. *Geophysics*, 0(11), 2024. DOI: [10.1190/geo2023-0463.1](https://doi.org/10.1190/geo2023-0463.1).
- [340] Owen, A. B. Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, 2(2), 1992. URL <https://www.jstor.org/stable/24304869>.
- [341] Paola, C. Quantitative models of sedimentary basin filling. *Sedimentology*, 47, 2000. DOI: [10.1046/j.1365-3091.2000.00006.x](https://doi.org/10.1046/j.1365-3091.2000.00006.x).
- [342] Patidar, A. K., Joshi, D., Dristant, U., and Choudhury, T. A review of tracer testing techniques in porous media specially attributed to the oil and gas industry. *Journal of Petroleum Exploration and Production Technology*, 12, 2022. DOI: [10.1007/s13202-022-01526-w](https://doi.org/10.1007/s13202-022-01526-w).

- [343] Pebesma, E. J. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30(7), 2004. DOI: [10.1016/j.cageo.2004.03.012](https://doi.org/10.1016/j.cageo.2004.03.012).
- [344] Perrone, A., Pennadoro, F., Tiani, A., Rossa, E. D., and Sætrom, J. Enhancing the geological models consistency in ensemble based history matching an integrated approach. In *Proceedings of the SPE Reservoir Characterisation and Simulation Conference and Exhibition, Abu Dhabi, UAE , 8–10 May*, number SPE-186049-MS, 2017. DOI: [10.2118/186049-MS](https://doi.org/10.2118/186049-MS).
- [345] Peters, E., Chen, Y., Leeuwenburgh, O., and Oliver, D. S. Extended Brugge benchmark case for history matching and water flooding optimization. *Computers & Geosciences*, 50:16–24, 2013. DOI: [10.1016/j.cageo.2012.07.018](https://doi.org/10.1016/j.cageo.2012.07.018).
- [346] Peters, L., Arts, R., Brouwer, G., Geel, C., Cullick, S., Lorentzen, R. J., Chen, Y., Dunlop, N., Vossepoel, F. C., Xu, R., Sarma, P., Alhuthali, A. H., and Reynolds, A. Results of the Brugge benchmark study for flooding optimisation and history matching. *SPE Reservoir Evaluation & Engineering*, 13(3):391–405, 2010. DOI: [10.2118/119094-PA](https://doi.org/10.2118/119094-PA).
- [347] Ping, J. and Zhang, D. History matching of channelized reservoirs with vector-based level-set parameterization. *SPE Journal*, 19(3):514–529, 2014. DOI: [10.2118/169898-PA](https://doi.org/10.2118/169898-PA).
- [348] Powell, M. J. D. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2), 1964. DOI: [10.1093/comjnl/7.2.155](https://doi.org/10.1093/comjnl/7.2.155).
- [349] Powell, M. J. D. The NEWUOA software for unconstrained optimization without derivatives. In Pillo, G. and Roma, M., editors, *Large-Scale Nonlinear Optimization. Nonconvex Optimization and Its Applications*, volume 83. Springer, 2006. DOI: [10.1007/0-387-30065-1_16](https://doi.org/10.1007/0-387-30065-1_16).
- [350] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, England, 3rd edition, 2007. ISBN 978-0521880688.
- [351] Ptak, T., Piepenbrink, M., and Martac, E. Tracer tests for the investigation of heterogeneous porous media and stochastic modelling of flow and transport-a review of some recent developments. *Journal of Hydrology*, 294, 2003. DOI: [10.1016/j.jhydrol.2004.01.020](https://doi.org/10.1016/j.jhydrol.2004.01.020).
- [352] Pyrcz, M. J., Sech, R. P., Covault, J. A., Willis, B. J., Sylvester, Z., and Sun, T. Stratigraphic rule-based reservoir modeling. *Bulletin of Canadian Petroleum Geology*, 63(4), 2015. DOI: [10.2113/gscpg-bull.63.4.287](https://doi.org/10.2113/gscpg-bull.63.4.287).
- [353] Raanes, P. N., Bocquet, M., and Carrassi, A. Adaptive covariance inflation in the ensemble Kalman filter by gaussian scale mixtures. *Quarterly Journal of the Royal Meteorological Society*, 145(718):53–75, 2019. DOI: [10.1002/qj.3386](https://doi.org/10.1002/qj.3386).

- [354] Raanes, P. N., Stordal, A. S., and Evensen, G. Revising the stochastic iterative ensemble smoother. *Nonlinear Processes in Geophysics*, 26(3), 2019. DOI: [10.5194/npg-2019-10](https://doi.org/10.5194/npg-2019-10).
- [355] Raanes, P. N., Stordal, A. S., and Lorentzen, R. J. Review of ensemble gradients for robust optimisation. *arXiv:2304.12136v1 [math.OC]*, 2023. DOI: [10.48550/arXiv.2304.12136](https://doi.org/10.48550/arXiv.2304.12136).
- [356] Rafiee, J. and Reynolds, A. C. Theoretical and efficient practical procedures for the generation of inflation factors for ES-MDA. *Inverse Problems*, 33(11):115003, 2017. DOI: [10.1088/1361-6420/aa8cb2](https://doi.org/10.1088/1361-6420/aa8cb2).
- [357] RamaRao, B. S., LaVenue, A. M., de Marsily, G., and Marietta, M. G. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. theory and computational experiments. *Water Resources Research*, 31(3):475–493, 1995. DOI: [10.1029/94WR02258](https://doi.org/10.1029/94WR02258).
- [358] Ranazzi, P. H., Luo, X., and Sampaio, M. A. Improving pseudo-optimal Kalman-gain localization using the random shuffle method. *Journal of Petroleum Science and Engineering*, 215, 2022. DOI: [10.1016/j.petrol.2022.110589](https://doi.org/10.1016/j.petrol.2022.110589).
- [359] Rasmussen, A. F., Sandve, T. H., Bao, K., Lauser, A., Hove, J., Skaflestad, B., Klöforn, R., Blatt, M., Rustad, A. B., vareid, O. S., Lie, K.-A., and Thune, A. The open porous media flow reservoir simulator. *Computers and Mathematics with Applications*, 81, 2020. DOI: [10.1016/j.camwa.2020.05.014](https://doi.org/10.1016/j.camwa.2020.05.014).
- [360] Reichle, R. H., McLaughlin, D. B., and Entekhabi, D. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1):103–114, 2002. DOI: [10.1175/1520-0493\(2002\)130<0103:HDAWTE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2).
- [361] Reichle, R. H., Walker, J. P., Koster, R. D., and Houser, P. R. Extended versus ensemble Kalman filtering for land data assimilation. *Journal of Hydrometeorology*, 3(6), 2002. DOI: [10.1175/1525-7541\(2002\)003<0728:EVEKFF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2).
- [362] Remy, N., Boucher, A., and Wu, J. *Applied Geostatistics with SGeMS – A User’s Guide*. 2009.
- [363] Reynolds, A. C., He, N., Chu, L., and Oliver, D. S. Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data. *SPE Journal*, 1(4):413–426, 1996. DOI: [10.2118/30588-PA](https://doi.org/10.2118/30588-PA).
- [364] Reynolds, A. C., Zafari, M., and Li, G. Iterative forms of the ensemble Kalman filter. In *Proceedings of 10th European Conference on the Mathematics of Oil Recovery, Amsterdam, 4–7 September*, 2006. DOI: [10.3997/2214-4609.201402496](https://doi.org/10.3997/2214-4609.201402496).
- [365] Ringrose, P. and Bentley, M. *Reservoir Model Design*. Springer Netherlands, 2015. DOI: [10.1007/978-94-007-5497-3](https://doi.org/10.1007/978-94-007-5497-3).
- [366] Ringrose, P. S. and Meckel, T. A. Maturing global CO₂ storage resources on offshore continental margins to achieve 2DS emissions re-

