



Universidade de Brasília
Faculdade de Tecnologia
Departamento de Engenharia Mecânica

Gated Data-Space Inversion for Reservoir Prediction

Thiago Tomás de Paula

DISSERTAÇÃO DE MESTRADO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MECÂNICAS

Brasília
2026

Universidade de Brasília
Faculdade de Tecnologia
Departamento de Engenharia Mecânica

Previsão de Reservatórios usando Inversão do Espaço de Dados Ponderada

Thiago Tomás de Paula

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Ciências Mecânicas como requisito parcial para obtenção do grau de Mestre.

Orientador: Prof. Dr. Eugênio Libório Feitosa Fortaleza

Brasília
2026

FICHA CATALOGRÁFICA

P324g Paula, Thiago Tomás de.
Gated Data-Space Inversion for Reservoir Prediction / Thiago Tomás de Paula; orientador Eugênio Libório Feitosa Fortaleza. -- Brasília, 2026.
89 p.

Dissertação de Mestrado (Programa de Pós-Graduação em Ciências Mecânicas) -- Universidade de Brasília, 2026.

1. Multiple Data Assimilation. 2. Reservoir prediction. 3. Data-Space Inversion. I. Fortaleza, Eugênio Libório Feitosa, orient. II. Título.

Universidade de Brasília
Faculdade de Tecnologia
Departamento de Engenharia Mecânica

Gated Data-Space Inversion for Reservoir Prediction

Thiago Tomás de Paula

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Ciências Mecânicas como requisito parcial para obtenção do grau de Mestre.

Trabalho aprovado. Brasília, 11 de agosto de 2026:

Prof. Dr. Eugênio Libório Feitosa Fortaleza
UnB/FT/ENM
Orientador

Prof. Dr. Aline Souza de Paula
UnB/FT/ENM
Examinador interno

Dr. Arthur Bizzi
Imperial College London
Examinador externo

To all fascinated with almost understanding things

Acknowledgements

I'd like to thank Matheus Teixeira de Sousa and Yan Tavares de Oliveira for the help given within and beyond the context of lectures, research projects and papers. I believe our small talk, discussions, and even arguments, have shaped me into a better researcher and person. I treasure our friendship and will count it for balancing out my opinions and setting me straight for as long as you will have me.

I'd like to thank professor José Oniram de Aquino Limaverde de Filho for both co-authoring my first paper in Control Engineering and for accepting me as his doctorate student moving forward. I cannot understate the appreciation I have for your faith in me and interest in my work, and I hope our doctorate journey will result in no small number of control papers.

*“Engineers use science to solve their problems if the science is available.
But available or not, the problem must be solved,
and whatever form the solution takes under these conditions
is called engineering.”
(from Shigley’s Mechanical Engineering Design)*

Artificial Intelligence Disclaimer

The author declares that the content of this dissertation was developed without the use of artificial intelligence tools, such as ChatGPT, Claude, Grammarly or similar platforms, applications, or agents, for the generation of text, figures, tables, or any other content. Unless stated otherwise, all content of this dissertation, such as text, figures, tables, etc., was solely produced and reviewed by the author, who takes full responsibility for any and all errors therein.

Sensitive Data Disclaimer

One of this dissertation's case studies is a real oilfield in the Brazilian Post-Salt, whose data are sensitive. Petrobras SA has given permission to use, and to some extent, disclose, data related to this field for the purposes of this work.

Abstract

Accurate prediction of reservoir performance under uncertainty is essential for decision-making in geological reservoir management. Data-Space Inversion (DSI) provides an attractive alternative to traditional history matching by updating ensemble forecasts directly in data space, thereby avoiding repeated subsurface simulations. Recently, DSI-ESMDA was proposed. It is an extension of DSI that employs the Ensemble Smoother with Multiple Data Assimilation (ESMDA) to replace the Randomized Maximum Likelihood minimizations of the original formulation with an iterative Kalman-type update, achieving significant computational efficiency while maintaining competitive posterior forecasts. Despite these advantages, DSI-ESMDA, like other ensemble smoothers, remains susceptible to excessive variance loss, which can lead to underestimation of posterior uncertainty. Methods such as covariance localization help to ameliorate this issue, but variance underestimation is still observed in practical applications. One contributing factor to this progressive loss of variance is that small production mismatches, often dominated by measurement noise or sampling effects, may trigger repeated micro-updates that reduce ensemble diversity without improving predictive skill. This work introduces an extension to DSI-ESMDA that modifies the effective influence of each observation through a nonlinear, misfit-dependent weighting of the innovation vector. The proposed formulation can be interpreted as introducing a smoothing influence function into the Kalman-type update, computed from normalized innovations using an error-function-based gating operator. In contrast to covariance localization, which acts on second-order statistics, innovation gating acts directly on the update term, selectively suppressing corrections associated with sufficiently-matched observations while preserving the full impact of significant mismatches. Match quality is thresholded by a gating parameter during innovation normalization. The gating methodology is inspired by adaptive filtering developed to mitigate actuator wear in closed-loop control systems. Although the present work focuses on DSI, the proposed formulation is general and can be applied to standard ensemble-based model-space inversion methods, including ESMDA and related iterative smoothers. The approach is evaluated on multiple synthetic reservoirs and one real field application using production rates from producers and injectors. Results show that innovation-gated DSI-ESMDA reduces variance loss during and produces more robust uncertainty envelopes in the prediction phase without degrading the data match. Furthermore, for moderate gating parameters, posterior forecasts remain close to those obtained with standard DSI-ESMDA, indicating that the proposed method acts as a stabilizing extension rather than a replacement.

Keywords: Multiple Data Assimilation; Reservoir prediction; Data-Space Inversion.

Resumo

A previsão precisa do desempenho de reservatórios mediante incerteza é essencial para a tomada de decisões no gerenciamento geológico de reservatórios. A Inversão no Espaço de Dados (DSI - Data-Space Inversion) oferece uma alternativa atraente ao ajuste de histórico tradicional na medida em que atualiza previsões de ensemble diretamente no espaço de dados e, assim, evita repetir simulações de subsuperfície. Recentemente, foi introduzido o DSI-ESMDA, uma extensão do DSI que emprega o Ensemble Smoother with Multiple Data Assimilation (ESMDA) para substituir as minimizações de Máxima Verossimilhança Aleatorizada (Randomized Maximum Likelihood) no DSI original por uma atualização iterativa similar ao filtro de Kalman, alcançando significativa eficiência computacional e mantendo previsões a posteriori competitivas. Apesar dessas vantagens, o DSI-ESMDA, assim como outros suavizadores de ensemble, permanece suscetível à perda excessiva de variância, o que pode levar à subestimação da incerteza a posteriori. Métodos como a localização de covariância ajudam a mitigar esse problema, mas a subestimação da variância ainda é observada em aplicações práticas. A perda progressiva de variância é em parte causada por pequenas discrepâncias na produção, frequentemente dominadas por ruído de medição ou efeitos de amostragem, podem desencadear microatualizações repetidas que reduzem a diversidade do ensemble sem melhorar a capacidade preditiva. Este trabalho apresenta uma extensão do DSI-ESMDA que modifica a influência efetiva de cada observação por meio de uma ponderação não linear do vetor de inovação, dependente da discrepância com os dados observados. A formulação proposta pode ser interpretada como a introdução de uma função de influência de suavização na atualização do ensemble. Ao contrário da localização de covariância, que atua sobre estatísticas de segunda ordem, a filtragem de inovações atua diretamente no termo de atualização, suprimindo correções associadas ajustes satisfatórios, e preservando correções de discrepâncias significativas. A formulação proposta é geral e pode ser aplicada a métodos padrão de inversão no espaço do modelo baseados em ensemble, incluindo o ESMDA e suavizadores correlatos. A abordagem é avaliada em reservatórios sintéticos e em uma aplicação em campo real utilizando taxas de produção de poços produtores e injetores. Resultados mostram que o DSI-ESMDA Ponderado reduz a perda de variância e gera envelopes de incerteza mais robustos na fase de previsão, sem comprometer o ajuste aos dados. Ponderações moderadas apresentam comportamento similar ao DSI-ESMDA padrão, indicando que o método proposto atua como uma extensão estabilizadora, ao invés de uma substituição.

Palavras-chave: Múltipla Assimilação de Dados; Predição de reservatório; Inversão do espaço de dados.

List of figures

Figure 1.1	Management steps.	22
Figure 1.2	Illustration of the inverse problem.	23
Figure 1.3	Overview of the history matching workflow.	23
Figure 2.1	Mindmap of reservoir prediction methods	27
Figure 2.2	Timeline of main works presented in the literature review	27
Figure 3.1	Synchronization between realization data vector and reference data. . .	37
Figure 3.2	Relationship between model parameters and data vectors.	38
Figure 3.3	Posterior distributions obtained with SDA (green) and MDA (yellow) for an observation vector with 2 points of data. Circles denotes means, and circumferences, covariances. True underlying posterior is denoted in blue, and the prior in gray.	40
Figure 3.4	Gaspari-Cohen 5th-order compact correlation function.	45
Figure 3.5	Important functions for understanding the proposed Data-Space Inversion algorithm.	48
Figure 3.6	Illustration of the effect of λ on the posterior ensemble for fixed $\mathbf{R}$ and N_a . .	49
Figure 4.1	Top view of selected realizations 1-10 from the Egg ensemble. Renderings were produced using the open-source software ResInsight.	52
Figure 4.2	Top view of selected realizations from the Olympus ensemble. Renderings were produced using the open-source software ResInsight.	54
Figure 4.3	Distribution of field oil production totals for each realization in the Egg, Olympus, and real field ensembles. Each ensemble is normalized by its maximum FOPT.	56
Figure 4.4	Outline of the evaluation process.	57
Figure 5.1	Average posterior coverage and MSE values found in the Egg ensemble. .	62
Figure 5.2	Summary of results for the Olympus ensemble. Posterior coverage and MSE are averaged across ground-truths.	63
Figure 5.3	Real field FOPT CDF.	64
Figure C.1	Prior and posterior ensembles obtained with $\lambda = 4$ gating in the real field. .	86
Figure C.2	Continuation of figure C.1	87
Figure C.3	Continuation of figure C.1	88
Figure C.4	End of figure C.1	89

List of tables

Table 4.1	Some geometry and fluid properties of the Egg ensemble.	52
Table 4.2	Types and distribution range intervals of facies properties in Olympus. .	53
Table 4.3	Olympus' fluid properties.	54
Table 4.4	Time-related simulation parameters per reservoir.	59
Table 4.5	Localization-related simulation parameters per reservoir. Lengths in grid-blocks.	59
Table 4.6	Data features used per reservoir well.	59
Table 5.1	Iteration and localization construction times after 10 runs.	61
Table 5.2	Component specifications of the machine used in order to run and benchmark experiments.	61
Table 5.3	DSI-ESMDA results for the real field.	64
Table 5.4	Gated DSI-ESMDA results for the real field with varying gating.	64
Table A.1	Summary of the disciplines completed by the author during their master's degree.	80
Table A.2	Summary of the author's publications during their master's degree. . . .	81
Table A.3	Summary of the software developed by the author during their master's degree.	81
Table C.1	DSI-ESMDA.	83
Table C.2	Gated DSI-ESMDA ($\lambda = 1$).	83
Table C.3	Gated DSI-ESMDA ($\lambda = 2$).	84
Table C.4	Gated DSI-ESMDA ($\lambda = 4$).	84
Table C.5	Gated DSI-ESMDA ($\lambda = 8$).	84
Table C.6	Gated DSI-ESMDA ($\lambda = 16$).	84
Table C.7	DSI-ESMDA.	85
Table C.8	Gated DSI-ESMDA ($\lambda = 1$).	85
Table C.9	Gated DSI-ESMDA ($\lambda = 2$).	85
Table C.10	Gated DSI-ESMDA ($\lambda = 4$).	85
Table C.11	Gated DSI-ESMDA ($\lambda = 8$).	86
Table C.12	Gated DSI-ESMDA ($\lambda = 16$).	86

List of abbreviations and acronyms

BHP	Bottomhole Pressure
CDF	Cumulative Distribution Function
CEM	Connection Element Method
DL	Deep Learning
DSI	Data-Space Inversion
DSI-ESMDA	Data-Space Inversion with Ensemble Smoother with Multiple Data Assimilation
DSIVC	Data-Space Inversion with Variable Controls
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
ENM	Department of Mechanical Engineering
ES	Ensemble Smoother
ESMDA	Ensemble Smoother with Multiple Data Assimilation
FOPT	Field Oil Production Total
FT	Faculty of Technology
IES	Iterative Ensemble Smoother
KF	Kalman Filter
MDA	Multiple Data Assimilation
ML	Machine Learning
MSE	Mean Square Error
NG	Net to Gross
OC	Observation Coverage
OPR	Oil Production Rate
PFA	Prediction Focused Analysis
RML	Randomized Maximum Likelihood
SDA	Single Data Assimilation
UnB	University of Brasília
WCTP	Water Cut Percentage
WIR	Water Injection Rate
WPR	Water Production Rate

List of symbols

Roman letters

ℓ	Data assimilation index
$\mathbb{1}$	Indicator function
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^{x \times y}$	Set of real $x \times y$ matrices
$\mathbf{C}_e^\ell$	Data error covariance matrix of the ℓ -th data assimilation
$\mathbf{C}_e$	Data error covariance matrix
$\mathbf{C}_p$	Prior distribution covariance matrix
$\mathbf{C}_{dh,p}$	Prior distribution history period cross-covariance matrix
$\mathbf{C}_{hh,p}$	Prior distribution history period covariance matrix
$\mathbf{d}$	Ground-truth data vector
$\mathbf{d}_{\text{obs}}^\ell$	Resampled/perturbed observed data vector of the ℓ -th data assimilation
$\mathbf{d}_{\text{obs}}$	Observed data vector
$\mathbf{d}_f$	Ground-truth feature f data vector
$\mathbf{d}_h$	History period ground-truth data vector
$\mathbf{d}_j$	Realization j data vector
$\mathbf{d}_v$	Validation period ground-truth data vector
$\mathbf{d}_{h,j}$	History period realization j data vector
$\mathbf{d}_{j,f}$	Realization j feature f data vector
$\mathbf{d}_{v,j}$	Validation period realization j data vector
$\mathbf{e}_j^\ell$	Realization j observed data perturbation of the ℓ -th data assimilation
$\mathbf{e}_d$	Measurement noise
$\mathbf{g}$	Gravitational acceleration vector
$\mathbf{g}(\cdot, \cdot)$	Prediction forward model
$\mathbf{H}$	History period extractor matrix
$\mathbf{I}$	Identity matrix
$\mathbf{K}$	Permeability tensor of the porous medium
$\mathbf{K}^\ell$	Kalman-like gain of the ℓ -th data assimilation
$\mathbf{m}$	Ground-truth model parameters
$\mathbf{m}_j$	Realization j model parameters
$\mathbf{R}$	Localization matrix
$\mathbf{u}$	Reservoir control vector
$\mathbf{x}_0$	Initial reservoir state
$\mathcal{L}(\cdot)$	Likelihood function

$\mathcal{L}^\ell(\cdot)$	Likelihood function of the ℓ -th data assimilation
$\mathcal{N}(\cdot, \cdot)$	Normal distribution
$O(\cdot)$	Data-mismatch objective function
$O(\cdot, \cdot, \cdot)$	Posterior objective function
$O_{\mathcal{L}}(\cdot, \cdot)$	Likelihood objective function
$O_{\mathcal{L}}^\ell(\cdot, \cdot)$	Likelihood objective function of the ℓ -th data assimilation
$O_p(\cdot, \cdot)$	Prior objective function
$\overline{\Delta \mathbf{D}}$	Ensemble spread
$\overline{\Delta \mathbf{D}}_h$	History period ensemble spread
$\bar{\mathbf{d}}$	Ensemble mean average
$\bar{\mathbf{d}}_h$	History period ensemble mean average
$\widetilde{\mathbf{d}}_f$	Ground-truth normalized feature f data vector
$\mathbf{d}_f$	Realization j normalized feature f data vector
L'_x	Critical localization distance along anisotropic x' -axis
L'_y	Critical localization distance along anisotropic y' -axis
N_a	Number of data assimilations
N_e	Number of ensemble realizations
N_h	Number of data points in the history period
N_v	Number of data points in the validation period
$p(\cdot)$	Probability distribution
$p^\ell(\cdot)$	Posterior distribution after ℓ data assimilations
$r(\cdot)$	Localization function
r_{go}	Ratio of dissolved gas to oil in the oleic phase
r_{og}	Ratio of vaporized oil to gas in the gaseous phase
R_{uv}	Localization factor between datum u and v
T_c	Critical localization time

Greek letters

α^ℓ	Covariance inflation factor of the ℓ -th data assimilation
$\alpha_{j,d}^\ell$	Gating coefficient at datum d for the j -th realization during the ℓ -th data assimilation
$\boldsymbol{\alpha}_j^\ell$	Gating coefficient vector for the j -th realization during the ℓ -th data assimilation
Δt_{uv}	Time difference between datum u and v
$\Delta x'_{uv}$	x' -axis distance between datum u and v
Δx_{uv}	x -axis distance between datum u and v
$\Delta y'_{uv}$	y' -axis distance between datum u and v
Δy_{uv}	y -axis distance between datum u and v

δ_{uv}	Localization distance between datum u and v
λ	Gating intensity
λ_α	Mobility of phase α
$\mathbf{u}_\alpha$	component velocity of component α
μ_α	Viscosity of phase α
ν	Noise magnitude
ρ_α	Density of phase α
$\rho_{S,\alpha}$	Surface density of phase α at one atmosphere
σ_d	Standard deviation of measurement noise on datum d
θ	Anisotropy angle
ϕ_{ref}	Reference porosity
A_α	Phase α 's component mass
b_ϕ	Shrinkage/expansion factor for phase α
$k_{r,\alpha}$	Relative permeability for phase α
m_ϕ	Pore volume multiplier
p_α	Phase pressure for phase α
$p_{c,\alpha\beta}$	Capillary pressure between phases α and β
q_α	Well outflux density of pseudo component α
s_ϕ	Saturation of phase α

Contents

1	Introduction	20
1.1	Work overview	20
1.2	Background context	21
1.2.1	What is oil?	21
1.2.2	What is a geological reservoir?	21
1.2.3	The inverse problem	22
1.3	Research problem	24
1.4	Research questions	24
1.5	Significance	24
1.6	Methodology overview	25
1.7	Dissertation outline	25
2	Literature review	26
2.1	Chapter outline	26
2.2	Model-space methods	28
2.2.1	Ensemble Kalman Filter	28
2.2.2	Iterative Ensemble Smoother	29
2.2.3	Ensemble Smoother with Multiple Data Assimilation	29
2.3	Data-space methods	30
2.3.1	Prediction Focused Analysis	30
2.3.2	Data Space Inversion	30
2.3.3	Data Space Inversion with ESM DA	31
2.3.4	Machine Learning and Deep Learning	32
2.4	Chapter summary	33
3	Theory framework	34
3.1	Chapter outline	34
3.2	Prediction forward model	34
3.3	DSI-ESMDA	36
3.3.1	Key concepts and definitions	36
3.3.2	Statistical Framework	38
3.3.3	Deriving the update equation	40
3.3.4	Improving data match with covariance localization	43
3.4	Gated DSI-ESMDA	45
3.4.1	Deriving the gating coefficient	45
3.4.2	Deriving the update equation	48

3.5	Chapter summary	49
4	Methods	51
4.1	Chapter outline	51
4.2	Reservoir case studies	51
4.2.1	Egg ensemble	51
4.2.2	Olympus ensemble	53
4.2.3	Real field	54
4.2.4	Field oil production totals	55
4.3	Evaluation workflow and parameters	56
4.3.1	Evaluation workflow	56
4.3.2	Noise modeling	57
4.3.3	Evaluation metrics	58
4.3.4	Evaluation parameters	59
4.4	Chapter summary	60
5	Results	61
5.1	Chapter outline	61
5.2	Egg ensemble	62
5.3	Olympus ensemble	63
5.4	Brazilian post-salt field	64
6	Discussion	65
6.1	Summary of key findings	65
6.2	Interpretation of findings	65
6.2.1	OC, MSE and mismatch increase	65
6.2.2	Mixed effects of localization	66
6.2.3	Mild and small gating intensities versus standard DSI-ESMDA	66
6.3	Evaluation of existing theories and models	67
7	Conclusion	68
7.1	Summary of findings	68
7.2	Contribution to knowledge	68
7.3	Practical implications	68
7.4	Limitations and future research	69
7.5	Closing statement	69
	References	70
	Appendices	79
	Appendix A Academic productions	80
A.1	Disciplines completed	80

A.2 Publications	80
A.3 Software	81
Appendix B Proof of the simplified Kalman gain formula	82
Appendix C Detailed results	83
C.1 Egg ensemble	83
C.2 Olympus ensemble	85
C.3 Real field	86

1 Introduction

This document is the author's master's dissertation submitted to the Mechanical Sciences Graduate Program at University of Brasília as a partial requisite in acquiring the master's degree. Remaining requisites the author has satisfied are presented in [appendix A](#).