- ductions. *Scientific Reports*, 9(17944), 2019. DOI: [10.1038/s41598-019-54363-z](https://doi.org/10.1038/s41598-019-54363-z).
- [367] Rodrigues, J. R. P. Calculating derivatives for automatic history matching. *Computational Geosciences*, 10(1):119–136, 2006. DOI: [10.1007/s10596-005-9013-3](https://doi.org/10.1007/s10596-005-9013-3).
 - [368] Roggero, F. and Hu, L. Y. Gradual deformation of continuous geostatistical models for history matching. In *Proceedings of the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 27–30 September*, number SPE-49004-MS, 1998. DOI: [10.2118/49004-MS](https://doi.org/10.2118/49004-MS).
 - [369] Rommelse, J. *Data Assimilation in Reservoir Management*. Ph.D. thesis, Technical University of Delft, Delft, The Netherlands, 2009.
 - [370] Rosa, D. R., Schiozer, D. J., and Davolio, A. Data assimilation of production and multiple 4D seismic acquisitions in a deepwater field using ensemble smoother with multiple data assimilation. *SPE Reservoir Evaluation & Engineering*, 26(4), 2023. DOI: [10.2118/215812-PA](https://doi.org/10.2118/215812-PA).
 - [371] Sætrom, J., Morell, E., Ravari, R. R., Maitre, C. L., and Seldal, M. Fast integrated reservoir modelling on the Gjøa field offshore Norway. In *Proceedings of the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 13–16 November*, number SPE-188557-MS, 2017. DOI: [10.2118/188557-MS](https://doi.org/10.2118/188557-MS).
 - [372] Sakov, P. and Bertino, L. Relation between two common localisation methods for the EnKF. *Computational Geosciences*, 15(2):225–237, 2011. DOI: [10.1007/s10596-010-9202-6](https://doi.org/10.1007/s10596-010-9202-6).
 - [373] Sakov, P. and Oke, P. R. Implications of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, 136:1042–1053, 2008. DOI: [10.1175/2007MWR2021.1](https://doi.org/10.1175/2007MWR2021.1).
 - [374] Sakov, P. and Oke, P. R. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A*, 60(2):361–371, 2008. DOI: [10.1111/j.1600-0870.2007.00299.x](https://doi.org/10.1111/j.1600-0870.2007.00299.x).
 - [375] Sakov, P., Counillon, F., Bertino, L., Lisæter, K. A., Oke, P. R., and Korabely, A. TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, 8(4), 2012. DOI: [10.5194/os-8-633-2012](https://doi.org/10.5194/os-8-633-2012).
 - [376] Sakov, P., Oliver, D. S., and Bertino, L. An iterative EnKF for strongly nonlinear systems. *Monthly Weather Review*, 140:1988–2004, 2012. DOI: [10.1175/MWR-D-11-00176.1](https://doi.org/10.1175/MWR-D-11-00176.1).
 - [377] Sarma, P. and Chen, W. H. Generalization of the ensemble Kalman filter using kernel for non Gaussian random fields. In *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 2–4 February*, number SPE-119177-MS, 2009. DOI: [10.2118/119177-MS](https://doi.org/10.2118/119177-MS).
 - [378] Sarma, P., Durlofsky, L. J., and Aziz, K. Kernel principal component analysis for efficient differentiable parameterization of multipoint geostatistics. *Mathematical Geosciences*, 40(1):3–32, 2008. DOI: [10.1007/s11004-007-9131-7](https://doi.org/10.1007/s11004-007-9131-7).
 - [379] Satija, A. and Caers, J. Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical func-

- tional principal component space. *Advances in Water Resources*, 77: 69–81, 2015. DOI: [10.1016/j.advwatres.2015.01.002](https://doi.org/10.1016/j.advwatres.2015.01.002).
- [380] Satija, A. and Caers, J. Direct forecasting of reservoir performance using production data without history matching. *Computational Geosciences*, 21(2):315–333, 2017. DOI: [10.1007/s10596-017-9614-7](https://doi.org/10.1007/s10596-017-9614-7).
- [381] Scheidt, C., Renard, P., and Caers, J. Prediction-focused subsurface modeling: Investigating the need for accuracy in flow-based inverse modeling. *Mathematical Geosciences*, 47(2):173–191, 2015. DOI: [10.1007/s11004-014-9521-6](https://doi.org/10.1007/s11004-014-9521-6).
- [382] Schneider, T., Stuart, A. M., and Wu, J.-L. Learning stochastic closures using ensemble Kalman inversion. *Transactions of Mathematics and Its Applications*, 5(1), 2021. DOI: [10.1093/imatrm/tnab003](https://doi.org/10.1093/imatrm/tnab003).
- [383] Seabra, G. S. and Vossepoel, F. C. resmda. Software. URL <https://tuda-geo.github.io/resmda/>.
- [384] Seabra, G. S., de Hoop, S., Voskov, D., and Vossepoel, F. C. Understanding of naturally fractured geothermal reservoirs using data assimilation. In *Proceedings of the 48th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California*, 2023.
- [385] Seabra, G. S., Mücke, N. T., Silva, V. L. S., Voskov, D., and Vossepoel, F. C. AI enhanced data assimilation and uncertainty quantification applied to geological carbon storage. *International Journal of Greenhouse Gas Control*, 136, 2024. DOI: [10.1016/j.ijggc.2024.104190](https://doi.org/10.1016/j.ijggc.2024.104190).
- [386] Sebacher, B., Hanea, R., and Stordal, A. S. An adaptive pluri-gaussian simulation model for geological uncertainty quantification. *Journal of Petroleum Science and Engineering*, 158, 2017. DOI: [10.1016/j.petrol.2017.08.038](https://doi.org/10.1016/j.petrol.2017.08.038).
- [387] Sebacher, B. M., Hanea, R., and Heemink, A. A probabilistic parametrization for geological uncertainty estimation using the ensemble Kalman filter (EnKF). *Computational Geosciences*, 17(5):813–832, 2013. DOI: [10.1007/s10596-013-9357-z](https://doi.org/10.1007/s10596-013-9357-z).
- [388] Sebacher, B. M., Stordal, A. S., and Hanea, R. Bridging multipoint statistics and truncated Gaussian fields for improved estimation of channelized reservoirs with ensemble methods. *Computational Geosciences*, 19(2):341–369, 2015. DOI: [10.1007/s10596-014-9466-3](https://doi.org/10.1007/s10596-014-9466-3).
- [389] Seiler, A., Aanonsen, S. I., Evensen, G., and Reivenæs, J. C. Structural uncertainty modeling and updating using the ensemble Kalman filter. *SPE Journal*, 15(4):1062–1076, 2010. DOI: [10.2118/125352-PA](https://doi.org/10.2118/125352-PA).
- [390] Shah, P. C., Gavalas, G. R., and Seinfeld, J. H. Error analysis in history matching: The optimum level of parameterization. *SPE Journal*, 18(6):219–228, 1978. DOI: [10.2118/6508-PA](https://doi.org/10.2118/6508-PA).
- [391] Shahkarami, A., Mohaghegh, S. D., Gholami, V., and Haghghat, S. A. Artificial intelligence (AI) assisted history matching. In *Proceedings of the SPE Western North American and Rocky Mountain Joint Meeting, Denver, Colorado*, number SPE-169507-MS, 2014. DOI: [10.2118/169507-MS](https://doi.org/10.2118/169507-MS).

- [392] Shariatnik, B., Gloaguen, E., Raymond, J., Boutin, L.-C., and Fabien-Ouellet, G. ERT data assimilation to characterize aquifer hydraulic conductivity heterogeneity through a heat-tracing experiment. *Near Surface Geophysics*, 22(3), 2023. DOI: [10.1002/nsg.12288](https://doi.org/10.1002/nsg.12288).
- [393] Shirangi, M. G. History matching production data and uncertainty assessment with an efficient TSVD parameterization algorithm. *Journal of Petroleum Science and Engineering*, 113:54–71, 2013. DOI: [10.1016/j.petrol.2013.11.025](https://doi.org/10.1016/j.petrol.2013.11.025).
- [394] Shirangi, M. G. and Durlofsky, L. J. A general method to select representative models for decision making and optimization under uncertainty. *Computers & Geosciences*, 96:109–123, 2016. DOI: [10.1016/j.cageo.2016.08.002](https://doi.org/10.1016/j.cageo.2016.08.002).
- [395] Shirangi, M. G. and Emerick, A. A. An improved TSVD-based Levenberg-Marquardt algorithm for history matching and comparison with Gauss-Newton. *Journal of Petroleum Science and Engineering*, 143:259–271, 2016. DOI: [10.1016/j.petrol.2016.02.026](https://doi.org/10.1016/j.petrol.2016.02.026).
- [396] Silva, T. M., Pesco, S., and Jr., A. B. Influences of the inflation factors generation in the main parameters of the ensemble smoother with multiple data assimilation. *Journal of Petroleum Science and Engineering*, 203, 2021. DOI: [10.1016/j.petrol.2021.108648](https://doi.org/10.1016/j.petrol.2021.108648).
- [397] Silva, V. L. S., Cardoso, M. A., Oliveira, D. F. B., and de Moraes, R. J. Stochastic optimization strategies applied to the OLYMPUS benchmark. *Computational Geosciences*, 24, 2020. DOI: [10.1007/s10596-019-09854-3](https://doi.org/10.1007/s10596-019-09854-3).
- [398] Skjaeveland, S. M., Siqveland, L. M., Kjosavik, A., Thomas, W. L. H., and Virnovsky, G. A. Capillary pressure correlation for mixed-wet reservoirs. *SPE Reservoir Evaluation & Engineering*, 23(1), 2000. DOI: [10.2118/60900-PA](https://doi.org/10.2118/60900-PA).
- [399] Skjervheim, J.-A., Evensen, G., Aanonsen, S. I., Ruud, B. O., and Johansen, T.-A. Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter. *SPE Journal*, 12(3):282–292, 2007. DOI: [10.2118/95789-PA](https://doi.org/10.2118/95789-PA).
- [400] Skjervheim, J.-A., Evensen, G., Hove, J., and Vabø, J. G. An ensemble smoother for assisted history matching. In *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA, 21–23 February*, number SPE-141929-MS, 2011. DOI: [10.2118/141929-MS](https://doi.org/10.2118/141929-MS).
- [401] Skjervheim, J.-A., Heanea, R. G., and Evensen, G. Fast model update coupled to an ensemble based closed loop reservoir management. In *Proceedings of the Petroleum Geostatistics, Biarritz, France, 7–11 September*, 2015. DOI: [10.3997/2214-4609.201413629](https://doi.org/10.3997/2214-4609.201413629).
- [402] Slowik, A. and Kwasnicka, H. Evolutionary algorithms and their applications to engineering problems. *Neural Computing and Applications*, 32, 2020. DOI: [10.1007/s00521-020-04832-8](https://doi.org/10.1007/s00521-020-04832-8).
- [403] Smola, A. J. and Schölkopf, B. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, 2004. DOI: [10.1023/B:STCO.0000035301.49549.88](https://doi.org/10.1023/B:STCO.0000035301.49549.88).