1.1 Work overview

The economic success of oil and gas fields is increasingly contingent on the quality of reservoir management decisions, which are made under uncertainty and based on limited data. Kalman-based ensemble methods ([Evensen, 1994](#); [Leeuwen; Evensen, 1996](#); [Emerick; Reynolds, 2013a](#); [Skjervheim et al., 2011](#)) are widely used to quantify uncertainty in reservoir forecasts by adjusting reservoir models to match production data, but require vast computational resources to perform repeated reservoir simulations. In an effort to reduce the computational burden, Data-Space Inversion (DSI) ([Sun; Durlafsky, 2017a](#)) was introduced as an alternative to traditional history matching by updating ensemble forecasts directly in data space, thereby avoiding repeated subsurface simulations.

[Lima, Emerick, and Ortiz \(2020\)](#) introduced DSI-ESMDA, an extension of DSI that employs the Ensemble Smoother with Multiple Data Assimilation (ESMDA) ([Emerick; Reynolds, 2013a](#)) to replace the Randomized Maximum Likelihood (RML) minimizations of the original formulation with an iterative Kalman-type update, achieving significant computational efficiency while maintaining competitive posterior forecasts. However, DSI-ESMDA, like related ensemble smoothers, remains susceptible to excessive variance loss, which can lead to underestimation of posterior uncertainty.

This work introduces an extension to DSI-ESMDA that modifies the effective influence of each observation through a nonlinear, misfit-dependent weighting of the innovation vector. The proposed formulation can be interpreted as introducing a smoothing influence function into the Kalman-type update, computed from normalized innovations using an error-function-based gating operator. Findings on both synthetic and real reservoirs show that Gated DSI-ESMDA reduces variance loss during and produces more robust uncertainty envelopes in the prediction phase without degrading the data match.

Gated DSI-ESMDA acts as a stabilizing extension rather than a replacement, and can be applied to standard ensemble-based model-space inversion methods, including ESMDA and related iterative smoothers. The power of this contribution is that it provides a simple yet effective mean of improving the robustness of ensemble-based inversion methods without requiring additional simulations or complex modifications to existing workflows.

1.2 Background context

This section provides a brief presentation of necessary background concepts, i.e., a base understanding of oil and the formation of a reservoir, as well as a literature review on methods employed for the reservoir prediction.

1.2.1 What is oil?

In the context of reservoir engineering, oil refers to the more precise term petroleum, a combination of the latin *petra* (rock) and *oleum* (oil in the general sense). Petroleum is an umbrella term for complex mixtures resulting from the deposition and decomposition of organic matter during a long period of time. It is composed of hundreds of chemical components, with the main constituent being hydrocarbons, i.e., organic compounds primarily composed of carbon and hydrogen atoms.

In its liquid phase, petroleum is inflammable, less dense than water, has a characteristic smell, and a color anywhere between black and light-brown. Petroleum is extracted from geological reservoirs, and its components, especially the hydrocarbons, used in applications that range from paving roads, to making plastic-based constructs, to fueling airplanes and cars, turning oil production into an extremely profitable economic activity.

1.2.2 What is a geological reservoir?

Geological reservoirs are subsurface formations capable of storing and transmitting fluids, such as hydrocarbons (oil and gas), water, hydrogen, carbon dioxide, and waste materials from industrial processes (Emerick, 2025, section 1.1).

In order for oil to exist within a reservoir, it must occupy empty spaces within appropriate rock formations. The amount of empty space a rock has is called porosity, since it comes in the form of small pores in the material. How much of that empty space is actually occupied is called saturation, and changes in time. Importantly, in order to oil to be extractable from inside the reservoir, these pores filled with oil must be sufficiently connected, allowing the fluid to come through it. The ease through which a fluid can flow through a porous media is called permeability.

In conventional petroleum reservoirs, oil is stored inside relatively porous and permeable rock formations that are trapped by much less permeable rock formations or geological faults. In contrast, unconventional reservoirs have the oil in porous rocks so impermeable the sealing property of a cap rock is unneeded.

If there is economic interest in a petroleum reservoir, i.e., it is sufficiently large (possibly by linking with other reservoirs), it can be referred to in the industry as an oil field. In this work, the terms reservoir, oil field, or simply field, are used interchangeably. When oil

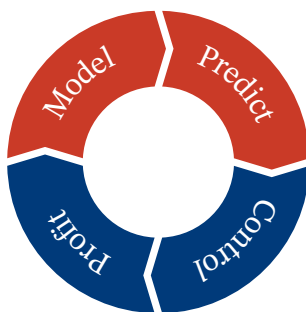
production is no longer economically feasible, the field is either abandoned or repurposed. This is the end of the reservoir's life, which started when wells were first completed.

1.2.3 The inverse problem

Oil production has been established as an extremely profitable, but extremely difficult economic activity. The scarcity of this natural resource pushes the search for new reservoirs offshore and in increasingly unfavorable geological formations, of which the discovery and extraction of oil from the Brazilian pre-salt is the prime example. As extraction becomes less profitable, strategies for oil production enhancement become attractive. This usually means closing (at least partially) producer wells, or injecting fluids into the reservoir (water and gas being the most common in this context) through injector wells in order to favorably manipulate the permeability and/or pressure within the field, whose aging means the plunging of pressure and oil saturation.

Controlling the reservoir often requires modeling and prediction of subsurface dynamics and rock properties (figure 1.1).

Figure 1.1 – Management steps.



Source: Self-elaboration.

More specifically, given the observations coming from an oil field (e.g., oil production total, pressure at a well-head), and other geological information on the reservoir (e.g., porosity and permeability of a rock sample), it is required to find the underlying parameters and states of the reservoir. Parameters are physical properties of the reservoir that remain constant throughout its life, their main representatives being rock permeability and porosity.

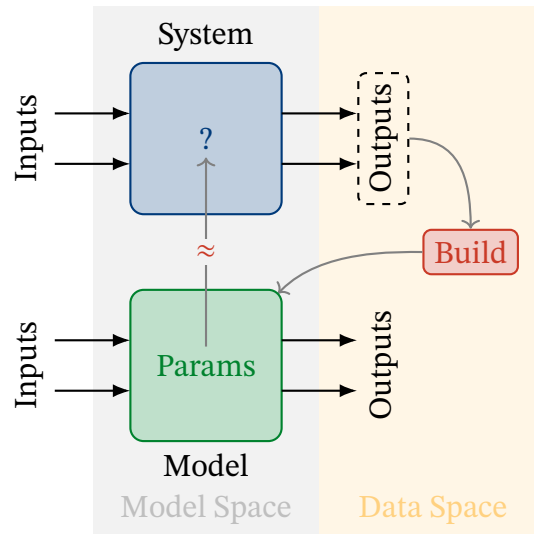
Conversely, states are physical properties that vary throughout reservoir life, their main representatives being saturation and pressure (Emerick, 2025, section 4.1). Once identified, parameters and states can be used to construct a model of the reservoir, which in turn can be deployed to generate predictions of its reference (figure 1.2).

Generally, inverse problems consist of calculating the causes of a phenomenon given observations of this phenomenon. In the context of geological reservoirs, the inverse problem is typically ill-posed. Hadamard (1902) postulates three properties a problem should satisfy in order to be deemed well-posed:

1. a solution to the equation exists (existence);
2. the solution is unique (uniqueness);
3. continuous changes in initial or boundary conditions cause continuous changes in the solution (stability).

A problem that does not satisfy the three criteria above is said ill-posed. Hadamard

Figure 1.2 – Illustration of the inverse problem.



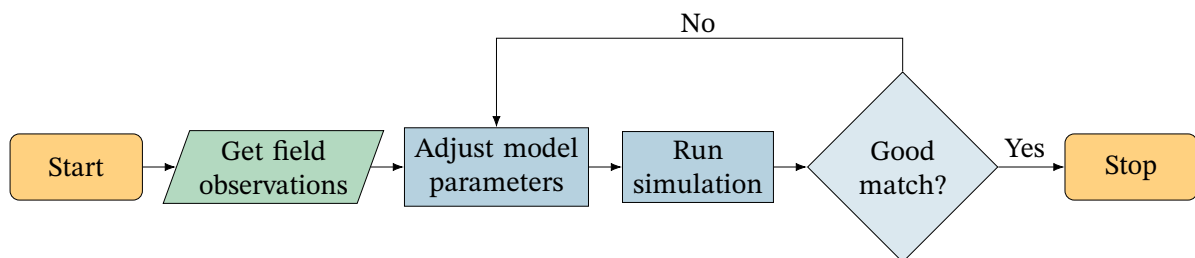
Source: Self-elaboration.

was concerned with partial differential equations in his work, but the notion of well- and ill-posedness is applicable to a wide range of applications, including the inverse problem.

At its most stringent formulation, the inverse problem tries to find model parameters that would reproduce, exactly, the field observations. However, since no modeling of dynamics is perfect, and field observations always come with uncertainty due to measurement noise, this formulation of the problem either has no solution, or the solution is sensitive to noise.

Intuitively, one could relax the stringent formulation so that a model that outputs *sufficiently* close matches to observations can be taken as a solution, adopting the workflow illustrated in [figure 1.3](#). This formulation mitigates model sensitivity to noise in observations,

Figure 1.3 – Overview of the history matching workflow.



Source: Self-elaboration.

but the problem remains ill-posed in account of the uniqueness criterion. Modern history matching processes adjust an ensemble of field models. These models are often referred to as realizations, a term borrowed from Probability Theory denoting that the object (model) is an instance, or sampling, from the same probability distribution that the reference (field) comes from.

Using an ensemble of realizations allows for the quantification of uncertainty in exchange of requiring a great deal of subsurface simulations. As [chapter 2](#) demonstrates, repeated model simulations remains a cost to be mitigated despite great progress done in making the history matching process computationally efficient.

1.3 Research problem

Literature on DSI and DSI-ESMDA take different approaches to mitigating variance loss, but they are either computationally expensive or require careful tuning of hyperparameters. Ideally, a solution would be simple to implement, require minimal tuning, and be computationally efficient. This work proposes introducing a straightforward gating mechanism on DSI-ESMDA's innovation vector, which selectively suppresses the influence of well-matched observations while preserving the impact of significant mismatches. Gated DSI-ESMDA thus modulates the posterior ensemble variance with a single parameter, the intensity of the gating function.

1.4 Research questions

This work addresses the following research questions:

1. Can a gating mechanism be introduced into DSI-ESMDA to selectively modulate the influence of observations based on their match quality?
2. Does the proposed innovation gating reduce variance loss during assimilation and produce more robust uncertainty envelopes in the prediction phase without degrading the data match?
3. How does the performance of Gated DSI-ESMDA compare to standard DSI-ESMDA across synthetic and real reservoir case studies?

1.5 Significance

This work introduces an extension to DSI-ESMDA that modifies the effective influence of each observation through a nonlinear, misfit-dependent weighting of the innovation vector. The proposed formulation can be interpreted as introducing a smoothing influence function into the Kalman-type update, computed from normalized innovations using an error-function-based gating operator. In contrast to covariance localization, which acts on second-order statistics, innovation gating acts directly on the update term, selectively suppressing corrections associated with sufficiently-matched observations while preserving the full impact of significant mismatches. Match quality is thresholded by a gating parameter during innovation normalization. The gating methodology is inspired by adaptive filtering

developed to mitigate actuator wear in closed-loop control systems (Fortaleza *et al.*, 2022) and is conceptually related to smooth shrinkage and noise-gating strategies in signal processing (Donoho; Johnstone, 1994; Donoho, 1995). Although the present work focuses on DSI, the proposed formulation is general and can be applied to standard ensemble-based model-space inversion methods, including ESM DA and related iterative smoothers. The proposed method inherits the computational efficiency of DSI-ESM DA while providing a robust, easily tunable mechanism for modulating posterior distribution variance, setting it as a practical and effective tool for uncertainty quantification in reservoir forecasting.

1.6 Methodology overview

The approach is evaluated against the original DSI-ESM DA on two synthetic reservoirs and one real field application using measurements from producers and injectors. The synthetic reservoirs are the Egg (Jansen *et al.*, 2014) and Olympus (Fonseca *et al.*, 2020) ensembles, which are widely used in the literature to benchmark history matching and uncertainty quantification methods, and the real field application is a mature oil field in the Brazilian Post-Salt. Posterior ensembles are compared across several metrics, including observation coverage and mean squared error (Emerick, 2025, subsection 9.4.2), to assess the impact of innovation gating on variance loss and uncertainty quantification. Results show that innovation-gated DSI-ESM DA reduces variance loss during and produces more robust uncertainty envelopes in the prediction phase without significantly degrading the data match. Furthermore, for moderate gating parameters, posterior forecasts remain close to those obtained with standard DSI-ESM DA, indicating that the proposed method acts as a stabilizing extension rather than a replacement.

1.7 Dissertation outline

Chapter 2 presents a review of the literature on reservoir uncertainty quantification, focusing on ensemble-based methods and data-space inversion. Chapter 3 introduces the theoretical framework underpinning the proposed methodology, from the relevant elements of reservoir engineering to the mathematical formulation of DSI-ESM DA and the innovation gating mechanism. Chapter 4 details the methods used to evaluate the proposed approach, including descriptions of the synthetic and real reservoir case studies, data collection, and analysis techniques. Chapter 5 presents the results of applying innovation-gated DSI-ESM DA to the case studies, including quantitative and qualitative assessments of performance relative to standard DSI-ESM DA. Chapter 6 discusses the implications of the findings and their relationship to the existing literature. Chapter 7 summarizes the main contributions and provides final remarks on the research and future work.

2 Literature review

2.1 Chapter outline

Aims and structure

This section provides a brief literature review of these methods, which are split in two major groups: model-space methods ([section 2.2](#)) and data-space methods ([section 2.3](#)). Details are kept as short as possible, specially in [section 2.2](#), and mathematical formulations omitted.

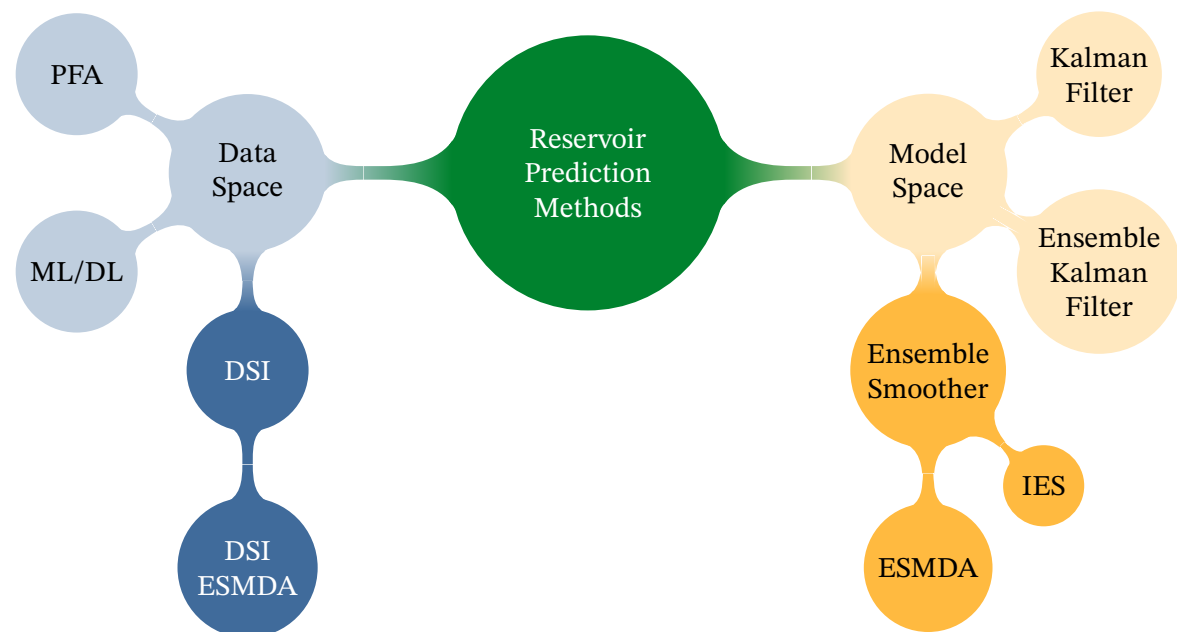
Most reservoir prediction methods are conceptually linked to the Kalman Filter (KF) ([Kalman, 1960](#)) as they attempt to, through the Bayesian theorem, improve a probability distribution given observations of the reservoir using linear update(s). The distribution *a priori* to data assimilation is called the prior distribution, and represents the belief one holds for the expected distribution. Conversely, the distribution *a posteriori* to data assimilation is called the posterior distribution.

The main difference between a model-space method and a data-space method is the distribution being updated. Model-space methods apply the Kalman-like update to an ensemble of model realizations, targeting distributions of model parameters. Subsurface simulations of the adjusted models are then performed to obtain predictions on reservoir states and flow rates. Data-space methods apply the Kalman-like update to an ensemble of predictions, targeting distributions of reservoir states and flow rates directly. Compared to model-space methods, the number of subsurface simulations is greatly reduced (only being required to establish the initial distribution), but do not offer insights to causes to reservoir behavior.

At the extreme, machine learning (ML) and deep learning (DL) methods can be applied directly to the ensemble of predictions, resulting in a posterior distribution not conceptually related to the Kalman update. Among model-space approaches, DL is also becoming increasingly popular due to its ability to estimate a model's simulation outputs, effectively proxying the simulator and consequently speeding up the history matching process (e.g., [Xiao et al. \(2024\)](#), [Tao et al. \(2025\)](#)).

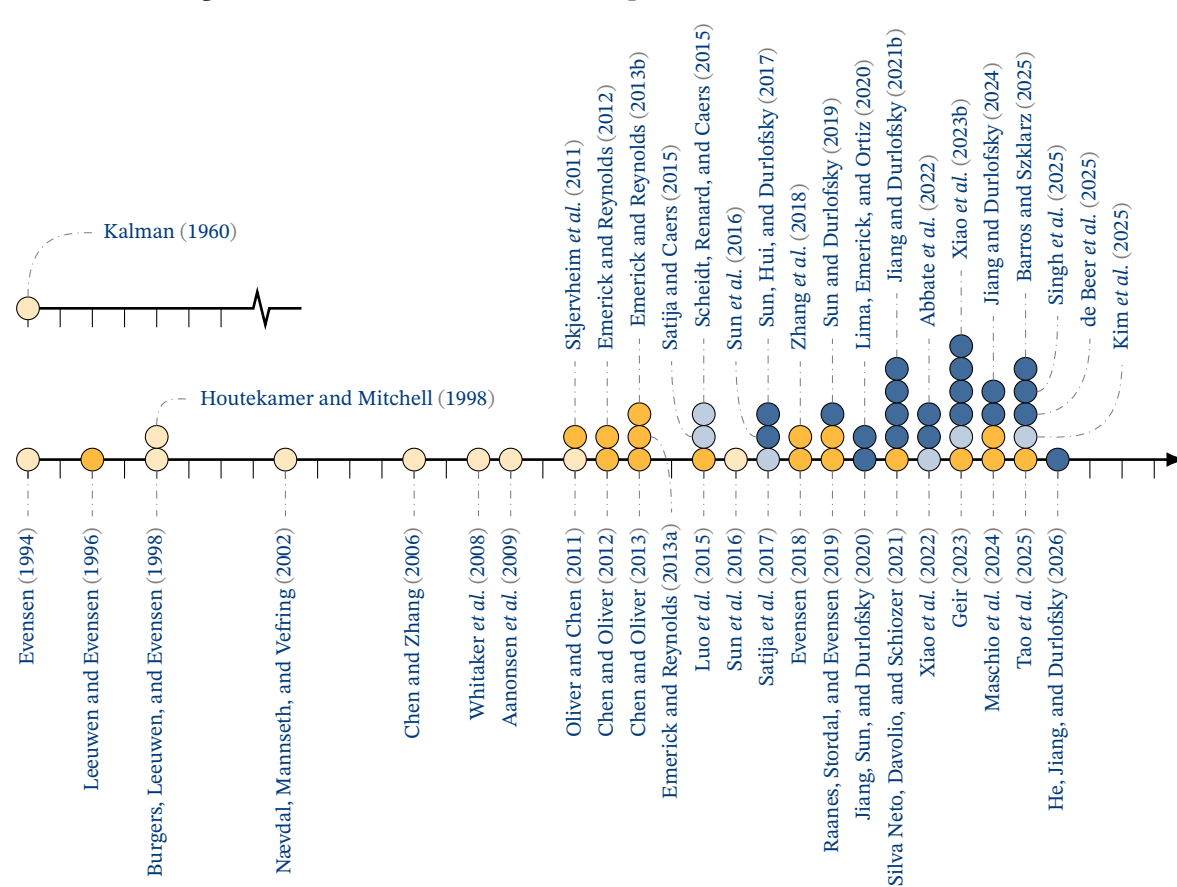
Ultimately, a reservoir prediction method is concerned with bringing the posterior of predictions as close as possible to reservoir observations without underestimating uncertainty. [Figure 2.1](#) mindmaps the main approaches mentioned in this section, and [figure 2.2](#) timelines most of the works discussed here.

Figure 2.1 – Mindmap of reservoir prediction methods



Source: Self-elaboration

Figure 2.2 – Timeline of main works presented in the literature review



Source: Self-elaboration

2.2 Model-space methods

2.2.1 Ensemble Kalman Filter

The cradle of modern model-space methods is the Ensemble Kalman Filter (EnKF) introduced by Evensen (1994), which in turn is an adaptation of KF.

KF is an optimal estimator for variables of a linear system (Emerick, 2025, section 4.1) that updates its estimation iteratively with each data assimilation. The result of a data assimilation step is the mean and covariance of the target variables posterior distribution. Its update equation, which can be derived from Bayes' theorem, consists of adding a correction to the mean. This correction is proportional to the difference between observed data and the prior distribution, commonly referred to as innovation. The proportionality constant, referred to as the Kalman gain, depends on the noise in observations as well as the prior's covariance¹.

In the context of history matching, the system is a reservoir, the variables being estimated are model parameters, and the system outputs are reservoir states and flow rates. KF can't be directly applied to the inverse problem on account of (i) the nonlinear relationship between system variables and outputs, and (ii) the size of the posterior covariance (trillions of entries if finer gridblocks are used)².

The first limitation can be addressed by running simulations of the reservoir model, substituting the linear relationship with a nonlinear function. EnKF addresses the second limitations by deploying an ensemble of model realizations. Instead of updating a single vector of model parameters, EnKF updates each parameter vector in the ensemble, approximating the prior covariance with the prior ensemble. Since the number of realizations is usually much smaller than the number of parameters (hundreds versus millions), this approximation is computationally feasible.

More precisely, EnKF estimates model parameters and reservoir states simultaneously, and its current formulation also applies perturbations to innovations in order to avoid underestimating uncertainty (Burgers; Leeuwen; Evensen, 1998; Houtekamer; Mitchell, 1998). While EnKF has been an influential tool for reservoir prediction (Aanonsen *et al.*, 2009; Oliver; Chen, 2011), specially after the work of Nævdal, Mannseth, and Vefring (2002), its framework is general and has been successfully applied in many fields, such as oceanography (Bertino; Evensen; Wackernagel, 2003; Keppenne; Rienecker, 2003), atmospheric modeling (Whitaker *et al.*, 2008; Houtekamer; Zhang, 2016), numerical weather prediction (Houtekamer *et al.*, 2005; Houtekamer; Mitchell, 2005; Szunyogh *et al.*, 2005; Otsuka; Kotsuki; Miyoshi, 2016), and hydrology (Reichle; McLaughlin; Entekhabi, 2002; Chen; Zhang,

¹ The posterior of an assimilation becomes the prior for the next assimilation, updating the Kalman gain.

² Knowledge on measurement noise is also a limitation of KF, albeit not particular to reservoir prediction.

2006; Liu; Chen; Zhang, 2008; Sun *et al.*, 2016).

2.2.2 Iterative Ensemble Smoother

In the EnKF framework, data assimilations occur one time step at a time, a process referred to as Sequential Data Assimilation (SDA). For systems with complicated relationships between estimated variables and system outputs, SDA means spending a great deal of time extracting these outputs.

This is case of geological reservoirs. Furthermore, EnKF estimates reservoir states and model parameters jointly, and might produce inconsistent reservoir states estimates. The Ensemble Smoother (ES) tackles these limitations by substituting SDA with a single data assimilation that considers all data points at once to obtain a posterior of model parameters.

ES was introduced by [Leeuwen and Evensen \(1996\)](#), and [Skjervheim *et al.* \(2011\)](#) were the first to apply it to reservoir models. Implementation wise, ES is effectively a single EnKF update on each model realization that uses all available observations at once.

Their success spurred researchers to develop similarly batching assimilations. Particularly, it was seen that the single ES update often produced data matches of insufficient quality, which lead to the development of iterative formulations of ES (IES). IES methods run the model update rule until convergence conditions have been satisfied, or a given number of iterations has been met. It is important to note that while IES iterations do require reservoir simulations, the number of iterations is usually much smaller than the number of time steps, which effectively guarantees computational time will be smaller than EnKF approaches.

Ensemble RML ([Chen; Oliver, 2012](#)), and its subsequent Levenberg-Marquardt ([Chen; Oliver, 2013](#)) were some of the first IES algorithms to be formulated. Since then, IES has remained an active research and study topic in history matching literature ([Luo *et al.*, 2015](#); [Zhang *et al.*, 2018](#); [Raanes; Stordal; Evensen, 2019](#); [Maschio *et al.*, 2024](#)). For more details on IES methods, the reader is referred to [Evensen \(2018\)](#), [Geir \(2023\)](#).

2.2.3 Ensemble Smoother with Multiple Data Assimilation

Of the proposed IES methods, the most influential is perhaps the Ensemble Smoother with Multiple Data Assimilation (ESMDA) introduced by [Emerick and Reynolds \(2013a\)](#), [Emerick and Reynolds \(2013b\)](#).

The Multiple Data Assimilation (MDA) mechanism was first developed to improve EnKF data assimilations ([Emerick; Reynolds, 2012](#)). In a word, MDA repeats the data assimilation step, but with perturbed observations. The covariance of the perturbation is inflated so that the greater the number of repetitions, the greater the covariance³

³ As such, ESMDA with a single MDA iteration reduces to ES.

By doing so, MDA improves the conditioning to data while usually avoiding underestimating uncertainty, specially in nonlinear model-data relationships. Therefore, ESMDA enjoys the computational speed of ES while also offering a straightforward mechanism through which one can mitigate the drawbacks of ES. However, MDA inflation coefficients and number of iterations must be defined in advance, and finding their optimal values remain open research topics (Emerick, 2025, subsection 5.3.5).

ESMDA's track record and simplicity of implementation make it a solid history matching algorithm that is amply employed both in industry and in research (Grana; Figueiredo; Azevedo, 2019; Silva Neto; Davolio; Schiozer, 2021; Tao *et al.*, 2025).

2.3 Data-space methods

2.3.1 Prediction Focused Analysis

Data-space methods for reservoir prediction are relatively new in literature, but have been gaining momentum with the Artificial Intelligence boom. The first algorithm proposed in this context capable of uncertainty quantification is due to Scheidt, Renard, and Caers (2015), who referred to the proposed method as Prediction Focused Analysis (PFA).

PFA uses low-dimensional representations of the ensemble prior in order to approximate conditional probability distributions to field observations. The approximated distribution is then projected back to the original dimension, allowing for predictions with quantification of uncertainty.

Since projections come from a bijective, smooth, nonlinear mapping, points close to the observations remain close after dimensionality reduction. This allows PFA to also serve as a tool for diagnosing ensemble quality, given that ensembles representative of data would clutter more around observed data in the low-dimension space.

However, PFA scales poorly with data dimensionality, and while changes would be made to the mapping process (Satija; Caers, 2015; Satija *et al.*, 2017), the method saw limited application in reservoir prediction.

2.3.2 Data Space Inversion

Sun and Durlafsky (2017b) introduce the Data-Space Inversion (DSI) algorithm, the first data-space method to allow for uncertainty quantification while also scaling reasonably with data dimension.

Similarly to PFA, DSI obtains a posterior distribution of data by sampling from a lower-dimensional space. While the mapping remains nonlinear, smooth and bijective, the conceptual difference from PFA is that the parameterization vector (or latent vector) is a random vector with fixed probability distribution.

For each realization in the prior ensemble, this vector is sampled and the mapping inverted in order to project the posterior back to the original data space. The sampling is defined by an optimization step known as Randomized Maximum Likelihood (RML).

DSI was shown to work adequately with reservoir forecasting since its conception (Sun; Durlofsky, 2017b; Sun; Hui; Durlofsky, 2017), and was also applied to carbon storage shortly after (Sun; Durlofsky, 2019).

By extending the prior vectors with well controls, Jiang, Sun, and Durlofsky (2020) managed to adapt the DSI framework beyond prediction and into closed-loop control. This approach, named Data-Space Inversion with Variable Controls (DSIVC), would be revisited by Liu *et al.* (2023), who coupled the connection element method (CEM) in order to improve subsurface simulation and models with complex geological formation.

Liu *et al.* (2021) improve DSI by solving the RML optimization with a stochastic approximation of the objective function gradient's, which lead to excellent predictions of reservoir states. Zhan *et al.* (2024) would later employ the version of DSI that uses this stochastic approach to RML as a proxy model to their proposed workflow of history matching horizontal wells.

de Beer *et al.* (2025) decomposed the DSI mapping in a way that makes sampling the posterior direct, avoiding the RML optimization. Their work has shown promising results, but lack applications on non-synthetic reservoirs.

2.3.3 Data Space Inversion with ESMDA

DSI's biggest bottleneck is the sampling of the latent variable, which is usually done with RML. In a method referred to as Data Space Inversion with ESMDA (DSI-ESMDA)⁴, Lima, Emerick, and Ortiz (2020) apply the ESMDA framework directly to data space, doing away with the requirements of RML. Since MDA iterations are much faster to compute, DSI-ESMDA enjoys faster performance than DSI. DSI-ESMDA's posteriors were seen to be competitive against, if not better than, the posteriors given by DSI and ESMDA. The use of a model-space framework applied directly to data space had profound effects on data-space methods.

Shortly after, Jiang and Durlofsky (2021a) would reformulate the original DSI approach in two fronts: first, DL algorithms would be used to compute the prior ensemble of latent variables; second, the ensemble of latent variables would be updated using ESMDA, or more precisely, the ESMDA rule applied to data space. The reformulated DSI, referred in this work as DSI with DL, presented good results on synthetic case studies, and was later put to the test on a complex real field (Jiang; Hui; Durlofsky, 2021), with good results. Jiang and Durlofsky

⁴ In this abbreviation, DSI refers to the fact that the method acts directly on the data space. DSI-ESMDA and the DSI proposed by (Sun; Durlofsky, 2017b) are not conceptually linked.

(2021b) introduced a new ensemble of latent variables to quantify model uncertainty in the distribution of data, presenting excellent results in synthetic reservoir case studies.

Jiang and Durlafsky (2024) refined the DL procedure used for finding the latent variables in DSI with DL, lending the method better for forecasting of reservoir states (e.g., field's pressure and saturation). He, Jiang, and Durlafsky (2026) expands the range of applications of DSI by deploying the aforementioned refined framework to forecast pressure, strain and stress fields in carbon storage problems.

In an approach similar to DSIVC, Szklarz and Barros (2022) apply DSI with DL on two synthetic problems: first, forecasting of carbon storage; second, optimization of water injection rates of a hydrocarbon reservoir. The authors would later broaden the application of DSI with DL to optimize well placement in a realistic hydrocarbon reservoir (Barros; Szklarz, 2025). Parallely, Singh *et al.* (2025) apply DSI with DL to optimize water injection rates of an oil field.

Abbate *et al.* (2022) use DSI-ESMDA as a basis for proposing a new method in forecasting carbon storage. The authors develop an IES tailor-made for this problem, and results were comparable to standard model-space approaches.

Lin *et al.* (2023) compare the performance of DSI-ESMDA against the original DSI algorithm and a logarithmic fit to data (Duong, 2010) in the forecast of two shale reservoir case studies. DSI-ESMDA presented both high robustness and high computational efficiency when compared to the other methods, validating the algorithm.

Some of these authors would move on to swap the Kalman-like gain in the DSI-ESMDA update for a DL algorithm, making the update nonlinear with respect to the innovation vector Xiao *et al.* (2023b)⁵. The proposed reformulation was tested on a few synthetic case studies of carbon storage problems, with good results.

Xiao *et al.* (2023a) expanded their approach to develop a shale reservoir control strategy. The proposed approach, named DSI with varying refracturing (DSIVRF), uses a DSI-ESMDA-like rule to obtain a posterior ensemble of reservoir predictions, and optimizes controls based on estimates of reservoir pressure. To avoid subsurface simulations, a DL algorithm is developed to approximate pressure distributions.

2.3.4 Machine Learning and Deep Learning

By now, it should have been clear to the reader that data-space is fertile soil for the development and application of ML and DL methods. While this review has so far presented them as auxiliary tools for DSI and DSI-ESMDA, the direct application of ML/DL algorithms is also possible. In this case, the posterior is not obtained via an iterative update, as was the

⁵ Confusingly enough, authors also called this approach “DSI with DL”.

case with DSI-ESMDA, nor does it require RML sampling, as was the case with the original DSI, but via the inference process of a deep neural network.

Detailing ML/DL algorithms is beyond the scope of this dissertation, but for examples of works in this line, the reader is referred to [Xiao *et al.* \(2022\)](#), [Xiao *et al.* \(2023c\)](#), and [Kim *et al.* \(2025\)](#).

2.4 Chapter summary

Key takeaways

This chapter has presented an overview of model-space and data-space methods for reservoir prediction, with a slight focus to data-space methods. Model-space methods are conceptually linked to the Kalman Filter, and developments in this area have historically focused on speeding up the data assimilation process. The state of the art in model-space methods are iterative ensemble smoothers, of which ESMDA is the most influential algorithm. Machine and deep learning algorithms are also often deployed as proxy models in this context. Data-space methods vary in their conceptual origins, but dimension-reduction tools are widely utilized. The state of the art in data-space methods consists of a synergy between deep learning algorithms for nonlinear estimation of an ensemble of latent variables, and ESMDA-like iterations to update this ensemble.

3 Theory framework

3.1 Chapter outline

Aims and structure

This chapter's main aim is to derive the ensemble update equations of DSI-ESMDA and the proposed Gated DSI-ESMDA. This requires, first and foremost, carrying out extensive groundwork on the notation and statistical framework underpinning DSI-ESMDA. [Section 3.2](#) defines the meaning of a forward model, a central concept in reservoir prediction, and introduces notation on data vectors. [Section 3.3](#) builds the statistical framework behind DSI-ESMDA and from it derives the ensemble update equation. [Section 3.4](#) presents and discusses the Gated DSI-ESMDA ensemble update equation, capping off the chapter.