- [404] Smolen, J. J. and Litsey, L. R. Formation evaluation using wireline formation tester pressure data. *Journal of Petroleum Technology*, 31, 1979. DOI: [10.2118/6822-PA](https://doi.org/10.2118/6822-PA).
- [405] Soares, R. V., Luo, X., Evensen, G., and Bhakta, T. Handling big models and big data sets in history-matching problems through an adaptive local analysis scheme. *SPE Journal*, 26(2), 2021. DOI: [10.2118/204221-PA](https://doi.org/10.2118/204221-PA).
- [406] Sobol', I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3), 2001. DOI: [10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6).
- [407] Soper, H. E. On the probable error of the correlation coefficient to a second approximation. *Biometrika*, 9(1/2), 1913. DOI: [10.2307/2331802](https://doi.org/10.2307/2331802).
- [408] Sousa, E. P. S. and Reynolds, A. C. Adaptive least squares support vector regression for history matching and Markov chain Monte Carlo uncertainty quantification. TUPREP research report, The University of Tulsa, 2018.
- [409] Spall, J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992. DOI: [10.1109/9.119632](https://doi.org/10.1109/9.119632).
- [410] Spall, J. C. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transactions on Aerospace and Electronic Systems*, 34(3):817–823, 1998. DOI: [10.1109/7.705889](https://doi.org/10.1109/7.705889).
- [411] Spremić, M., Eidsvik, J., and Avseth, P. Bayesian rock-physics inversion using a localized ensemble-based approach – with an application to the alvheim field. *Geophysics*, 89(2), 2024. DOI: [10.1190/geo2022-0764.1](https://doi.org/10.1190/geo2022-0764.1).
- [412] Srivastava, R. M. *Handbook of Mathematical Geosciences: Fifty Years of IAMG*, chapter The Origins of the Multiple-Point Statistics (MPS) Algorithm. Springer International Publishing, 2018. DOI: [10.1007/978-3-319-78999-6_32](https://doi.org/10.1007/978-3-319-78999-6_32).
- [413] Stordal, A. S. Iterative Bayesian inversion with Gaussian mixtures: finite sample implementation and large sample asymptotics. *Computational Geosciences*, 19(1):1–15, 2014. DOI: [10.1007/s10596-014-9444-9](https://doi.org/10.1007/s10596-014-9444-9).
- [414] Stordal, A. S. and Elsheikh, A. H. Iterative ensemble smoothers in the annealed importance sampling framework. *Advances in Water Resources*, 86:231–239, 2015. DOI: [10.1016/j.advwatres.2015.09.030](https://doi.org/10.1016/j.advwatres.2015.09.030).
- [415] Stordal, A. S. and Lorentzen, R. J. An iterative version of the adaptive Gaussian mixture filter. *Computational Geosciences*, 18(3–4):579–595, 2014. DOI: [10.1007/s10596-014-9402-6](https://doi.org/10.1007/s10596-014-9402-6).
- [416] Stordal, A. S., Karlses, H. A., Nævedal, G., Skaug, H. J., and Vallés, B. Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Computational Geosciences*, 15(2):293–305, 2011. DOI: [10.1007/s10596-010-9207-1](https://doi.org/10.1007/s10596-010-9207-1).

- [417] Strebelle, S. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–21, 2002. DOI: [10.1023/A:1014009426274](https://doi.org/10.1023/A:1014009426274).
- [418] Sun, L., Seidou, O., Nistor, I., and Liu, K. Review of the Kalman-type hydrological data assimilation. *Hydrological Sciences Journal*, 61(13), 2016. DOI: [10.1080/02626667.2015.1127376](https://doi.org/10.1080/02626667.2015.1127376).
- [419] Sun, W. and Durlofsky, L. J. A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. *Mathematical Geosciences*, 49:679–715, 2017. DOI: [10.1007/s11004-016-9672-8](https://doi.org/10.1007/s11004-016-9672-8).
- [420] Sun, W. and Durlofsky, L. J. Data-space approaches for uncertainty quantification of CO₂ plume location in geological carbon storage. *Advances in Water Resources*, 123, 2019. DOI: [10.1016/j.advwatres.2018.10.028](https://doi.org/10.1016/j.advwatres.2018.10.028).
- [421] Sun, W., Hui, M.-H., and Durlofsky, L. J. Production forecasting and uncertainty quantification for naturally fractured reservoirs using a new data-space inversion procedure. *Computational Geosciences*, Online, 2017. DOI: [10.1007/s10596-017-9633-4](https://doi.org/10.1007/s10596-017-9633-4).
- [422] Sun, W., Vink, J. C., and Gao, G. A practical method to mitigate spurious uncertainty reduction in history matching workflows with imperfect reservoir models. In *Proceedings of the SPE Reservoir Simulation Conference, Montgomery, Texas, USA, 20-22 February*, number SPE-182599-MS, 2017. DOI: [10.2118/182599-MS](https://doi.org/10.2118/182599-MS).
- [423] Suykens, J. A. K. and Vandewalle, J. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999. DOI: [10.1023/A:1018628609742](https://doi.org/10.1023/A:1018628609742).
- [424] Sylvester, Z., Pirmez, C., and Cantelli, A. A model of submarine channel-levee evolution based on channel trajectories: Implications for stratigraphic architecture. *Marine and Petroleum Geology*, 28(3), 2011. DOI: [10.1016/j.marpetgeo.2010.05.012](https://doi.org/10.1016/j.marpetgeo.2010.05.012).
- [425] Szunyogh, I., Kostelich, E. J., Gyarmati, G., Patil, D. J., Hunt, B. R., Kalnay, E., Ott, E., and Yorke, J. A. Assessing a local ensemble Kalman filter: perfect model experiments with the national centers for environmental prediction global model. *Tellus A*, 57:528–545, 2005. DOI: [10.1111/j.1600-0870.2005.00136.x](https://doi.org/10.1111/j.1600-0870.2005.00136.x).
- [426] Tang, H., Fu, P., Sherman, C. S., Zhang, J., Ju, X., ois Hamon, F., Azzolina, N. A., Burton-Kelly, M., and Morris, J. P. A deep learning-accelerated data assimilation and forecasting workflow for commercial-scale geologic carbon storage. *International Journal of Greenhouse Gas Control*, 112, 2021. DOI: [10.1016/j.ijggc.2021.103488](https://doi.org/10.1016/j.ijggc.2021.103488).
- [427] Tang, M., Liu, Y., and Durlofsky, L. J. A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. *Journal of Computational Physics*, 413, 2020. DOI: [10.1016/j.jcp.2020.109456](https://doi.org/10.1016/j.jcp.2020.109456).