3.2 Prediction forward model

The aims of this section are twofold: to introduce the concept of prediction forward model and other related concepts; to introducing notation that will be assumed known after this section.

As discussed in [subsection 1.2.3](#), the inverse problem is part of the process in obtaining forecasts. Denoting by $\mathbf{m}$ the vector of model parameters, and by $\mathbf{u}$ the set of controls applied on the field, and by $\mathbf{d}$ the reservoir production within the time frame of the controls, one can write

$$\mathbf{d} = \mathbf{g}(\mathbf{m}, \mathbf{u}; \mathbf{x}_0) \quad (3.1)$$

with $\mathbf{x}_0$ the initial value of reservoir states. Since $\mathbf{x}_0$ is constant, it is omitted in the remainder of this work for simplification.

Function $\mathbf{g}$ is called the *prediction forward model*, or simply forward model ([Emerick, 2025](#), section 1.4). The forward model is the underlying mapping between a model's parameters, controls and initial states to its observable features. In this work, the features employed are oil and water production rates (OPR and WPR); water injection rates (WIR); bottom-hole pressures (BHP), i.e., the pressure at a well's completion point; water cut percentages (WCTP), i.e., the volume of water relative to the volume of fluid extracted by a producer well, and; field oil production total (FOPT), i.e., the total volume of oil extracted from the field within the given time period.

The key assumption behind the forward model is that it is deterministic. Implementation wise, $\mathbf{g}$ consists of evaluating a set of equations at each gridblock of the model. Each

evaluation evolves the model a little further in time, until the scheduled time period is appreciated, concluding the process. For simplicity, the described process will be referred to as a simulation of the reservoir.

In truth, no reservoir simulator can be an unerrant representation of underlying reservoir dynamics, unless the reservoir itself is extremely simple. Still, the mathematical equations that constitute the forward model are give reasonable approximations to these dynamics, so that data mismatches are mostly caused by model parameters unrepresentative of their true values.

With that said, simulators vary in which equations are used. Of the options available, the most commonly used are black-oil equations (Rasmussen *et al.*, 2021). Based on the premise that there are three different fluid phases (aqueous, oleic, and gaseous) and three (pseudo)components (water, oil, and gas), the equations can be deduced from conservation of mass for each component with suitable closure relationships. A given phase α 's component mass A_α satisfies the conservation equation

$$\frac{\partial}{\partial t} (\varphi_{\text{ref}} A_\alpha) + \nabla \cdot \mathbf{u}_\alpha + q_\alpha = 0, \quad (3.2)$$

where

$$A_w = m_\varphi b_w s_w, \quad \mathbf{u}_w = b_w \mathbf{v}_w, \quad (3.3a)$$

$$A_o = m_\varphi (b_o s_o + r_{og} b_g s_g), \quad \mathbf{u}_o = b_o \mathbf{v}_o + r_{og} b_g \mathbf{v}_g, \quad (3.3b)$$

$$A_g = m_\varphi (b_g s_g + r_{go} b_o s_o), \quad \mathbf{u}_g = b_g \mathbf{v}_g + r_{go} b_o \mathbf{v}_o. \quad (3.3c)$$

and the phase fluxes are given by Darcy's law:

$$\mathbf{v}_\alpha = -\lambda_\alpha \mathbf{K} (\nabla p_\alpha - \rho_\alpha \mathbf{g}). \quad (3.4)$$

Finally, the following closure equations are imposed:

$$s_w + s_o + s_g = 1, \quad (3.5)$$

$$p_{c,ow} = p_o - p_w, \quad (3.6)$$

$$p_{c,og} = p_o - p_g. \quad (3.7)$$

The description of each symbol is given below (Rasmussen *et al.*, 2021, adapted).

m_φ pore volume multiplier as function of pressure.

φ_{ref} reference porosity, a constant in time but varying in space.

p_α phase pressure for phase α .

b_α shrinkage/expansion factor for phase α defined as the ratio of surface volume at standard conditions to reservoir volume for a given amount of fluid: $b_\alpha = V_{\text{surface},\alpha} / V_{\text{reservoir},\alpha}$. For the oil phase, b_o is called shrinkage factor, whereas b_g is called expansion factor.

r_{go} ratio of dissolved gas to oil in the oleic phase.

r_{og} ratio of vaporized oil to gas in the gaseous phase.

s_α saturation of phase α , the pore volume fraction occupied by the phase. The saturations of all phases sum to 1.

$p_{c,\alpha\beta}$ capillary pressure between phases α and β .

K permeability of the porous medium.

$k_{r,\alpha}$ relative permeability for phase α , modeling the reduction in effective permeability for a fluid phase in the presence of other phases.

μ_α viscosity of phase α , typically a function of phase pressure and composition.

λ_α mobility of phase α , given by $\lambda_\alpha = k_{r,\alpha}/\mu_\alpha$.

$\mathbf{v}_\alpha$ phase velocity of phase α .

$\mathbf{u}_\alpha$ component velocity of component α .

$\rho_{S,\alpha}$ surface density of phase α at one atmosphere, a given constant.

ρ_α density of phase α in the reservoir. For water, $\rho_w = b_w \rho_{S,w}$. For the oleic phase the relationship is more complex since it must include dissolved gas: $\rho_o = b_o(\rho_{S,o} + r_{go}\rho_{S,g})$, similar for the gaseous phase: $\rho_g = b_g(\rho_{S,g} + r_{og}\rho_{S,o})$.

g gravitational acceleration vector.

q_α well outflux density of pseudo component α , (negative for well inflows).

3.3 DSI-ESMDA

3.3.1 Key concepts and definitions

This subsection presents the key mathematical objects and concepts used by both the standard DSI-ESMDA and the proposed Gated DSI-ESMDA, many of which are related to the forward model presented in [section 3.2](#). Important nuances, such as the distinction between prediction and forecasting, are also pointed here.

3.3.1.1 Predicting versus forecasting

Prediction and forecasting, usually interchangeable terms, have an important distinction in this work. Prediction refers to the estimation of any datum at a given time, while forecasting refers to the estimation of any datum beyond the observation period. Forecasting is a type of prediction, but in particular prediction of new data.

In Data Science literature, a common way of estimating forecast quality is splitting available data into a training partition and a test partition. The training partition is used to freely adjust predictions within a given period, and the test, to verify how close predictions will be on new data.

In the context of this work, data from a given reference will be split into two disjointed periods in time: history period, and validation period, mirroring the common train and test splits respectively.

3.3.1.2 Data vectors

Let $\mathbf{d}$ denote the true production data vector of a reference reservoir across time, free of observation error. As discussed above, data is separated into a history period and a validation period, so that $\mathbf{d} = \begin{bmatrix} \mathbf{d}_h^\top & \mathbf{d}_v^\top \end{bmatrix}^\top$ with $\mathbf{d}_h$ the ground-truth history data, and $\mathbf{d}_v$ the ground-truth validation data. Analogously, let $\mathbf{u}$ and $\mathbf{m}$ be the controls applied to the ground-truth and the parameters of the ground-truth, respectively. Denoting the prediction forward model by $\mathbf{g}(\cdot, \cdot)$, ground-truth data is given by $\mathbf{d} = \mathbf{g}(\mathbf{m}, \mathbf{u})$.

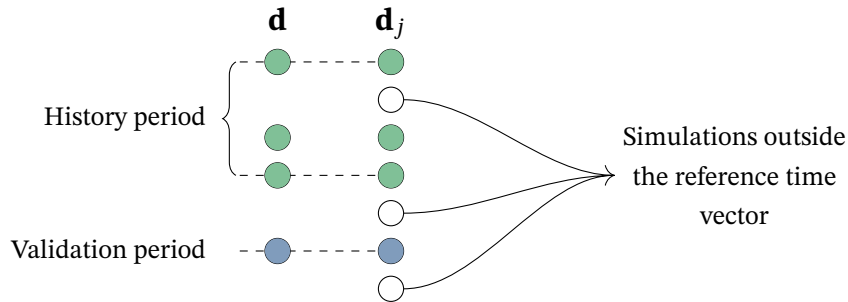
Let $\mathbf{d}_j$ denote the production data vector produced by running the forward model with parameters $\mathbf{m}_j$ from the start of the history period up to (at least) the end of the validation period, i.e., $\mathbf{d}_j = \mathbf{g}(\mathbf{m}_j, \mathbf{u})$. For an ensemble of N_e realizations, there are N_e different such vectors, each of which associated to its own parameters vector.

Following the pattern established in the ground-truth, $\mathbf{d}_{h,j}$ denotes the history period data output by realization j . Assume $\mathbf{d}_h$ is corrupted by additive measurement noise $\mathbf{e}_d \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_e)$, where $\mathbf{C}_e$ is the noise covariance matrix. The corrupted vector is the observed data during the history period, denoted $\mathbf{d}_{\text{obs}}$. Figure 3.2 summarizes the concepts and notations in this section.

Note 3.1: Time vectors

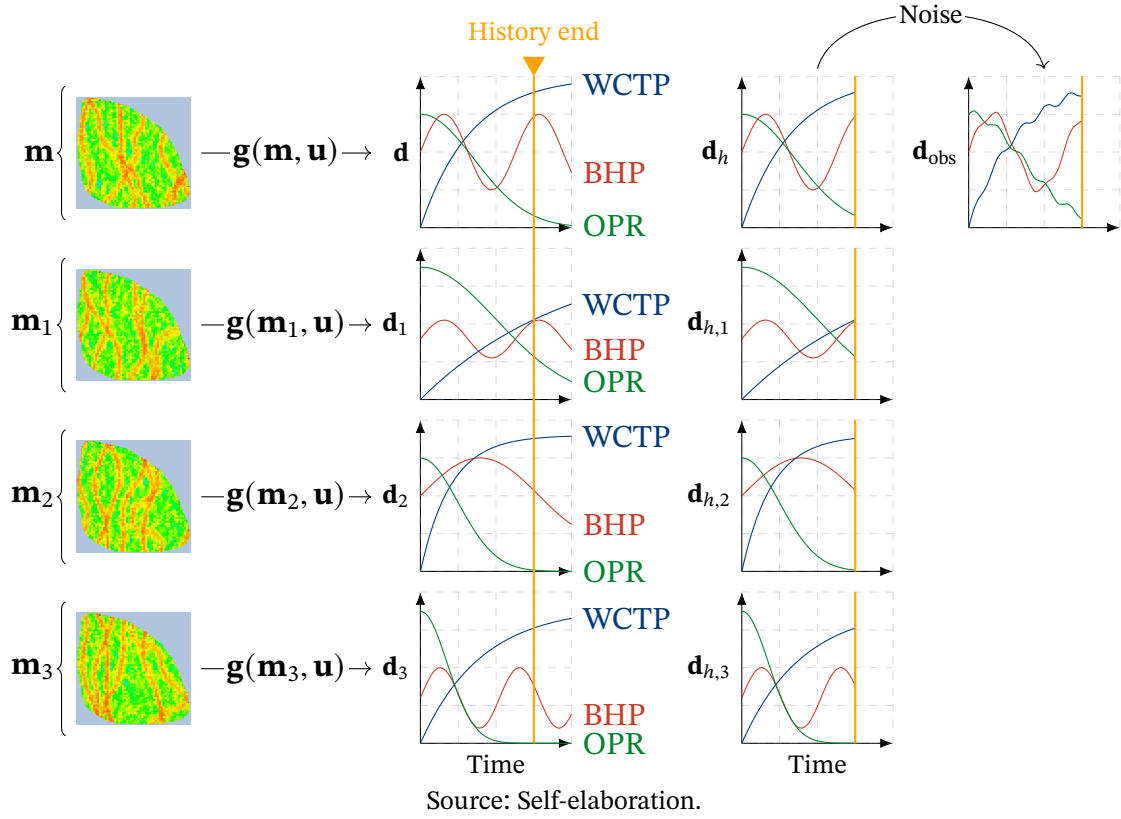
Note that all realization data vectors are implicitly assumed to have a synchronized time vector. Ultimately, this means simulation scheduling was the same for all realizations. Reference data, however, is under no obligation to follow this scheduling. As such, in order for equation (3.44) to work, the history period of the realization data is constrained to match the observed data's. Figure 3.1 illustrates this process for data vectors with a single feature (e.g., field oil production total).

Figure 3.1 – Synchronization between realization data vector and reference data.



Source: Self-elaboration.

Figure 3.2 – Relationship between model parameters and data vectors.



3.3.2 Statistical Framework

This subsection details the statistical frameworks underpinning the DSI-ESMDA algorithm, and expects the reader to hold knowledge of probability theory, namely: the axioms of probability; what is a conditional probability; statement of Bayes' theorem. what is a probability distribution; formulation of Gaussian probability density function. Subsection 4.2.4 also assume knowledge of cumulative distribution function. Emerick (2025, appendix C) provides a brief overview of all these topics.

DSI-ESMDA can be conceptualized as a Bayesian approach to the forecasting problem, aiming to estimate the underlying conditional probability distribution of predictions given the reference data. For given data, the algorithm returns samples from this posterior distribution, outputting a range of values rather than a single value. While this makes evaluating the algorithm less straightforward, it allows for quantification of uncertainty.

Let $p(\mathbf{d}^* | \mathbf{d}_{obs})$ be the conditional *a posteriori* probability distribution of true reservoir production being $\mathbf{d}^*$ given observations $\mathbf{d}_{obs}$. Note that, as with ground-truth production $\mathbf{d}$, $\mathbf{d}^*$ can also be split into history ($\mathbf{d}_h^*$) and validation ($\mathbf{d}_v^*$) periods. From Bayes' theorem, one can write the posterior as

$$p(\mathbf{d}^* | \mathbf{d}_{obs}) = c_{pos} \mathcal{L}(\mathbf{d}_{obs} | \mathbf{d}^*) p(\mathbf{d}^*) \quad (3.8)$$

where c_{pos} is, for the purposes of this work, a normalization constant; $\mathcal{L}(\mathbf{d}_{\text{obs}} | \mathbf{d}^*)$, the likelihood function¹, is the probability of $\mathbf{d}_{\text{obs}}$ being the true reservoir production if $\mathbf{d}^*$ was observed, and; $p(\mathbf{d}^*)$, the prior, is the probability of observing $\mathbf{d}^*$ before observations $\mathbf{d}_{\text{obs}}$ were made.

Maximizing [equation \(3.8\)](#) means finding reservoir production that is best suited to match field observations tempered by prior beliefs on the behavior of data. While this approach is statistically sensible, in practice a single maximization tends to produce poor matches to data. This situation could be mitigated by assimilating more data, since then the posterior would be adjusted Bayes' theorem, $p(\mathbf{d}^* | \mathbf{d}_{\text{obs}}^{\text{new}}, \mathbf{d}_{\text{obs}}) = c_{\text{pos}}^{\text{new}} \mathcal{L}(\mathbf{d}_{\text{obs}}^{\text{new}} | \mathbf{d}^*) p(\mathbf{d}^* | \mathbf{d}_{\text{obs}})$, but it is assumed that $\mathbf{d}_{\text{obs}}$ already holds all the data available for analysis.

Therefore, in order to improve data match, the effect of posterior updating is obtained by successively assimilating disturbed versions of observed data. This is called Multiple Data Assimilation (MDA), contrasting with the previous Single Data Assimilation (SDA). Denoting the number of assimilations by N_a , assimilation $1 \leq \ell \leq N_a$ with disturbed observations $\mathbf{d}_{\text{obs}}^\ell$ produces posterior p^ℓ such that²

$$p^1(\mathbf{d}^* | \mathbf{d}_{\text{obs}}^1) = c_{\text{pos}}^1 \mathcal{L}(\mathbf{d}_{\text{obs}}^1 | \mathbf{d}^*) p(\mathbf{d}^*) \quad (3.9)$$

$$p^2(\mathbf{d}^* | \mathbf{d}_{\text{obs}}^2, \mathbf{d}_{\text{obs}}^1) = c_{\text{pos}}^2 \mathcal{L}(\mathbf{d}_{\text{obs}}^2 | \mathbf{d}^*) p^1(\mathbf{d}^* | \mathbf{d}_{\text{obs}}^1) \quad (3.10)$$

$$\vdots \quad (3.11)$$

$$p^{N_a}(\mathbf{d}^* | \mathbf{d}_{\text{obs}}^{N_a}, \dots, \mathbf{d}_{\text{obs}}^1) = c_{\text{pos}, \mathcal{L}_{\text{pos}}}^{N_a} \mathcal{L}(\mathbf{d}_{\text{obs}}^{N_a} | \mathbf{d}^*) p^{N_a-1}(\mathbf{d}^* | \mathbf{d}_{\text{obs}}^{N_a-1}, \dots, \mathbf{d}_{\text{obs}}^1) \quad (3.12)$$

or

$$p^{N_a}(\mathbf{d}^* | \mathbf{d}_{\text{obs}}^{N_a}, \dots, \mathbf{d}_{\text{obs}}^1) = p(\mathbf{d}^*) \prod_{\ell=1}^{N_a} c_{\text{pos}}^\ell \mathcal{L}(\mathbf{d}_{\text{obs}}^\ell | \mathbf{d}^*). \quad (3.13)$$

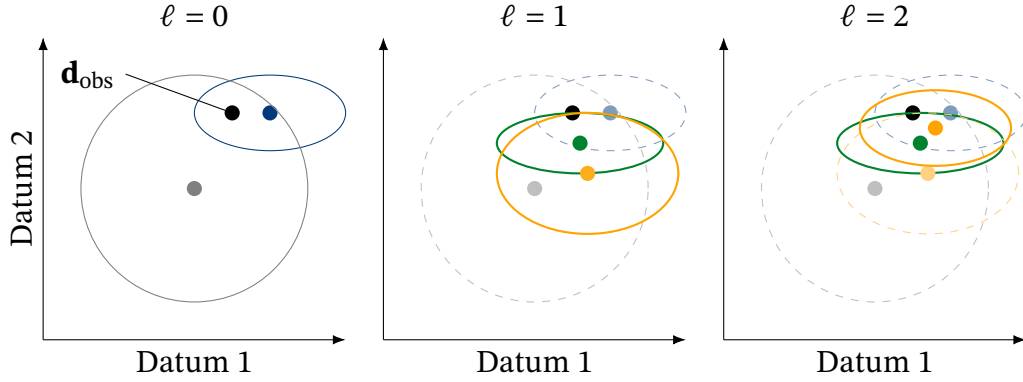
By successively updating the posterior with disturbed data, the posterior's maximum is also successively updated. If maximizing [equation \(3.8\)](#) was akin to trying to reach optimum p^* in a single fixed step, maximizing [equation \(3.13\)](#) is akin to reaching p^* with multiple steps, as illustrated in [figure 3.3](#). The number of steps is exactly equal to N_a , and their size dependent on the perturbations done to the observed data.

While SDA and MDA can be made to be equivalent if likelihoods are Gaussian (more details in [subsection 3.3.3](#)), in general MDA produces better matches to data than SDA. This is because MDA attempts to reduce variance in the posterior distribution without underestimating uncertainty in observations, which often results in tighter posteriors whose means are closer aligned to observations. Furthermore, the iterative search of MDA for a maximum better accommodates for underlying likelihoods that are not Gaussian.

¹ Note that the likelihood is denoted by a $\mathcal{L}$ instead of the usual p . This is deliberate, since likelihoods are not necessarily probability distributions, i.e., they need not sum up to 1.

² Writing a conditional distribution as the product of simpler conditional distributions is sometimes called the chain rule of conditional probabilities ([Emerick, 2025](#), subsection C.1.8).

Figure 3.3 – Posterior distributions obtained with SDA (green) and MDA (yellow) for an observation vector with 2 points of data. Circles denotes means, and circumferences, covariances. True underlying posterior is denoted in blue, and the prior in gray.



3.3.3 Deriving the update equation

This subsection derives the base update equation used by DSI-ESMDA (and by extension, Gated DSI-ESMDA) by extracting a practical implementation from the general, more abstract statistical framework outlined in [subsection 3.3.2](#). Chiefly, this process requires making assumptions on prior and likelihoods and then using model ensembles to sample from the distributions. It will be assumed that the reader is familiar with linear algebra manipulations, and the gradient operation ∇ .

In the context of reservoir engineering, it is standard to assume that data (and model parameters, for that matter) have underlying Gaussian distributions. In this case, the general likelihood in [equation \(3.8\)](#) becomes

$$\mathcal{L}(\mathbf{d}_{\text{obs}} | \mathbf{d}^*) = c_{\mathcal{L}} \exp(-O_{\mathcal{L}}(\mathbf{d}_{\text{obs}}, \mathbf{d}^*)) \quad (3.14)$$

where $O_{\mathcal{L}}$ is defined as

$$O_{\mathcal{L}}(\mathbf{d}_{\text{obs}}, \mathbf{d}) = \frac{1}{2}(\mathbf{d}_{\text{obs}} - \mathbf{H}\mathbf{d}^*)^{\top} \mathbf{C}_e^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{H}\mathbf{d}^*) \quad (3.15)$$

$$= \frac{1}{2}(\mathbf{d}_{\text{obs}} - \mathbf{d}_h)^{\top} \mathbf{C}_e^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{d}_h), \quad (3.16)$$

with $\mathbf{C}_e$ the measurement noise in observed data and $\mathbf{H}$ a matrix that extracts $\mathbf{d}_h^*$ from $\mathbf{d}^*$. The associated posterior can then be rewritten as

$$p(\mathbf{d}^* | \mathbf{d}_{\text{obs}}) = c_{\text{pos}} c_{\mathcal{L}} p(\mathbf{d}^*) \exp(-O_{\mathcal{L}}(\mathbf{d}_{\text{obs}}, \mathbf{d}^*)). \quad (3.17)$$

Similarly, the likelihood at assimilation l in [equation \(3.13\)](#) becomes

$$\mathcal{L}^{\ell}(\mathbf{d}_{\text{obs}}^{\ell} | \mathbf{d}_h^*) = c_{\mathcal{L}}^{\ell} \exp(-O_{\mathcal{L}}^{\ell}(\mathbf{d}_{\text{obs}}^{\ell}, \mathbf{d}_h^*)) \quad (3.18)$$

where $O_{\mathcal{L}}^{\ell}$ is defined as

$$O_{\mathcal{L}}^{\ell}(\mathbf{d}_{\text{obs}}^{\ell}, \mathbf{d}^*) = \frac{1}{2}(\mathbf{d}_{\text{obs}}^{\ell} - \mathbf{H}\mathbf{d}^*)^{\top} (\mathbf{C}_e^{\ell})^{-1}(\mathbf{d}_{\text{obs}}^{\ell} - \mathbf{H}\mathbf{d}^*), \quad (3.19)$$

with $\mathbf{C}_e^\ell$ the measurement noise in the disturbed observed data. The associated posterior can then be rewritten as

$$p^{N_a}(\mathbf{d}^* | \mathbf{d}_{\text{obs}}^{N_a}, \dots, \mathbf{d}_{\text{obs}}^1) = p(\mathbf{d}) \prod_{l=1}^{N_a} c_{\text{pos}}^\ell \mathcal{L}(\mathbf{d}_{\text{obs}}^\ell | \mathbf{d}^*) \quad (3.20)$$

$$= p(\mathbf{d}) \prod_{l=1}^{N_a} c_{\text{pos}}^\ell c_{\mathcal{L}}^\ell \exp(-O_{\mathcal{L}}^\ell(\mathbf{d}_{\text{obs}}^\ell, \mathbf{d}^*)) \quad (3.21)$$

$$= c_{\text{pos}, \mathcal{L}}^{N_a} p(\mathbf{d}^*) \exp\left(-\sum_{\ell=1}^{N_a} O_{\mathcal{L}}^\ell(\mathbf{d}_{\text{obs}}^\ell, \mathbf{d}^*)\right) \quad (3.22)$$

where $c_{\text{pos}, \mathcal{L}}^{N_a}$ is a normalization constant that absorbed the intermediary posterior and likelihood normalization constants. Then, [equation \(3.22\)](#) is equivalent to [equation \(3.17\)](#) if and only if

$$c_{\text{pos}, \mathcal{L}}^{N_a} = c_{\text{pos}} c_{\mathcal{L}}, \quad (3.23a)$$

$$\sum_{l=1}^{N_a} (\mathbf{C}_e^\ell)^{-1} = \mathbf{C}_e^{-1}. \quad (3.23b)$$

Now, DSI-ESMDA models the covariance at assimilation l proportionately to $\mathbf{C}_e$:

$$\mathbf{C}_e^\ell = \alpha^\ell \mathbf{C}_e, \quad \alpha^\ell > 0, \quad (3.24)$$

so that [equation \(3.23b\)](#) reduces to

$$\sum_{l=1}^{N_a} \frac{1}{\alpha^\ell} = 1. \quad (3.25)$$

Constants α^ℓ parameterize the covariance in assimilation l , and by extension how much the intermediary posterior distribution should tighten around the data. Invoking the analogy at the end of [subsection 3.3.2](#) and the associated illustration in [figure 3.3](#), $\mathbf{C}_e$ influences the direction of the adjustment, facilitating corrections on less uncertain data, and α^ℓ influences the size of the adjustment, so that assimilations with larger α^ℓ incur smaller corrections.

The next step towards an implementable update equation is assuming the prior distribution is also Gaussian, i.e.,

$$p(\mathbf{d}^*) = c_{\text{prior}} \exp(-O_p(\mathbf{d}^*, \mathbf{d}_{\text{prior}})) \quad (3.26)$$

with c_{prior} a normalization constant and

$$O_p(\mathbf{d}^*, \mathbf{d}_{\text{prior}}) = \frac{1}{2}(\mathbf{d}^* - \mathbf{d}_{\text{prior}})^\top \mathbf{C}_p^{-1}(\mathbf{d}^* - \mathbf{d}_{\text{prior}}) \quad (3.27)$$

where $\mathbf{d}_{\text{prior}}$ is the prediction with highest belief of occurring prior to observing $\mathbf{d}_{\text{obs}}$, and $\mathbf{C}_p$ is the prior covariance matrix, i.e., it is assumed that $\mathbf{d}^* \sim \mathcal{N}(\mathbf{d}_{\text{prior}}, \mathbf{C}_p)$.

Substituting [equation \(3.26\)](#) in [equation \(3.22\)](#), it follows that the first assimilation ([equation \(3.9\)](#)) can be written as

$$p^1(\mathbf{d}^* | \mathbf{d}_{\text{obs}}^1) = c^1 \exp \left(-[O_p^1(\mathbf{d}^*, \mathbf{d}_{\text{prior}}^0) + O_{\mathcal{L}}^1(\mathbf{d}_{\text{obs}}^1, \mathbf{d}^*)] \right) \quad (3.28)$$

$$= c^1 \exp(-O(\mathbf{d}^*, \mathbf{d}_{\text{obs}}^1, \mathbf{d}_{\text{prior}}^0)) \quad (3.29)$$

where $c^1 = c_{\text{prior}} c_{\text{pos}}^1 c_{\mathcal{L}}^1$, $O(\mathbf{d}^*, \mathbf{d}_{\text{obs}}^1, \mathbf{d}_{\text{prior}}^0) = O_p(\mathbf{d}^*, \mathbf{d}_{\text{prior}}^0) + O_{\mathcal{L}}^1(\mathbf{d}_{\text{obs}}^1, \mathbf{d}^*)$, and $\mathbf{d}_{\text{prior}}^0$ is the mean of the prior distribution after 0 assimilations, i.e., before any assimilation was performed. O_p^0 is exactly equal to [equation \(3.26\)](#), but with superscript 1 denoting the use of $\mathbf{C}_p^0$, the covariance of $\mathbf{d}_{\text{prior}}^0$. The maximum *a posteriori* (MAP) of the first assimilation, denoted $\mathbf{d}_{\text{map}}^1$, is thus given by

$$\mathbf{d}_{\text{map}}^1 = \underset{\mathbf{d}^*}{\text{argmin}}(O(\mathbf{d}^*, \mathbf{d}_{\text{obs}}^1, \mathbf{d}_{\text{prior}}^0)) \quad (3.30)$$

which works out to

$$0 = \nabla_{\mathbf{d}^*} O(\mathbf{d}_{\text{map}}^1, \mathbf{d}_{\text{obs}}^1, \mathbf{d}_{\text{prior}}^0) \quad (3.31)$$

$$0 = (\mathbf{C}_p^0)^{-1}(\mathbf{d}_{\text{map}}^1 - \mathbf{d}_{\text{prior}}^0) - \mathbf{H}^\top (\mathbf{C}_e^1)^{-1}(\mathbf{d}_{\text{obs}}^1 - \mathbf{H}\mathbf{d}_{\text{map}}^1) \quad (3.32)$$

$$0 = (\mathbf{C}_p^0)^{-1}(\mathbf{d}_{\text{map}}^1 - \mathbf{d}_{\text{prior}}^0) - \mathbf{H}^\top (\mathbf{C}_e^1)^{-1}(\mathbf{d}_{\text{obs}}^1 - \mathbf{H}\mathbf{d}_{\text{prior}}^0 - \mathbf{H}[\mathbf{d}_{\text{map}}^1 - \mathbf{d}_{\text{prior}}^0]) \quad (3.33)$$

$$0 = (\mathbf{C}_p^0)^{-1}(\mathbf{d}_{\text{map}}^1 - \mathbf{d}_{\text{prior}}^0) - \mathbf{H}^\top (\mathbf{C}_e^1)^{-1}(\mathbf{d}_{\text{obs}}^1 - \mathbf{H}\mathbf{d}_{\text{prior}}^0) + \mathbf{H}^\top (\mathbf{C}_e^1)^{-1}\mathbf{H}(\mathbf{d}_{\text{map}}^1 - \mathbf{d}_{\text{prior}}^0) \quad (3.34)$$

or

$$((\mathbf{C}_p^0)^{-1} + \mathbf{H}^\top (\mathbf{C}_e^1)^{-1}\mathbf{H})(\mathbf{d}_{\text{map}}^1 - \mathbf{d}_{\text{prior}}^0) = \mathbf{H}^\top (\mathbf{C}_e^1)^{-1}(\mathbf{d}_{\text{obs}}^1 - \mathbf{d}_{h,\text{prior}}^0) \quad (3.35)$$

$$\mathbf{d}_{\text{map}}^1 = \mathbf{d}_{\text{prior}}^0 + \mathbf{K}^1(\mathbf{d}_{\text{obs}}^1 - \mathbf{d}_{h,\text{prior}}^0). \quad (3.36)$$

where

$$\mathbf{K}^1 = ((\mathbf{C}_p^0)^{-1} + \mathbf{H}^\top (\mathbf{C}_e^1)^{-1}\mathbf{H})^{-1}\mathbf{H}^\top (\mathbf{C}_e^1)^{-1}. \quad (3.37)$$

The general structure of [equation \(3.36\)](#), where a gain is applied to *a priori* prediction error (called the innovation, or, in this case, innovation vector) in order to improve data mismatch, is characteristic of Kalman filtering³. Hence, in this work, $\mathbf{K}$ is referred to as “Kalman-like gain”, and [equation \(3.36\)](#), as the “Kalman-like update”.

While [equation \(3.37\)](#) is a valid expression, it requires several matrix inversions, which can be computationally expensive. [Appendix B](#) shows that the Kalman gain is equivalent to

$$\mathbf{K}^1 = \mathbf{C}_p^0 \mathbf{H}^\top (\mathbf{H}(\mathbf{C}_p^0) \mathbf{H}^\top + \mathbf{C}_e^1)^{-1}. \quad (3.38)$$

Denoting the prior’s covariance in the history period by $\mathbf{C}_{hh,p}^0$, and the prior’s cross-covariance of the history period by $\mathbf{C}_{hd,p}^0$, it follows that

$$\mathbf{C}_{hh,p}^0 = \mathbf{H}(\mathbf{C}_p^0) \mathbf{H}^\top, \quad (3.39a)$$

$$\mathbf{C}_{dh,p}^0 = \mathbf{C}_p^0 \mathbf{H}^\top. \quad (3.39b)$$

³ Indeed, the Bayesian framework can be used in order to derive the standard Kalman filter equations.

In practice, $\mathbf{C}_p^0$ can't be known, and [equation \(3.39b\)](#) must be approximated by sampling the prior distribution. This is done through the prediction forward model of an ensemble of model realizations (hence the “Ensemble Smoother” in DSI-ESMDA). Denoting by $\mathbf{d}_j^0$ the prediction vector obtained by running the forward model on realization j , $1 \leq j \leq N_e$, the following approximations of [equation \(3.39b\)](#) are obtained:

$$\mathbf{C}_{hh,p}^0 \approx \Delta \mathbf{D}_h^0 \Delta \mathbf{D}_h^0, \quad (3.40a)$$

$$\mathbf{C}_{dh,p}^0 \approx \Delta \mathbf{D}^0 (\Delta \mathbf{D}_h^0)^\top. \quad (3.40b)$$

with

$$\Delta \mathbf{D}^0 = \frac{1}{\sqrt{N_e - 1}} \begin{bmatrix} \mathbf{d}_1^0 - \bar{\mathbf{d}}^0 & \cdots & \mathbf{d}_{N_e}^0 - \bar{\mathbf{d}}^0 \end{bmatrix}, \quad (3.41a)$$

$$\bar{\mathbf{d}}^0 = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{d}_j^0, \quad (3.41b)$$

and

$$\Delta \mathbf{D}_h^0 = \frac{1}{\sqrt{N_e - 1}} \begin{bmatrix} \mathbf{d}_{h,1}^0 - \bar{\mathbf{d}}_h^0 & \cdots & \mathbf{d}_{h,N_e}^0 - \bar{\mathbf{d}}_h^0 \end{bmatrix}, \quad (3.42a)$$

$$\bar{\mathbf{d}}_h^0 = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{d}_{h,j}^0. \quad (3.42b)$$

At the next data assimilation, $\mathbf{d}_{\text{prior}}^1 = \mathbf{d}_{\text{map}}^1$, $\mathbf{d}_{\text{obs}}^2 \sim \mathcal{N}(\mathbf{d}_{\text{obs}}, \mathbf{C}_e^\ell)$, and the minimization [equation \(3.30\)](#) is repeated. Following the pattern, the update equation after $1 \leq \ell \leq N_a$ assimilations becomes

$$\mathbf{d}_j^{\ell+1} = \mathbf{d}_j^\ell + \mathbf{K}^\ell (\mathbf{d}_{\text{obs}} + \sqrt{\alpha^\ell} \mathbf{e}_j^\ell - \mathbf{d}_{h,j}^\ell), \quad (3.43a)$$

$$\mathbf{K}^\ell = \Delta \mathbf{D}^\ell \Delta \mathbf{D}_h^\ell (\Delta \mathbf{D}_h^\ell \Delta \mathbf{D}_h^\ell + \alpha^\ell \mathbf{C}_e)^\ell)^{-1}. \quad (3.43b)$$

with $\mathbf{d}_{\text{obs}} + \sqrt{\alpha^\ell} \mathbf{e}_j^\ell$, $\mathbf{e}_j^\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_e)$, implementing the sampling $\mathbf{d}_{\text{obs}}^\ell \sim \mathcal{N}(\mathbf{d}_{\text{obs}}, \mathbf{C}_e^\ell) = \mathcal{N}(\mathbf{d}_{\text{obs}}, \alpha^\ell \mathbf{C}_e)$.

[Equation \(3.43b\)](#) is the base DSI-ESMDA equation. Compared to the equation originally introduced by [Lima, Emerick, and Ortiz \(2020\)](#), a covariance localization factor was omitted. This factor improves the Kalman-like update by diminishing the Kalman gain between weakly correlated data, and is addressed in the following subsection.

3.3.4 Improving data match with covariance localization

The quality of the approximations in [equation \(3.40b\)](#) increase as the amount of members in the ensemble increase, but the larger the ensemble, the more computational time is

spent processing it. To make matters worse, the quality of the approximations deteriorate as data dimension increase, since more spurious correlations are introduced. Hence, small ensembles tend to benefit from some process of correlation adjustment, specially when applied to a complex reservoir.

Covariance localization, or simply localization, is an *ad hoc* procedure employed to mitigate the effects of small ensembles (Emerick, 2025, section 7.5). Its idea is weighting the influence a datum has on another datum in a manner inverse to the distance between the points. Localization is applied through Schur product between this matrix of weights $\mathbf{R}$, henceforth referred to as the localization matrix, and either the Kalman-like gain or the correlation matrices themselves.