- [428] Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia, USA, 2005. ISBN 978-0898715729.
- [429] Tavakoli, R. and Reynolds, A. C. History matching with parameterization based on the SVD of a dimensionless sensitivity matrix. *SPE Journal*, 15(12):495–508, 2010. DOI: [10.2118/118952-PA](#).
- [430] Tavakoli, R. and Reynolds, A. C. Monte Carlo simulation of permeability fields and reservoir performance predictions with SVD parameterization in RML compared with EnKF. *Computational Geosciences*, 15(1):99–116, 2011. DOI: [10.1007/s10596-010-9200-8](#).
- [431] Tavakoli, R., Yoon, H., Delshad, M., ElSheikh, A. H., Wheeler, M. F., and Arnold, B. W. Comparison of ensemble filtering algorithms and null-space monte carlo for parameter estimation and uncertainty quantification using CO₂ sequestration data. *Water Resources Research*, 49(12), 2013. DOI: [10.1002/2013WR013959](#).
- [432] Tavakoli, R., Srinivasan, S., and Wheeler, M. F. Rapid updating of stochastic models by use of an ensemble-filter approach. *SPE Journal*, 19(3):500–513, 2014. DOI: [10.2118/163673-PA](#).
- [433] Tavassoli, Z., Carter, J. N., and King, P. R. An analysis of history matching errors. *Computational Geosciences*, 9, 2005. DOI: [10.1007/s10596-005-9001-7](#).
- [434] Tetteh, M., Li, L., and Davis, A. Leveraging deep learning with progressive growing GAN and ensemble smoother with multiple data assimilation for inverse modeling. *Advances in Water Resources*, 187, 2024. DOI: [10.1016/j.advwatres.2024.104680](#).
- [435] Thulin, K., Li, G., Aanonsen, S. I., and Reynolds, A. C. Estimation of initial fluid contacts by assimilation of production data with EnKF. In *Proceedings of the SPE Annual Technical Conference and Exhibition, Anaheim, California, 11–14 November*, number SPE-109975-MS, 2007. DOI: [10.2118/109975-MS](#).
- [436] Thurin, J., Brossier, R., and Métivier, L. Ensemble-based uncertainty estimation in full waveform inversion. *Geophysical Journal International*, 219, 2019. DOI: [10.1093/gji/ggz384](#).
- [437] Tian, X., Volkov, O., and Voskov, D. An advanced inverse modeling framework for efficient and flexible adjoint-based history matching of geothermal fields. *Geothermics*, 116, 2024. DOI: [10.1016/j.geothermics.2023.102849](#).
- [438] Tillier, E., Veiga, S. D., and Derfoul, R. Appropriate formulation of the objective function for the history matching of seismic attributes. *Computers & Geosciences*, 51, 2013. DOI: [10.1016/j.cageo.2012.07.031](#).
- [439] Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S. Ensemble square-root filters. *Monthly Weather Review*, 131:1485–1490, 2003. DOI: [10.1175/1520-0493\(2003\)131<1485:ESRF>2.0.CO;2](#).
- [440] Todaro, V., D’Oria, M., Tanda, M. G., and Gómez-Hernández, J. J. genES-MDA: a generic open-source software package to solve inverse

- problems via the ensemble smoother with multiple data assimilation. *Computers and Geosciences*, 167, 2022. DOI: [10.1016/j.cageo.2022.105210](https://doi.org/10.1016/j.cageo.2022.105210).
- [441] Tolstukhin, E., Lyngnes, B., and Sudan, H. H. Ekofisk 4D seismic – seismic history matching workflow. In *Proceedings of the SPE Europec/EAGE Annual Conference, 4-7 June, Copenhagen, Denmark*, number SPE-154347-MS, 2012. DOI: [10.2118/154347-MS](https://doi.org/10.2118/154347-MS).
- [442] Torczon, V. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7(1), 1997. DOI: [10.1137/S1052623493250780](https://doi.org/10.1137/S1052623493250780).
- [443] Trani, M., Arts, R., and Leeuwenburgh, O. Seismic history matching of fluid fronts using the ensemble Kalman filter. *SPE Journal*, 18(1), 2012. DOI: [10.2118/163043-PA](https://doi.org/10.2118/163043-PA).
- [444] Tso, C.-H. M., Johnson, T. C., Song, X., Chen, X., Kuras, O., Wilkinson, P., Uhlemann, S., Chambers, J., and Binley, A. Integrated hydrogeophysical modelling and data assimilation for geoelectrical leak detection. *Journal of Contaminant Hydrology*, 234, 2020. DOI: [10.1016/j.jconhyd.2020.103679](https://doi.org/10.1016/j.jconhyd.2020.103679).
- [445] Tversky, A. and Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974. DOI: [10.1126/science.185.4157.1124](https://doi.org/10.1126/science.185.4157.1124).
- [446] van Essen, G. M., Jimenez, E. A., Przybysz-Jarnut, J. P., Horesh, L., Douma, S. G., van den Hoek, P. J., Conn, A., and Mello, U. T. Adjoint-based history-matching of production and time-lapse seismic data. In *Proceedings of the SPE Europec/EAGE Annual Conference, Copenhagen, Denmark*, number SPE-154375-MS, 2012. DOI: [10.2118/154375-MS](https://doi.org/10.2118/154375-MS).
- [447] van Leeuwen, P. J. A consistent interpretation of the stochastic version of the ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 146(731):2815–2825, 2020. DOI: [10.1002/qj.3819](https://doi.org/10.1002/qj.3819).
- [448] van Leeuwen, P. J. and Evensen, G. Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review*, 124:2898–2913, 1996. DOI: [10.1175/1520-0493\(1996\)124<2898:DAAIMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2).
- [449] Vasco, D. W., Yoon, S., and Datta-Gupta, A. Integrating dynamic data into high-resolution reservoir models using streamline-based analytic sensitivity coefficients. *SPE Journal*, 4(4):389–399, 1999. DOI: [10.2118/59253-PA](https://doi.org/10.2118/59253-PA).
- [450] Vink, J. C., Gao, G., and Chen, C. Bayesian style history matching: Another way to under-estimate forecast uncertainty? In *Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, 28-30 September*, number SPE-175121-MS, 2015. DOI: [10.2118/175121-MS](https://doi.org/10.2118/175121-MS).
- [451] Virieux, J. and Operto, S. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 76(6), 2009. DOI: [10.1190/1.3238367](https://doi.org/10.1190/1.3238367).