Lima, Emerick, and Ortiz (2020) apply localization to the Kalman-like gain, so that equation (3.43b) becomes:

$$\mathbf{d}_j^{\ell+1} = \mathbf{d}_j^\ell + (\mathbf{R} \circ \mathbf{K}^\ell)(\mathbf{d}_{\text{obs}} + \sqrt{\alpha^\ell} \mathbf{e}_j^\ell - \mathbf{d}_{h,j}^\ell), \quad (3.44)$$

Intuitively, a correlation is strong if measures of a same quantity were taken closer in space and/or in time, and weaker otherwise. Denoting this “distance” by δ_{uv} , where u is an index from $\mathbf{d}_j$ and v an index from $\mathbf{d}_h$, $\mathbf{R}$ is given by

$$R_{uv} = r(\delta_{uv}), \quad (3.45)$$

$$r(\delta) = \begin{cases} -\frac{1}{4}\delta^5 + \frac{1}{2}\delta^4 + \frac{5}{8}\delta^3 - \frac{5}{3}\delta^2 + 1, & 0 \leq \delta \leq 1 \\ \frac{1}{12}\delta^5 - \frac{1}{2}\delta^4 + \frac{5}{8}\delta^3 + \frac{5}{3}\delta^2 - 5\delta + 4 - \frac{2}{3}\delta^{-1}, & 1 < \delta \leq 2, \\ 0, & \delta > 2 \end{cases}, \quad (3.46)$$

$$\delta_{uv} = \sqrt{\left(\frac{\Delta x'_{uv}}{L_{x'}}\right)^2 + \left(\frac{\Delta y'_{uv}}{L_{y'}}\right)^2 + \left(\frac{\Delta t_{uv}}{T_c}\right)^2}, \quad (3.47)$$

$$\begin{bmatrix} \Delta x'_{uv} \\ \Delta y'_{uv} \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \Delta x_{uv} \\ \Delta y_{uv} \end{bmatrix}, \quad (3.48)$$

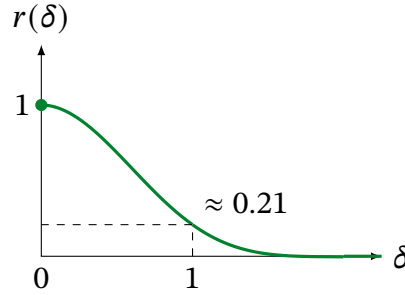
where r , a common choice in the context of localization, is the Gaspari-Cohn 5th-order correlation function (Gaspari; Cohn, 1999).

Δx_{uv} and Δy_{uv} are the x and y differences, respectively, between the points in space from where measure u and v were taken. In this context, measurements always come from a well, and so these differences are always distances between wells. Δt_{uv} is the difference between the time when measurement u was taken, and the time when v was taken. Note that while v is restrained to the history matching period by definition, u is not. The inclusion of Δt_{uv} guarantees that measurements far apart in time will hold little correlation, even if taken from the same well.

A rotation by θ is applied to take into account the anisotropy of reservoir permeability, and are $L_{x'}$, $L_{y'}$, T_c positive constants that regulate the critical points of differences in x , y

and t separately. By critical, it is understood the inflection point $\delta = 1$ in $r(\delta)$. As illustrated by [figure 3.4](#), the greater the critical value, the more lax the localization, allowing more weakly correlated changes to hold more impact.

Figure 3.4 – Gaspari-Cohen 5th-order compact correlation function.



Source: Self-elaboration.

3.4 Gated DSI-ESMDA

A well tuned localization tends to greatly improve the quality of the posterior, but if the prior zeroes in on an average that is far mismatched from the reference, it is hard to tune it in order to inject some variance on the collapsed curves. This is partly the reason behind the random disturbance in injecting variance on the ensemble fit.

Even when the localization $\mathbf{R}$ is well tuned (i.e., θ , $L_{x'}$, $L_{y'}$ and T_c generate adequate posteriors), DSI-ESMDA may suffer the inherent history matching problem of making poor predictions despite good fit to past data. The ensemble posterior narrows around a curve that is not representative of the reservoir.

The inherent problem of history matching may be impossible to solve, but it may be mitigated by allowing the posterior to spread more. The proposed Gated DSI-ESMDA introduces a gating mechanism into [equation \(3.44\)](#) in order to increase posterior variance. This section derives the proposal's update equation by first deriving the gating coefficient ([subsection 3.4.1](#)) and then coupling it with the standard DSI-ESMDA equation ([subsection 3.4.2](#)).

3.4.1 Deriving the gating coefficient

This subsection provides a self-contained derivation of the alpha parameter used in the innovation gating mechanism. The derivation follows closely the work of [Fortaleza et al. \(2022\)](#), from where the gating mechanism was inspired.

Let $Y \in \mathbb{R}$ be a random variable representing the output of a given process in time, and Y_I the observation of Y . It is assumed that the observation is affected by a Gaussian noise

$N \sim \mathcal{N}(0, \sigma_n^2)$, such that

$$Y_I = Y + N. \quad (3.49)$$

Let y_I be a realization of Y_I . Denoting by y_M the prediction of y from a given model, y_O is defined to be the adjusted prediction after the observation y_I is taken into account. For the sake of simplicity, y_O is given by the following exponential moving average (EMA) relationship:

$$\begin{cases} y_O[0] = y_I[0], & k = 0, \\ y_O[k] = \alpha[k]y_I[k] + (1 - \alpha[k])y_M[k-1], & k > 0 \end{cases} \quad (3.50)$$

where $\alpha[k] \in [0, 1]$ is the smoothing factor of the EMA at time step index k .

Parameter $\alpha[k]$ represents the statistical confidence that the filter output reflects the same direction of the process output at sample k . More precisely, $\alpha[k]$ is given by the following expression:

$$\begin{aligned} \alpha[k] = & P(\text{sign}(Y_M[k-1] - Y_I[k]) = \text{sign}(\mu_{Y_M[k-1]} - \mu_{Y_I[k]})) \\ & - P(\text{sign}(Y_M[k-1] - Y_I[k]) \neq \text{sign}(\mu_{Y_M[k-1]} - \mu_{Y_I[k]})), \end{aligned} \quad (3.51)$$

where $Y_M[k-1]$ is an observation of Y_M with mean $\mu_{Y_M[k-1]}$ and Y_I an observation of Y_I with mean $\mu_{Y_I[k]}$. Here $\text{sign}(x)$ represents the sign function, which is defined as:

$$\text{sign}(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases} \quad (3.52)$$

Equation (3.51) can be rewritten as follows:

$$\alpha[k] = 2P(\text{sign}(Y_M[k-1] - Y_I[k]) = \text{sign}(\mu_{Y_M[k-1]} - \mu_{Y_I[k]})) - 1. \quad (3.53)$$

Defining a random variable Y_{MI} whose observation at time k have mean and variance given by

$$\mu_{Y_{MI}}[k] = |\mu_{Y_M[k-1]} - \mu_{Y_I[k]}|, \quad (3.54)$$

$$\sigma_{Y_{MI}}^2[k] = \sigma_{Y_M[k-1]}^2 + \sigma_{Y_I[k]}^2 - 2\text{Cov}(Y_M[k-1], Y_I[k]), \quad (3.55)$$

where $\text{Cov}(Y_M[k-1], Y_I[k])$ is the covariance between $Y_M[k-1]$ and $Y_I[k]$, equation (3.53) can be expressed as follows:

$$\alpha[k] = 2P(Y_{MI}[k] > 0) - 1. \quad (3.56)$$

In order to find a practical implementation of $\alpha[k]$, equation (3.56) is analyzed under the following technical assumptions.

Assumption 3.1. The random variable $Y_{MI}[k]$ can be approximated by a normal distribution with mean $\mu_{Y_{MI}[k]}$ and variance $\sigma_{Y_{MI}[k]}^2$.

Assumption 3.2. Both high frequency and random behavior of the measurement noise implies that the observations $Y_M[k-1]$ and $Y_I[k]$ can be considered approximately statistically independent of each other. Mathematically, $\text{Cov}(Y_M[k-1], Y_I[k]) \approx 0$.

Assumption 3.3. The variance of the model's prediction is much smaller than the variance of the measured observation. Mathematically, $\sigma_{Y_M[k-1]}^2 \ll \sigma_{Y_I[k]}^2$.

Assumption 3.4. By considering an additive zero-mean measurement noise, the samples $y_M[k-1]$ and $y_I[k]$ can be used to estimate $\mu_{Y_M[k-1]}$ and $\mu_{Y_I[k]}$, respectively.

From [assumption 3.1](#), [equation \(3.56\)](#) can be rewritten as:

$$\begin{aligned}\alpha[k] &= 2P(\mathcal{N}(\mu_{Y_{MI}[k]}, \sigma_{Y_{MI}[k]}^2) > 0) - 1 \\ &= 2P(\mathcal{N}(0, \sigma_{Y_{MI}[k]}^2) \leq \mu_{Y_{MI}[k]}) - 1,\end{aligned}\quad (3.57)$$

which can be expressed in terms of the Gauss error function $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} dx$:

$$\begin{aligned}\alpha[k] &= \text{erf}\left(\frac{\mu_{Y_{MI}[k]}}{\sqrt{2\sigma_{Y_{MI}[k]}^2}}\right) \\ &= \text{erf}\left(\frac{|\mu_{Y_M[k-1]} - \mu_{Y_I[k]}|}{\sqrt{2[\sigma_{Y_M[k-1]}^2 + \sigma_{Y_I[k]}^2 - 2\text{Cov}(Y_M[k-1], Y_I[k])]}}\right).\end{aligned}\quad (3.58)$$

Applying [assumptions 3.2](#) to [3.4](#), [equation \(3.58\)](#) becomes

$$\alpha[k] = \text{erf}\left(\frac{|y_M[k-1] - y_I[k]|}{\sqrt{2\sigma_{Y_I[k]}^2}}\right) = \text{erf}\left(\frac{|y_M[k-1] - y_I[k]|}{\sqrt{2\sigma_n^2[k]}}\right).\quad (3.59)$$

Since knowledge of the measurement noise variance $\sigma_n^2[k]$ is not always available in practice, a fixed $\lambda \geq 1$ is introduced to scale the noise standard deviation. Thus, the final expression for $\alpha[k]$ is given by

$$\alpha[k] = \text{erf}\left(\frac{1}{\lambda\sqrt{2}} \frac{|y_M[k-1] - y_I[k]|}{\sigma_n[k]}\right)\quad (3.60)$$

which is equivalent to [equation \(3.61b\)](#) in [chapter 3](#).

3.4.2 Deriving the update equation

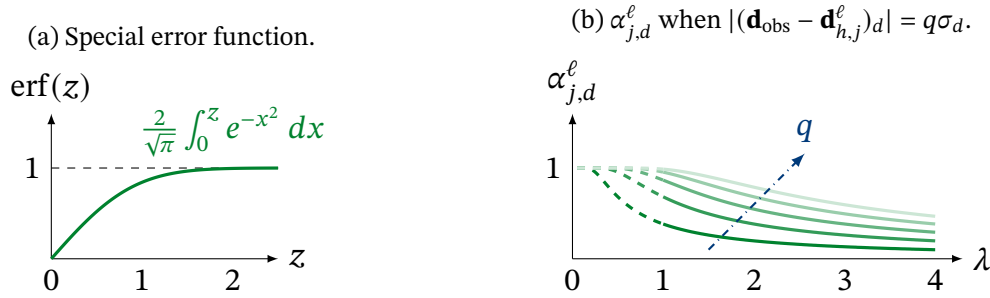
Gated DSI-ESMDA applies the gating coefficient directly into the innovation vector:

$$\mathbf{d}_j^{\ell+1} = \mathbf{d}_j^\ell + (\mathbf{R} \circ \mathbf{K}^\ell) [\alpha_j^\ell \circ (\mathbf{d}_{\text{obs}} - \mathbf{d}_{h,j}^\ell)], \quad (3.61a)$$

$$\alpha_{j,d}^\ell = \text{erf} \left(\frac{1}{\lambda \sqrt{2}} \frac{|(\mathbf{d}_{\text{obs}} - \mathbf{d}_{h,j}^\ell)_d|}{\sigma_d} \right) \quad (3.61b)$$

Constant $\lambda \geq 1$ is named the gating intensity, and it inflates the estimate of the standard deviation of the measurement error at datum d . The gating recovers variance loss in the posterior ensemble by diminishing Kalman-like updates on reasonably matched data. As [figure 3.5b](#) shows, the smaller the mismatch of a measurement relative to the noise at that datum, the smaller the coefficients, and thus the smaller the update. Gating intensity further softens the update by bringing the coefficient closer to zero. At the extreme ($\lambda \rightarrow \infty$), $\alpha_j^\ell = \mathbf{0}$ and the Kalman-like update is fully undone.

Figure 3.5 – Important functions for understanding the proposed Data-Space Inversion algorithm.



Source: Self-elaboration.

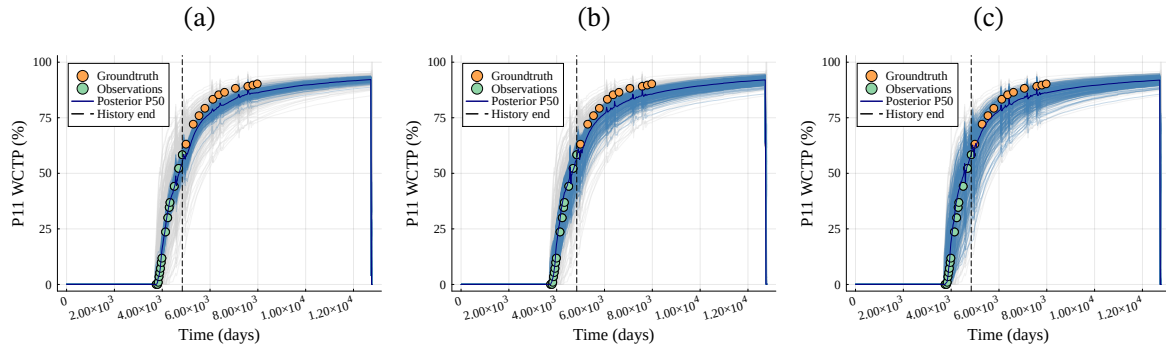
Note 3.2: Index d

Data vectors stack time series of a fields' measurements, so index d in this chapter's equations specify a datum in place and time simultaneously. For example, if DSI is being applied on the oil production of two wells of a field, each with a simulation time vector of 100 elements, then for any $1 \leq j \leq N_e$ $\mathbf{d}_j$ would be 200×1 , with elements $d = 1$ through $d = 100$ the oil production history of the first well, and elements $d = 101$ through $d = 200$ the oil production history of the second well.

[Figure 3.6](#) illustrates the effect of gating. Prior ensemble is shown in gray, while posterior ensemble is shown in blue. The 50th percentile of the posterior is drawn as a darker blue line under label P50. The reference data is shown in teal during the history period (dashed vertical line), and yellow beyond it. The left panel resulted from a base application of DSI-ESMDA, the center panel, from Gated DSI-ESMDA with $\lambda = 4$, and the right panel, from Gated DSI-ESMDA with $\lambda = 16$.

The posterior recovers variance as λ grows, gradually reverting to the prior distribution. Unlike localization, whose effect on posterior variance tend to target designed periods, the effects of gating are roughly uniform.

Figure 3.6 – Illustration of the effect of λ on the posterior ensemble for fixed $\mathbf{R}$ and N_a .



The Kalman-like update in [equations \(3.44\) and \(3.61a\)](#) acts on the entirety of the realization data vector by projecting the history match correction through the cross-covariance of data in the history period and data in the forecast period. That being the case, the quality of the update is heavily dependent on the quality of the prior ensemble; a prior not being representative of the reference, or a history period with too little data, are significant obstacles in the pursuit of a well adjusted posterior.

Note 3.3: Noise modeling.

Technically speaking, this work will apply the standard DSI-ESMDA and the proposed algorithm on normalized data vectors. As [section 4.3](#) clarifies, this is done to facilitate noise injection and modeling, and does not affect the validity of the theory and formulas presented in this chapter.

3.5 Chapter summary

Key takeaways

This chapter has presented the statistical framework underpinning DSI-ESMDA, and, by extension, the proposed Gated DSI-ESMDA. It was shown that DSI-ESMDA iteratively updates the maximum *a posteriori* solution by assimilating resamplings of observed data multiple times. While the method requires (i) assuming Gaussian prior and likelihood and (ii) knowledge of the data error covariance $\mathbf{C}_e$, it is computationally efficient and statistically sound. It was seen that the negative effects on covariance estimation due to a small ensemble can be mitigated via localization, an *ad hoc* strategy that restricts

the influence of spurious correlations. As an extension of DSI-ESMDA, the proposed approach enjoys and suffers the same benefits and limitations. However, its innovation gating greatly facilitates the recovery of posterior variance by reverting the posterior to the prior in a roughly uniform fashion. The intensity of the reversal depends on the gating intensity, and in the best case a well tuned gating could bypass the need for localization.

4 Methods

4.1 Chapter outline

Aims and structure

This chapter characterizes the case studies used in this work, and explains the procedure used to obtain the findings in [chapter 5](#), and what values were set for its parameters. Case study characterization comprises: describing how the ensemble was generated; what are the geometrical and geological features of the reservoir; how many wells each realization has; what fluid is in the reservoir. [Section 4.2](#) characterizes each case study, capping its discussion with a commentary on the cumulative distributions of field oil production totals. [Section 4.3](#) presents the workflow designed to generate and evaluate results on each case study, including what metrics were used, how experiment parameters were set, and what data features were used.

4.2 Reservoir case studies

Three reservoirs were used as case studies for this work, two of which are synthetic reservoirs and one is a real reservoir. This section presents them, highlighting their most significant structural and geological characteristics within the scope of this work. However, as information regarding the real reservoir is sensitive, it is only briefly discussed.

The choice of these reservoirs was based on the availability of data and their suitability for testing the proposed methods: the Egg and Olympus ensembles are widely used in the literature for benchmarking purposes (with the Olympus ensemble being the more complex of the two), while the real field provides a practical application scenario for the algorithms employed in this work.

Simulations of the Egg and Olympus ensembles to generate the associated priors were performed using the JutulDarcy.jl simulator ([Møyner, 2025](#)).

4.2.1 Egg ensemble

Originally introduced as a deterministic model in a study aimed at exploring the reasons behind reservoir flooding problems and identifying conditions for optimal solutions ([Zandvliet *et al.*, 2007](#)), a stochastic version was later proposed as a benchmark in a robust waterflooding optimization problem ([Essen *et al.*, 2009](#)). This reservoir consists of 18,533

active cells with dimensions of 8x8x4 m. The arrangement of these cells forms an egg-shaped grid, hence the name (Jansen *et al.*, 2014).

4.2.1.1 Realizations

The standard Egg ensemble comprises a single deterministic realization and 100 permeability realizations of a channelized reservoir. In addition to eight injectors labeled INJECT1 through INJECT8, each realization includes four vertical producer wells labeled PROD1 through PROD4, which produce oil and water but not gas. Table 4.1 shows some of the properties shared by the realizations.

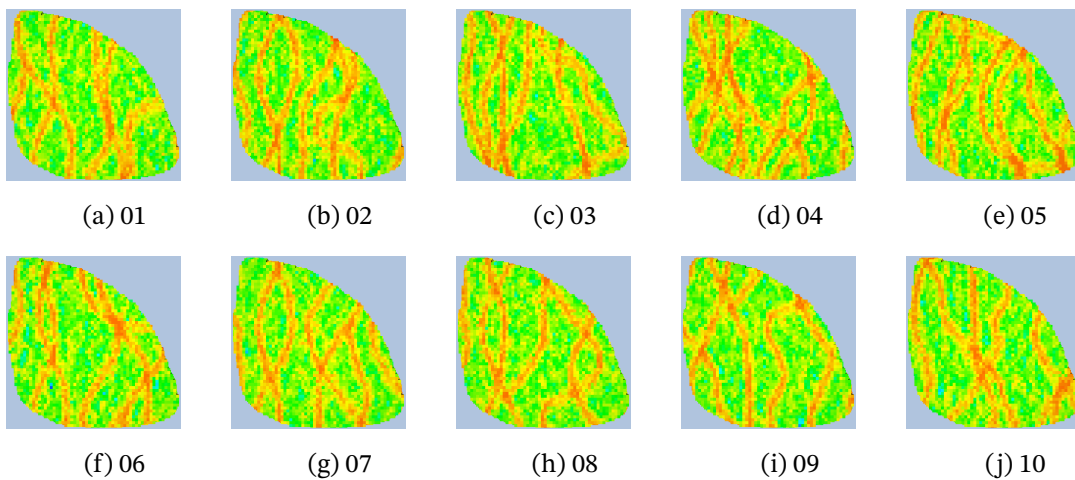
Table 4.1 – Some geometry and fluid properties of the Egg ensemble.

Property	Value	Unit
Rock Compressibility	0.0	1/Pa
Water Compressibility	1.0×10^{-10}	1/Pa
Oil Dynamic Viscosity	5.0×10^{-3}	Pa·s
Water Dynamic Viscosity	1.0×10^{-3}	Pa·s
Well Diameter	20	cm
Simulation Time	3600	days

Adapted from Jansen *et al.* (2014).

In figure 4.1, renderings of permeability along the x-axis in some of Egg’s realizations are depicted, with colors assigned according to a logarithmic scale. To enhance visualization, the z-axis was magnified fivefold.

Figure 4.1 – Top view of selected realizations 1-10 from the Egg ensemble. Renderings were produced using the open-source software ResInsight.



Source: Filho *et al.* (2025).

4.2.2 Olympus ensemble

Launched in early 2017, the Olympus reservoir model was designed as a benchmark environment to problems concerning control optimization and reservoir management within the framework of the Integrated Systems Approach for Petroleum Production 2 research consortium (Fonseca *et al.*, 2018; Fonseca *et al.*, 2020). In what follows, the depiction of Olympus mirrors the one outlined by Fonseca *et al.* (2020), with adjustments made where deemed necessary.

Inspired by an oil field in the North Sea, this synthetic reservoir is approximately rectangular, measuring 9x3 km, with a depth of 50 m divided into 16 layers. The 3D grid comprises cells with dimensions of approximately 50x50x3 m, totaling 341,728 cells, of which 192,750 are active. The perimeter of the reservoir is defined by a fault, with six additional smaller faults present within it. Furthermore, Olympus features two distinct zones separated by an impermeable shale layer: The upper zone contains fluvial channel sands embedded in floodplain shales, while the lower zone consists of alternating layers of coarse, medium, and fine sands that are inclined relative to the general structural dip of the field.

The reservoir includes four facies types in different layers (see table 4.2), and geological properties, such as porosity, permeability, and the net-to-gross (NG) ratio, were generated using standard geostatistical techniques for the different facies types. In the absence of sufficient early-stage data, no porosity/permeability ratio was used. Permeabilities in the x and y directions were identical, while in the z direction, it corresponded to 10% of that in the x direction.

Table 4.2 – Types and distribution range intervals of facies properties in Olympus.

Type	Presence zones	NG	Porosity	Permeability (μm^2)
Channel sand	Upper	0.8-1	0.2-0.35	0.4-1
Shale	Upper, perimeter	0.0	0.03	0.001
Coarse sand	Lower	0.7-0.9	0.2-0.3	0.15-0.4
Medium sand	Lower	0.75-0.95	0.1-0.2	0.075-0.15
Fine sand	Lower	0.9-1	0.05-0.1	0.01-0.05

Source: Fonseca *et al.* (2020).

Based on available exploration well records, it was determined that the oil/water contact depth was 2092 m, with an in-situ hydrostatic pressure of 205 bar. The initial water saturation distribution was modeled using capillary pressure curves. Different capillary pressure curves were assigned to different facies types. Therefore, changes in the facies resulted in different initial water saturation distributions. The reservoir's fluid was similar to oil without dissolved gas, and some of its properties are summarized in table 4.3.

Table 4.3 – Olympus' fluid properties.

Property	Value	Unit
Rock compressibility	1.42×10^{-10}	1/Pa
Water compressibility	3.97×10^{-10}	1/Pa
Oil dynamic viscosity	$2.8-3.5 \times 10^{-3}$	Pa·s
Water dynamic viscosity	0.398×10^{-3}	Pa·s
Well diameter	19	cm
Simulation Time	7380	days

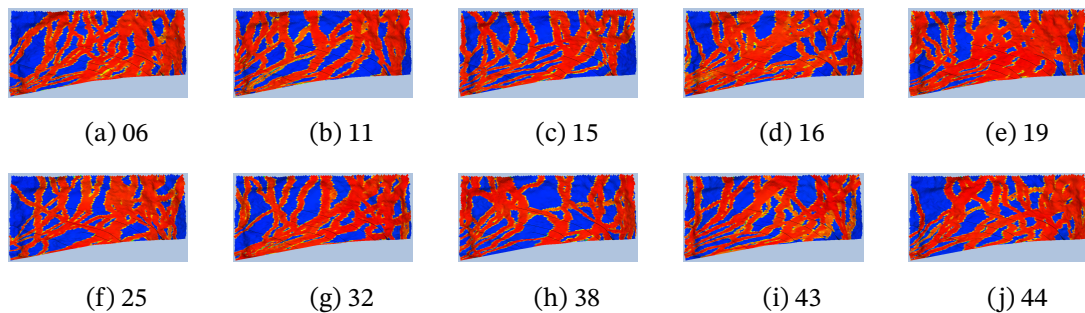
Source: Adapted from [Fonseca et al. \(2020\)](#).

4.2.2.1 Realizations

A set of 50 realizations was generated by varying facies through random seed changes, while keeping consistent mesh geometry, faults, and oil-water contact. Uncertain properties such as facies, porosity, permeability, NG ratio, initial water saturation distribution, and transmissibility within faults were considered.

By first generating a high-fidelity base model with approximately 5 million grid cells, 5 wells were drilled into this base case model and synthetic records were generated for each of these wells. These records were then used to constrain the generation of 50 high-fidelity realizations to capture uncertainty. Each was upscaled for flow simulations using the flow-based upscaling method, featuring a total of 18 wells in each realization, of which 11 are producers and 7 are injectors. Each realization includes 7 injector wells, labeled INJ-1 to INJ-7, and 11 producer wells, labeled PROD-1 to PROD-11. Producers produce oil and water, but not gas. [Figure 4.2](#) illustrates the permeability in the x direction for selected realizations, depicted with colors on a logarithmic scale and a fivefold exaggeration of the z -axis.

Figure 4.2 – Top view of selected realizations from the Olympus ensemble. Renderings were produced using the open-source software ResInsight.



Source: [Filho et al. \(2025\)](#).

4.2.3 Real field

Given that most parameters characterizing the properties of the carbonate reservoir in question cannot be disclosed due to ongoing confidentiality agreements with industrial

partners, only essential information is provided.

The real field is a mature reservoir located in the Brazilian Post-Salt, with a production history spanning over 20 years. It is comprised of 8 injector wells, labeled I1 to I8, and 16 producer wells, labeled P1 to P16. Producers produce oil and water, but not gas.

An ensemble of 200 realizations was generated from observed production data, with a total of 510 simulation time steps of roughly 30 days each. Importantly, observed data has a time vector that is not identical to the simulation's but is a subset of it.

4.2.4 Field oil production totals

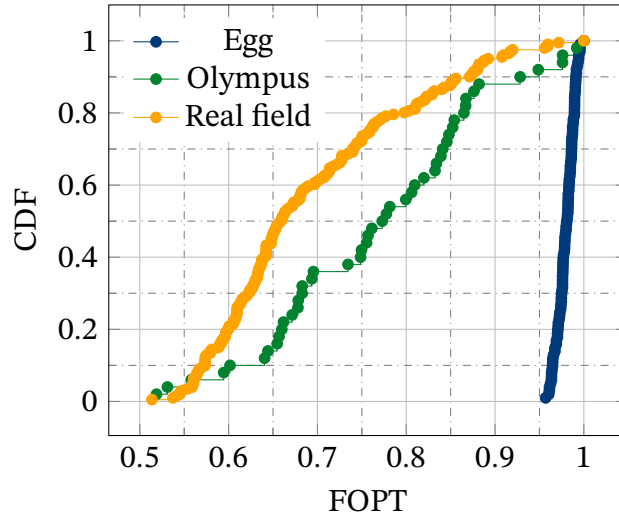
The field oil production total (FOPT) is a measurement commonly used to compare ensembles since ground-truth and realizations are each represented by a single value: how much oil was produced within the recorded data period. Furthermore, the FOPT range in an ensemble is correlated with reservoir complexity, so that simpler reservoirs are expected to exhibit a narrower range of FOPT in their ensemble than a more complex reservoir. [Figure 4.3](#) shows the cumulative distribution function (CDF) of field oil production total (FOPT) of the case studies ensembles. FOPT values are normalized by the maximum across all a case study's realizations.

The Egg ensemble has a relatively narrow, roughly uniform range of FOPT values, suggesting that the priors should be reasonable representatives of the ground-truths, especially those that are closer to the median of the distribution.

On the other hand, the Olympus ensemble has a much wider (about 10 times that of the Egg ensemble), albeit still roughly uniform, range of FOPT values. The picked ground-truths (realizations 5, 10, etc.) are spread within the 5-95th percentiles, with half of them quite distant from the median. Since this reservoir's FOPTs thin out at the percentile ranges of 0-5% and 90-100%, it would be expected that ground-truths picked from these ranges would be more difficult to match, as they are less likely to be represented in the prior ensemble. This effect is further aggravated by the large range of FOPT values in the Olympus ensemble, which is roughly 10 times larger than that of the Egg ensemble.

The real ensemble has a similarly wide range of FOPT values. However, the distribution quite lopsided, with a very populated 5-80th percentile range and a thin 0-5% and 80-100% range. As with the Olympus case, a broad range of FOPT values coupled with poorly-represented tails can lead to inadequate priors depending on where the ground-truth sits.

Figure 4.3 – Distribution of field oil production totals for each realization in the Egg, Olympus, and real field ensembles. Each ensemble is normalized by its maximum FOPT.



Source: Self-elaboration.

4.3 Evaluation workflow and parameters

While the outline of both DSI-ESMDA and the proposed Gated DSI-ESMDA methods have been presented in [chapter 3](#), it remains to be explained how experiments will be set up ([subsection 4.3.1](#)), how the error covariance matrix $\mathbf{C}_e$ is constructed ([subsection 4.3.2](#)), what metrics will be used to evaluate the performance of the methods ([subsection 4.3.3](#)), and what parameters and features are used ([subsection 4.3.4](#)).

4.3.1 Evaluation workflow

The evaluation pipeline, illustrated in [figure 4.4](#) is executed once per ground-truth data vector $\mathbf{d}$. In synthetic ensembles, 10 ground-truths are picked. The average of these results, as well as the results found with the real field case study, are presented in the following chapter. More detailed results that would not have fit with the flow of the text are exhibited in [appendix C](#).

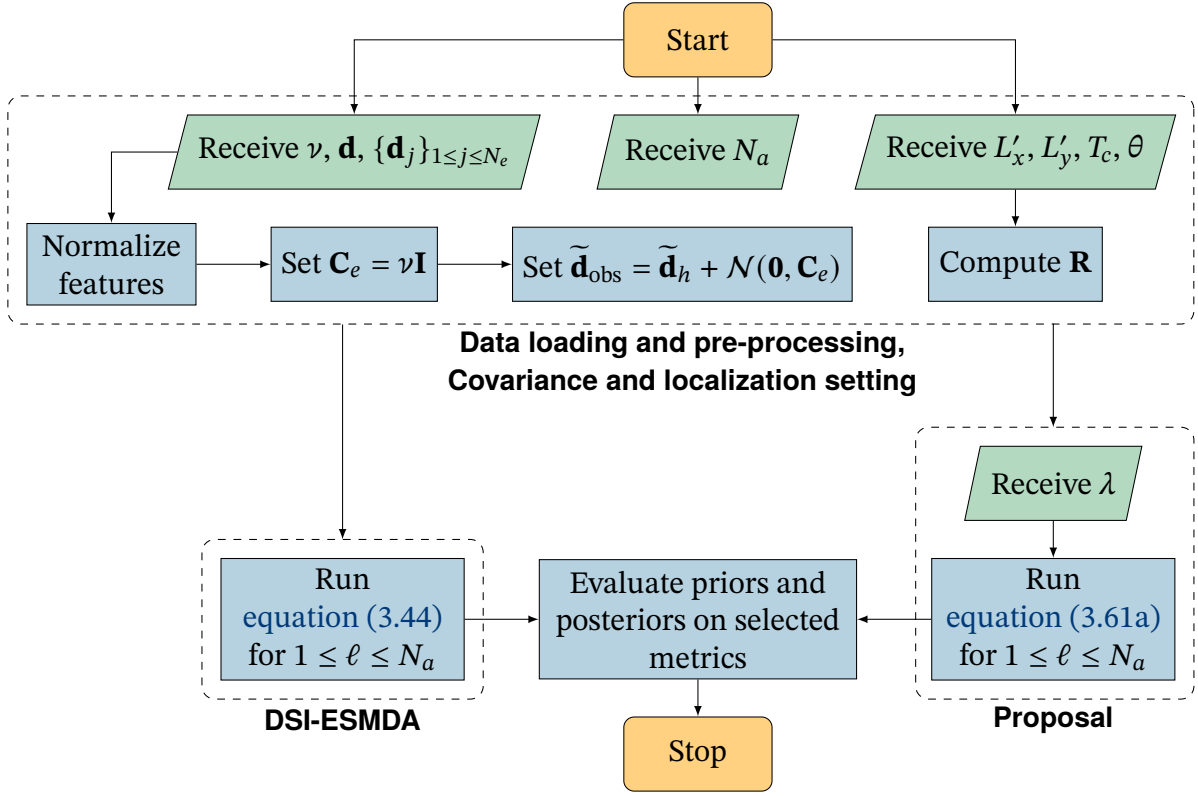
First, all data is normalized to the range $[0, 1]$ using the minimum and maximum values of the ground-truth production data, as follows:

$$\tilde{\mathbf{d}}_{j,f} = \frac{\mathbf{d}_{j,f} - \min(\mathbf{d}_f)}{\max(\mathbf{d}_f) - \min(\mathbf{d}_f)} \quad (4.1)$$

$$\tilde{\mathbf{d}}_f = \frac{\mathbf{d}_f - \min(\mathbf{d}_f)}{\max(\mathbf{d}_f) - \min(\mathbf{d}_f)}, \quad (4.2)$$

where index j denotes the realization, f denotes the production data feature, and the tilde indicates normalization. This process is repeated for all non-constant features and

Figure 4.4 – Outline of the evaluation process.



Source: Self-elaboration.

realizations to improve the conditioning of the data and facilitate noise modeling, as shown next.

4.3.2 Noise modeling

The error covariance matrix $\mathbf{C}_e$ is assumed to be diagonal, with its non-zero entries uniformly sampled from $(0, 1]$ and scaled by a constant $\nu > 0$. This constant is named the noise magnitude and is a hyperparameter of the Evaluation. However, ν is only means to easier explore the effects of different noise magnitudes on the performance of the methods, and not an essential parameter of the methods themselves.

Next, observed data is constructed by adding noise to the normalized ground-truth data restricted to the history period:

$$\tilde{\mathbf{d}}_{\text{obs}} = \tilde{\mathbf{d}}_h + \mathbf{e}_d, \quad \mathbf{e}_d \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_e). \quad (4.3)$$

Since real field data already contains noise, this step is skipped for that case, and observed data is simply the normalized ground-truth data restricted to the history period.

Finally, the methods are applied to the pre-processed data, and their priors and posteriors are evaluated using the metrics described in [subsection 4.3.3](#).

4.3.3 Evaluation metrics

Both the prior and posterior ensembles are evaluated on the history and validation periods using the following metrics (Emerick, 2025, sections 9.4.2 and 9.4.4):

- observation coverage (OC),
- mean squared error (MSE),
- data-mismatch objective function (denoted O).

OC measures the percentage of observed data that falls within the range of an ensemble, while MSE quantifies the weighted squared difference between the observed and predicted data. Data-mismatch objective function, or simply data mismatch, quantifies the discrepancy between observed and predicted data. Metrics are defined as follows:

$$OC_h(\mathbf{Y}) = \frac{1}{N_h} \sum_{d=1}^{N_h} \mathbb{1}(\min(\mathbf{Y})_d \leq d_{h,d} \leq \max(\mathbf{Y})_d), \quad (4.4)$$

$$OC_v(\mathbf{Y}) = \frac{1}{N_v} \sum_{d=1}^{N_v} \mathbb{1}(\min(\mathbf{Y})_d \leq d_{v,d} \leq \max(\mathbf{Y})_d), \quad (4.5)$$

$$MSE(\mathbf{Y}) = \frac{1}{N_h} (\mathbf{d}_h - \bar{\mathbf{y}})^\top \mathbf{C}^{-1} (\mathbf{d}_h - \bar{\mathbf{y}}), \quad \mathbf{C} = \Delta \mathbf{Y} \Delta \mathbf{Y}^\top + \mathbf{C}_e \quad (4.6)$$

$$O(\mathbf{y}_j) = \frac{1}{2N_h} (\mathbf{d}_h - \mathbf{y}_{j,h})^\top \mathbf{C}_e^{-1} (\mathbf{d}_h - \mathbf{y}_{j,h}), \quad (4.7)$$

where

- $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{N_e}]$ denotes an ensemble (posterior or prior),
- N_h is the number of data points in the history period,
- N_v is the number of data points in the validation period,
- $\bar{\mathbf{y}}$ is the mean of the ensemble,
- $\max(\mathbf{Y})_d$ is the ensemble maximum at datum d ,
- $\min(\mathbf{Y})_d$ is the ensemble minimum at datum d ,
- $\Delta \mathbf{Y}$ is the deviation of the predicted data from its mean (see equation (3.41a)),
- $\mathbb{1} : \{0, 1\} \rightarrow \{0, 1\}$ returns 1 if its argument is true and 0 otherwise.