- [452] Vishny, D., Morzfeld, M., Gwirtz, K., Bach, E., Dunbar, O. R. A., and Hodyss, D. High-dimensional covariance estimation from a small number of samples. *Journal of Advances in Modeling Earth Systems*, 16(9), 2024. DOI: [10.1029/2024MS004417](https://doi.org/10.1029/2024MS004417).
- [453] Vo, H. X. and Durlafsky, L. J. A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models. *Mathematical Geosciences*, 46(7):775–813, 2014. DOI: [10.1007/s11004-014-9541-2](https://doi.org/10.1007/s11004-014-9541-2).
- [454] Vono, M., Dobigeon, N., and Chainais, P. High-dimensional Gaussian sampling: A review and a unifying approach based on a stochastic proximal point algorithm. *SIAM Review*, 64(1), 2022. DOI: [10.1137/20M1371026](https://doi.org/10.1137/20M1371026).
- [455] Voskov, D., Saifullin, I., Novikov, A., Wapperom, M., Orozco, L., Seabra, G. S., Chen, Y., Khait, M., Lyu, X., Tian, X., de Hoop, S., and Palha, A. Open Delft advanced research terra simulator (open-DARTS). *The Journal of Open Source Software*, 9(99), 2024. DOI: [10.21105/joss.06737](https://doi.org/10.21105/joss.06737).
- [456] Wackernagel, H. *Multivariate Geostatistics: An Introduction with Applications*. Springer Verlag, 3rd edition, 2003. DOI: [10.1007/978-3-662-05294-5](https://doi.org/10.1007/978-3-662-05294-5).
- [457] Waggoner, J. R., Cominelli, A., and Seymour, R. H. Improved reservoir modeling with time-lapse seismic in a Gulf of Mexico gas condensate reservoir. In *SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 29 September–2 October*, number SPE-77514-MS, 2002. DOI: [10.2118/77514-MS](https://doi.org/10.2118/77514-MS).
- [458] Wang, Y. and Kovscek, A. R. Streamline approach for history matching production data. *SPE Journal*, 5(4):353–362, 2000. DOI: [10.2118/58350-PA](https://doi.org/10.2118/58350-PA).
- [459] Wang, Y., Li, G., and Reynolds, A. C. Estimation of depths of fluid contacts by history matching using iterative ensemble-Kalman smoothers. *SPE Journal*, 15(2), 2010. DOI: [10.2118/119056-PA](https://doi.org/10.2118/119056-PA).
- [460] Watanabe, S., Han, J., Hetz, G., Datta-Gupta, A., King, M. J., and Vasco, D. W. Streamline-based time-lapse-seismic-data integration incorporating pressure and saturation effects. *SPE Journal*, 22(4), 2017. DOI: [10.2118/166395-PA](https://doi.org/10.2118/166395-PA).
- [461] Webb, S. J., Bayless, J. S., and Dunlop, K. N. B. Enabling the “big loop” – ensuring consistency of geological and reservoir simulation models. In *Proceedings of the AAPG Annual Convention and Exhibition, Long Beach, California, April*, 2007.
- [462] Wen, X.-H. and Gómez-Hernández, J. J. Upscaling hydraulic conductivities in heterogeneous media: An overview. *Journal of Hydrology*, 183. DOI: [10.1016/S0022-1694\(96\)80030-8](https://doi.org/10.1016/S0022-1694(96)80030-8).
- [463] Whitaker, J. S. and Hamill, T. M. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913–1924, 2002. DOI: [10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).

- [464] Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., and Toth, Z. Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review*, 136:463–482, 2008. DOI: [10.1175/2007MWR2018.1](https://doi.org/10.1175/2007MWR2018.1).
- [465] White, J. T. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. *Environmental Modelling and Software*, 109, 2018. DOI: [10.1016/j.envsoft.2018.06.009](https://doi.org/10.1016/j.envsoft.2018.06.009). URL <https://github.com/dwelter/pestpp>.
- [466] Wu, H., Fu, P., Hawkins, A. J., Tang, H., and Morris, J. P. Predicting thermal performance of an enhanced geothermal system from tracer tests in a data assimilation framework. *Water Resources Research*, 57(12), 2021. DOI: [10.1029/2021WR030987](https://doi.org/10.1029/2021WR030987).
- [467] Wu, Z., Reynolds, A. C., and Oliver, D. S. Conditioning geostatistical models to two-phase production data. *SPE Journal*, 3(2):142–155, 1999.
- [468] Xiao, C., Leeuwenburgh, O., Lin, H. X., and Heemink, A. Non-intrusive subdomain POD-TPWL for reservoir history matching. *Computational Geosciences*, 23, 2019. DOI: [10.1007/s10596-018-9803-z](https://doi.org/10.1007/s10596-018-9803-z).
- [469] Xiao, C., Zhang, S., Ma, X., Zhou, T., Hou, T., and Chen, F. Deep-learning-generalized data-space inversion and uncertainty quantification framework for accelerating geological CO₂ plume migration monitoring. *Geoenergy Science and Engineering*, 224, 2023. DOI: [10.1016/j.geoen.2023.211627](https://doi.org/10.1016/j.geoen.2023.211627).
- [470] Xu, T. and Gómez-Hernández, J. J. Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble Kalman filtering. *Water Resources Research*, 52(8), 2016. DOI: [10.1002/2016WR019111](https://doi.org/10.1002/2016WR019111).
- [471] Xu, W., Tran, T. T., Srivastava, R. M., and Journel, A. G. Integrating seismic data in reservoir modeling: the collocated cokriging approach. In *Proceedings of the SPE Annual Technical Conference and Exhibition*, number SPE-24742-MS, 1992. DOI: [10.2118/24742-MS](https://doi.org/10.2118/24742-MS).
- [472] Yao, Y., Vehtari, A., Simpson, D., and Gelman, A. Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis*, 13(3), 2018. DOI: [10.1214/17-BA1091](https://doi.org/10.1214/17-BA1091).
- [473] Yeh, W. W.-G. Review of parameter identification in groundwater hydrology: The inverse problem. *Water Resources Research*, 22(2): 95–108, 1986. DOI: [10.1029/WR022i002p00095](https://doi.org/10.1029/WR022i002p00095).
- [474] Yeung, Y.-H., Barajas-Solano, D. A., and Tartakovsky, A. M. Physics-informed machine learning method for large-scale data assimilation problems. *Water Resources Research*, 58(5), 2022. DOI: [10.1029/2021WR031023](https://doi.org/10.1029/2021WR031023).
- [475] Yin, Z., Feng, T., and MacBeth, C. Fast assimilation of frequently acquired 4D seismic data for reservoir history matching. *Computers & Geosciences*, 128:30–40, 2019. DOI: [10.1016/j.cageo.2019.04.001](https://doi.org/10.1016/j.cageo.2019.04.001).

- [476] Zachariassen, E., Skjervheim, J.-A., Vabø, J. G., Lunt, I., Hove, J., and Evensen, G. Integrated work flow for model update using geophysical monitoring data. In *Proceedings of the 73rd EAGE Conference & Exhibition, Vienna, Austria, 23–26 May, 2011*. DOI: [10.3997/2214-4609.20149329](https://doi.org/10.3997/2214-4609.20149329).
- [477] Zemel, B. *Tracers in the Oil Field*. Elsevier, first edition, 1995. ISBN 978-0444541512.
- [478] Zhang, F., Reynolds, A. C., and Oliver, D. S. Evaluation of the reduction in uncertainty obtained by conditioning a 3D stochastic channel to multiwell pressure data. *Mathematical Geology*, 34(6):713–740, 2002. DOI: [10.1023/A:1019805310025](https://doi.org/10.1023/A:1019805310025).
- [479] Zhang, F., Reynolds, A. C., and Oliver, D. S. The impact of upscaling errors on conditioning a stochastic channel to pressure data. *SPE Journal*, 8(1):13–21, 2003. DOI: [10.2118/83679-PA](https://doi.org/10.2118/83679-PA).
- [480] Zhang, J., Li, G. L. W., Wu, L., and Zeng, L. An iterative local updating ensemble smoother for estimation and uncertainty assessment of hydrologic model parameters with multimodal distributions. *Water Resources Research*, 54(3):1716–1733, 2018. DOI: [10.1002/2017WR020906](https://doi.org/10.1002/2017WR020906).
- [481] Zhang, X.-L., Michelén-Ströfer, C., and Xiao, H. Regularized ensemble Kalman methods for inverse problems. *Journal of Computational Physics*, 416, 2020. DOI: [10.1016/j.jcp.2020.109517](https://doi.org/10.1016/j.jcp.2020.109517).
- [482] Zhang, Y. and Oliver, D. S. Improving the ensemble estimate of the Kalman gain by bootstrap sampling. *Mathematical Geosciences*, 42: 327–345, 2010. DOI: [10.1007/s11004-010-9267-8](https://doi.org/10.1007/s11004-010-9267-8).
- [483] Zhang, Y., Stordal, A. S., and Lorentzen, R. J. A natural Hessian approximation for ensemble based optimization. *Computational Geosciences*, 27, 2023. DOI: [10.1007/s10596-022-10185-z](https://doi.org/10.1007/s10596-022-10185-z).
- [484] Zhao, Y., Reynolds, A. C., and Li, G. Generating facies maps by assimilating production data and seismic data with the ensemble Kalman filter. In *Proceedings of the SPE Improved Oil Recovery Symposium, Tulsa, Oklahoma, 20–23 April*, number SPE-113990-MS, 2008. DOI: [10.2118/113990-MS](https://doi.org/10.2118/113990-MS).
- [485] Zhou, H., Gómez-Hernández, J. J., Franssen, H.-J. H., and Li, L. An approach to handling non-gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources*, 34 (7), 2011. DOI: [10.1016/j.advwatres.2011.04.014](https://doi.org/10.1016/j.advwatres.2011.04.014).
- [486] Zhu, Y. and Zabaras, N. Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification. *Journal of Computational Physics*, 366, 2018. DOI: [10.1016/j.jcp.2018.04.018](https://doi.org/10.1016/j.jcp.2018.04.018).
- [487] Zubarev, D. I. Pros and cons of applying proxy-models as a substitute for full reservoir simulations. In *Proceedings of the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 4–7 October*, number SPE-124815-MS, 2009. DOI: [10.2118/124815-MS](https://doi.org/10.2118/124815-MS).

Index

Symbols

4D (time-lapse) seismic 11, 24, 169, 172,
229, 252–254, 256, 270, 275, 345, 347

A

a posteriori 14
a priori 9
adjoint method 36, 72, 348
annealed importance sampling 148
assisted history matching 22, 351

B

batch-EnRML 342
Bayes' rule 14, 19, 26, 27, 39, 114, 115,
117, 120, 121, 147, 148, 329, 330
Bayesian estimator 114
best guess approach 233
big-loop 101, 234
Broyden-Fletcher-Goldfarb-Shanno 68,
69
Brugge case 156, 177, 342, 349, 352

C

capillary pressure 237, 241
Cholesky decomposition 41–43, 169,
238, 301
conjugate gradient method 64
cookie-cutter approach 237
Corey model 239
correlation-based localization 209, 211,
344
cosine similarity 266
covariance inflation 194
covariance localization 196, 215
covariance matrix adaptation evolution
strategy 75
critical length 205

cumulative density function 48, 100,
316
curse of dimensionality 52, 87, 233

D

Darcy's law 6
data-error covariance 25, 31, 245, 254
data-mismatch objective function 269
data-space inversion 348
design of experiments 79
deterministic ensemble Kalman filter
168, 340
deterministic ensemble smoother with
multiple data assimilation 168
discrete cosine transform 100, 346
distance-based localization 200, 202,
204
domain localization 214, 215, 344

E

ensemble adjustment Kalman filter 340
ensemble Kalman filter 15, 103, 122,
133, 339
ensemble Kalman inversion 349
ensemble randomized maximum
likelihood 16, 175, 341
ensemble smoother 16, 139, 342
ensemble smoother with multiple data
assimilation 140, 142, 342
ensemble square root filter 137, 340
ensemble transform Kalman filter 340
ensemble-based optimization 349
evolutionary algorithm 74
extended Kalman filter 121, 339

F

facies 10, 100, 237, 336, 342, 346
facies overlap coefficient 267
faults 241

forward model 9, 20, 35, 50, 71, 79, 183, 252
forward problem 5
full waveform inversion 355
Furrer-Bengtsson pseudo-optimal taper 200, 211

G

Gaspari-Cohn correlation function 200, 204, 235, 345
Gauss-Newton 66, 175, 341
generative adversarial network 347
generative models 347
genetic algorithm 73–75
geological carbon storage 348, 349, 354
geological reservoirs 1, 2, 4, 10, 11
geostatistics 11, 13, 47, 169, 233, 235, 258, 331, 333–335
gradual deformation 95, 348
gradzone 99
grid parameters 234, 237, 241

H

half-iteration EnKF 341
hard data 47, 235, 236, 258
history matching 11, 22, 351

I

importance sampling 148, 233, 351
integrated workflows 234
inverse problem 8
iterative adaptive Gaussian mixture smoother 175, 342
iterative local updating ensemble smoother 175, 351

K

Kalman filter 15, 16, 104, 112, 121
Kalman gain 107, 110, 114, 217
Kalman gain localization 196, 199, 217, 228, 230, 271, 282
kernel principal component analysis 100, 346
kriging 11, 47, 81, 331, 334, 335

L

Latin hypercube sampling 80, 348
LET model 239
level-set function 100, 346
Levenberg-Marquardt 67, 68, 177, 342

Levenberg-Marquardt ensemble randomized maximum likelihood 342
likelihood 20, 21, 329
line search 71
local analysis 214
local ensemble transform Kalman filter 340, 355
localization 16, 134, 156, 168, 184, 187, 196–198, 207, 228, 232, 234, 235, 238, 242, 343
locally varying anisotropy 336
Lorenz 96 model 343

M

Mahalanobis distance 262
manual history matching 94
Markov chain Monte Carlo 15, 48, 342
maximum a posteriori 27
maximum likelihood 21, 329
mean squared error 262
model parameters 5, 7, 20, 103, 114, 115, 140
Monte Carlo 15, 19, 40, 80, 88, 122, 148, 339, 355
Moore-Penrose pseudoinverse 220, 298
Morris sensitivity analysis 88
multiple deterministic approach 233
multiple stochastic approach 233
multiple-point statistics 101, 337, 346

N

net present value 352
net-to-gross ratio 236
Newton's method 65
noise-informed covariance estimation 345
nonlinear conjugate gradient 65, 67
nonlinear conjugate gradient method 64
normal score transform 332
normalized model change 266
normalized objective function 57
normalized root-mean-squared error 255
normalized variance 31, 267
numerical weather prediction 122, 137, 340

O

object-based simulation 101, 337, 346
objective function 22, 27, 44, 175, 182

observation coverage 261
 ocean bottom nodes 275
 optimal hard threshold 225, 312
 optimization 35, 51
 optimization-based principal component analysis 100

P

parametrization 92, 99, 101, 231
 particle filter 342
 particle swarm optimization 73, 76, 348
 Pearson correlation 89, 320, 321
 permeability 1, 3, 4, 6, 7, 10, 32, 35, 234, 236
 perturbed observations 126, 127, 340
 petroelastic model 253, 254
 pilot-point 94
 porosity 1, 4, 6, 10, 234, 236
 posterior covariance 30, 33, 126, 143, 144, 170, 189, 191, 267
 prediction-focused analysis 348
 pressure transient analysis 236, 256
 principal component analysis 96, 346
 prior covariance 26, 97, 177, 190, 191
 probability density function 14, 20, 26, 35, 40, 43, 47, 49, 54, 96, 119, 120, 122, 123, 144, 147, 190, 316–318, 323–326, 328, 331, 332
 probability maps 100, 346
 probability mass function 315
 process-based modeling 337
 production data 243
 production-logging tools 257
 proxy modeling 50, 79, 80, 86
 PUNQ-S3 case 91, 226, 287

Q

quasi-Newton 35, 36, 56, 67, 68

R

randomized maximum likelihood 50
 rank deficiency 187
 regularized Levenberg-Marquardt for minimum-average-cost 175, 343
 relative permeability 4, 6, 237, 239, 271, 275, 282, 285
 repeat formation testers 257
 representative models 287
 rule-based modeling 101, 337

S

sampling errors 197

scalar parameters 207, 213, 234, 237, 241
 scaling 70
 Schur or Hadamard product 196, 197, 343
 Schur product theorem 197, 198
 sensitivity analysis 213
 sequential data assimilation 103, 119, 194, 341
 sequential Gaussian simulation 41, 169, 335
 sequential indicator simulation 336
 side lobes 253
 simplex gradient 350
 simulated annealing 73, 77, 78, 348
 simultaneous perturbation stochastic approximation 72, 350
 singular value decomposition 164, 179, 219, 297
 Sobol sensitivity analysis 88
 Spearman correlation 89
 spurious correlations 187
 steepest descent method 62
 stochastic optimization 73
 stochastic simplex approximate gradient 350
 stratigraphic rule-based modeling 346
 streamline simulation 72
 structural parameters 243
 subspace iterative ensemble smoother 16, 182, 343
 support vector machine 83
 support vector regression 83

T

termination criteria 70
 tracer 11, 255, 256
 transmissibility multipliers 241
 truncated Gaussian 100, 336
 truncated plurigaussian 100, 237, 337, 346
 turbidite 9, 267, 271, 282
 two-point geostatistics 335

U

uncertainty 4, 23, 339, 341, 345, 347, 349

V

variational autoencoder 347
 variogram 102, 333, 335, 346

W

water-alternating-gas 256, 275

Z

zonation 93