Both history and validation splits of ground-truth $\mathbf{d}$ are used to compute observation coverages, while MSE and misfit exist only for the history period. Despite the similarity between the MSE and the misfit, the latter is computed for each realization in the ensemble, while the former is computed for the entire ensemble. In this work, the misfit will be presented as an average over all realizations in the ensemble, with an associated standard deviation.

4.3.4 Evaluation parameters

While a straightforward application of DSI-ESMDA requires knowledge of the error covariance matrix (addressed in [subsection 4.3.2](#)) and the number of Kalman-like iterations N_a , the amount of parameters to tune balloon if localization is utilized. The proposed Gated DSI-ESMDA method introduces an additional hyperparameter, the gating intensity λ . In practice, the gating is introduced after the standard DSI-ESMDA has been applied and tuned, making the tuning of λ incremental and independent of the other parameters.

[Table 4.4](#) compares the time-related simulation parameters and features across reservoirs, while [table 4.5](#) compares the localization-related parameters. Note that N_e represents the number of realizations after a ground-truth is separated from the ensemble.

Experiments were performed with the number of iterations fixed at $N_a = 1$ and noise magnitude $\nu = 0.01$ for all experiments. Both standard DSI-ESMDA and Gated DSI-ESMDA were run with and without localization. The value of gating intensity λ varied in $\{1, 2, 4, 8, 16\}$. Finally, [table 4.6](#) lists the data features used for each case study. Feature selection was based on economic potential and relevance in subsurface dynamics.

Table 4.4 – Time-related simulation parameters per reservoir.

Reservoir	N_e	Number of time steps	Time step size	Simulation life		History split	
Egg	100	120	30 days	3600 days	(9.8 years)	60%	(5.9 years)
Olympus	49	20	1 calendar year	7305 days	(20 years)	60%	(12 years)
Real field	200	510	1 calendar month	12785 days	(35 years)	60%	(21 years)*

Source: Self-elaboration.

*: Real data time vectors aren't synchronized across features, so 60% of a feature may lead to a shorter time scope than another feature.

Table 4.5 – Localization-related simulation parameters per reservoir. Lengths in gridblocks.

Reservoir	Critical length in x	Critical length in y	Critical time	Anisotropy
Egg	100	100	3000 days	0°
Olympus	100	100	7200 days	0°
Real field	-	-	10500 days	0°

Source: Self-elaboration.

Table 4.6 – Data features used per reservoir well.

Reservoir	Well features used		Number of wells		Total features
	Producer	Injector	Producer	Injector	
Egg	OPR, WPR	BHP	4	8	16
Olympus	OPR, WPR	WIR	11	7	29
Real field	BHP, WCTP	BHP	16	8	39 (a WCTP was constant)

Source: Self-elaboration.

4.4 Chapter summary

Key takeaways

This chapter has presented the necessary tools for understanding how this works findings were obtained. From the CDFs of FOPTs, it was seen that Egg realizations are very simple, while Olympus is closer in complexity to the real field. Posterior distributions of reservoir predictions are evaluated across four metrics, observation coverage (of the history and validation periods), mean square error (history period only), and data mismatch (history period only). The general workflow of an experiment consists of loading data, normalizing it, constructing the localization matrix (if required), and running a single iteration of the ensemble update equations

5 Results

5.1 Chapter outline

Aims and structure

This chapter presents the main findings of each of the three case studies: the Egg ensemble ([section 5.2](#)), the Olympus ensemble ([section 5.3](#)), and the real field in the Brazilian post-salt ([section 5.4](#)). Brief commentaries are made on the trend of findings, and the computational time of the experiments are shown in [table 5.1](#). Experiments were run on a commercial personal use laptop whose specifications are presented in [table 5.2](#). More detailed results are presented in [appendix C](#).

Table 5.1 – Iteration and localization construction times after 10 runs.

Case study	Loc.	Gating	Time (s)	
			Avg.	Std. dev.
Egg	No	No	1.495	0.168
	No	Yes	0.188	0.085
	Yes	-	2.683	0.201
Olympus	No	No	0.067	0.080
	No	Yes	0.007	0.002
	Yes	-	0.230	0.002
Real field	No	No	0.715	0.136
	No	Yes	0.512	0.122
	Yes	-	8.543	0.184

Source: Self-elaboration.

Table 5.2 – Component specifications of the machine used in order to run and benchmark experiments.

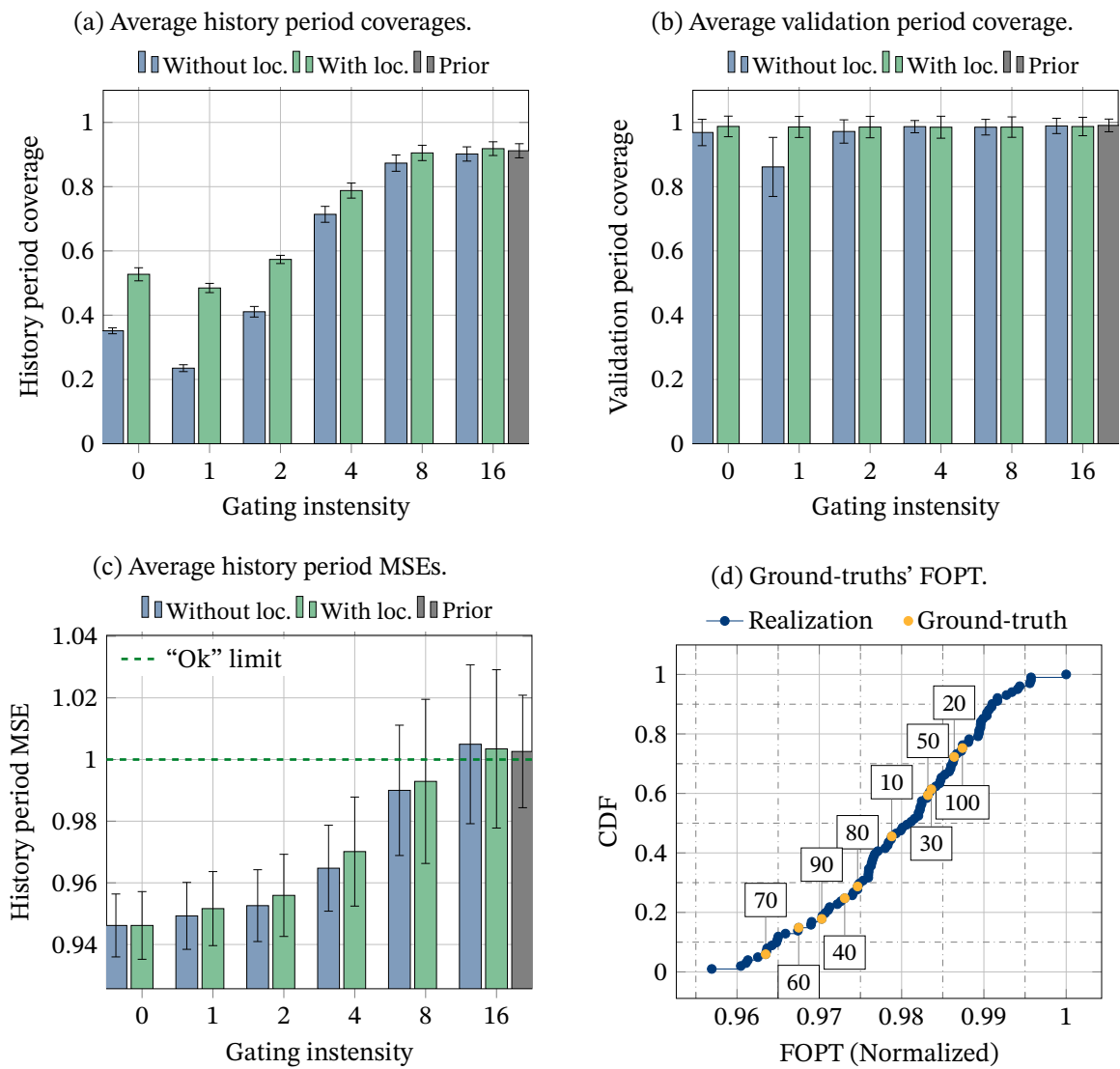
Hardware model	Lenovo ThinkBook 14 G6 IRL
CPU	13th Gen Intel® Core™ i5-13420H x 12
Threads per core	2
CPU max. clock	4.6 GHz
GPU	Intel Corporation Raptor Lake-P [UHD Graphics]
Cache memory	320 kiB (L1d) 384 kiB (L1i) 7 MiB (L2) 12 MiB (L3)
RAM	2x8 GiB DDR5 SODIMM (5600 MT/s)
Disk memory	512 GB
Operating system	Arch Linux (64-bit)

Source: Self-elaboration.

5.2 Egg ensemble

Figure 5.1 summarizes the findings of the Egg ensemble study case by averaging posterior observation coverages and mean square errors across the 10 ground-truths selected. The FOPT CDF suggests that ground-truths are fairly unbiased. Standard DSI-ESMDA is indicated by gating intensity 0. For more details on this case study’s results, see [section C.1](#). OC and MSE clearly tended to increase with gating intensity, and with relatively small standard deviations. MSE, in particular, remained fairly close a threshold deemed “okay” in reservoir prediction literature ([Emerick, 2025](#), subsection 9.4.2). OC and MSE were found insensitive to localization (green bars) for intense enough gating.

Figure 5.1 – Average posterior coverage and MSE values found in the Egg ensemble.

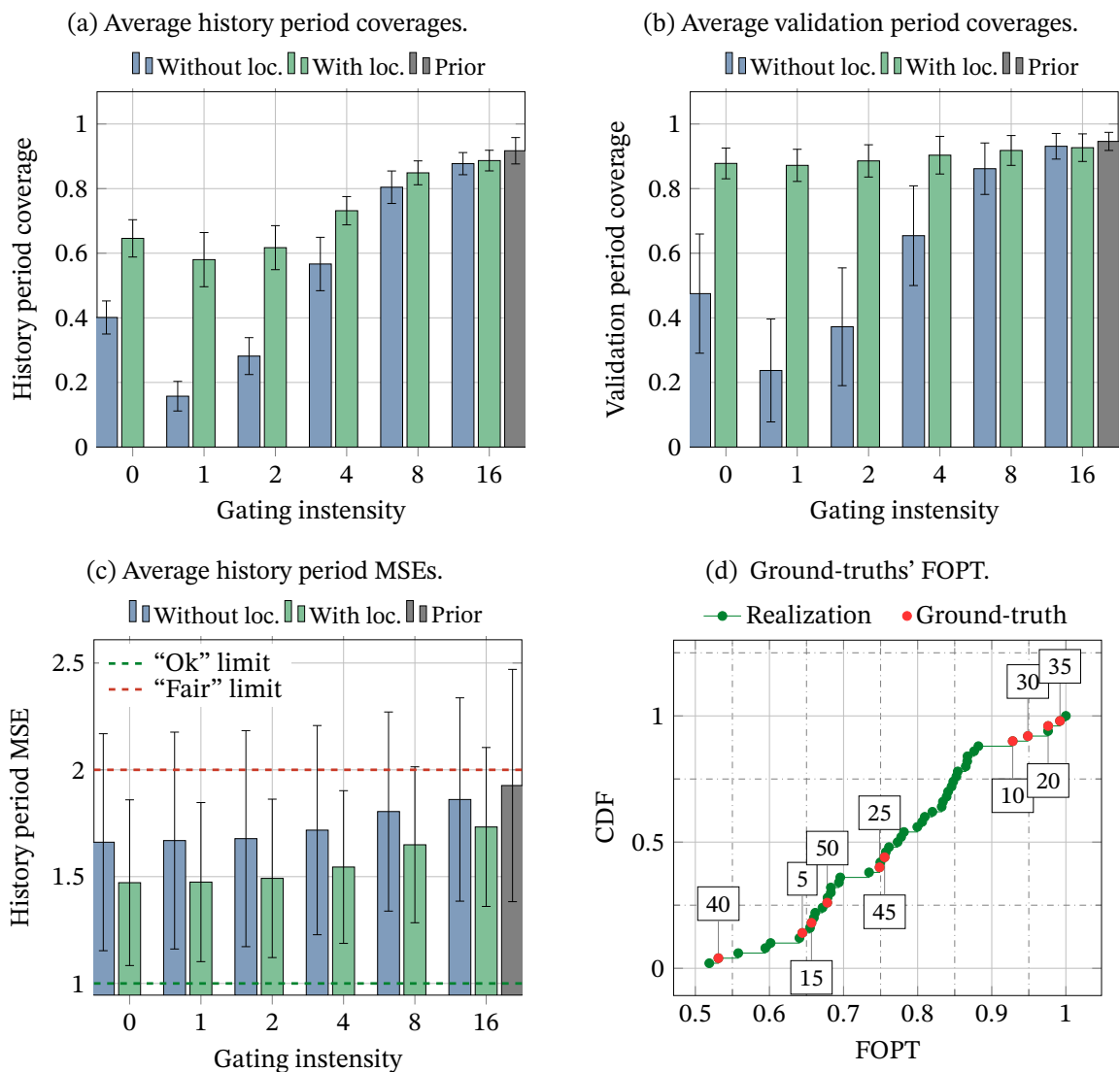


Source: Self-elaboration.

5.3 Olympus ensemble

Figure 5.2 summarizes the findings of the Olympus ensemble study case by averaging posterior observation coverages and mean square errors across the 10 ground-truths selected. The FOPT CDF suggests that ground-truths are somewhat biased towards extreme values. Standard DSI-ESMDA is indicated by gating intensity 0. For more details on this case study's results, see [section C.2](#). OC and MSE clearly tended to increase with gating intensity, albeit with much larger deviations than the Egg case. MSE, in particular, remained well beyond the okay limit, and in worst cases did not even satisfy the “fair” limit. Also contrasting with the Egg case, localization was shown to significantly improve OC and, to a lesser extent, reduce MSE.

Figure 5.2 – Summary of results for the Olympus ensemble. Posterior coverage and MSE are averaged across ground-truths.

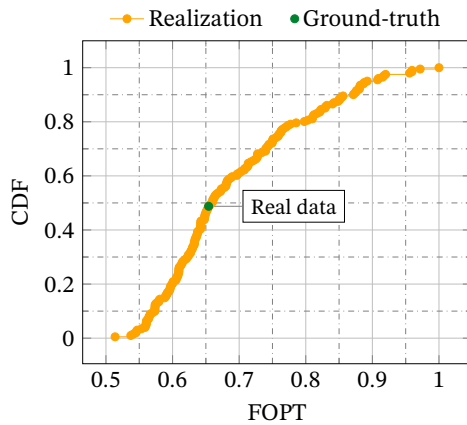


Source: Self-elaboration.

5.4 Brazilian post-salt field

For this case study analysis, only the real field was picked as ground-truth for running the experiments, so results need not be averaged/summarized. Section C.3 plots the posterior ensemble from Gated DSI-ESMDA with $\lambda = 4$. Figure 5.3 shows that real data nearly squarely at the 50th percentile of FOPT. Main findings for the standard DSI-ESMDA are presented in table 5.3, while results for the proposed Gated DSI-ESMDA for varying gating are shown in table 5.4. As in the Egg case, localization did not have a significant impact on observation coverages or data mismatches, and both OC and MSE tended to grow with gating intensity. MSE of the posteriors tended not to be much lower than the priors.

Figure 5.3 – Real field FOPT CDF.



Source: Self-elaboration.

Table 5.3 – DSI-ESMDA results for the real field.

Loc.	Observation coverages		MSE	Data mismatch	
	History	Validation		Avg.	Std. dev.
No	0.63	0.57	2.28	1.24	0.02
Yes	0.63	0.57	2.28	1.24	0.02

Source: Self-elaboration.

Table 5.4 – Gated DSI-ESMDA results for the real field with varying gating.

Gating intensity	Loc.	Observation coverages				MSE		Data mismatch			
		History		Validation				Avg.		Std. dev.	
		Pr.	Po.	Pr.	Po.	Pr.	Po.	Pr.	Po.	Pr.	Po.
$\lambda = 1$	No	0.86	0.53	0.95	0.52	2.71	2.32	5.44	1.21	2.19	0.01
	Yes	0.86	0.54	0.95	0.55	2.71	2.32	5.44	1.21	2.19	0.01
$\lambda = 2$	No	0.86	0.56	0.95	0.55	2.71	2.36	5.44	1.27	2.19	0.02
	Yes	0.86	0.57	0.95	0.58	2.71	2.36	5.44	1.27	2.19	0.02
$\lambda = 4$	No	0.86	0.63	0.95	0.64	2.71	2.44	5.44	1.52	2.19	0.04
	Yes	0.86	0.64	0.95	0.66	2.71	2.44	5.44	1.53	2.19	0.05
$\lambda = 8$	No	0.86	0.73	0.95	0.74	2.71	2.55	5.44	2.15	2.19	0.19
	Yes	0.86	0.73	0.95	0.75	2.71	2.55	5.44	2.16	2.19	0.19
$\lambda = 16$	No	0.86	0.80	0.95	0.81	2.71	2.66	5.44	3.10	2.19	0.57
	Yes	0.86	0.80	0.95	0.81	2.71	2.66	5.44	3.12	2.19	0.57

Source: Self-elaboration.

Pr.: Prior.

Po.: Posterior.

6 Discussion

It's tough to make predictions, especially about the future.

Yogi Berra

6.1 Summary of key findings

Increase in gating intensity tended to increase observation coverage, mean squared error, and data mismatch, albeit with case-dependent sensitivities. Localization yielded mixed results: for the Egg and real field cases, effects on coverage and MSE were minimal, especially for intense enough gatings, while the effect in the Olympus case was drastic. Cumulative distribution plots of the field oil production total suggest that ground-truths are unbiased in the Egg and real field cases, but somewhat biased for the Olympus case. Mild gating intensities (2 in the Egg and Olympus cases, 4 in the real field case) produced similar results to the standard DSI-ESMDA for same localization configuration, while small gating (1 for the Egg case, 1 and 2 for the Olympus and real field cases) had worse coverages than the standard DSI-ESMDA.

6.2 Interpretation of findings

6.2.1 OC, MSE and mismatch increase

That the metrics employed tended to increase with the increase in gating intensity is not surprising, as the objective of Gated DSI-ESMDA lies exactly in increasing posterior variance, i.e., its spread. This naturally leads to higher coverages at the cost of higher mismatch and MSE.

As for the differing standard deviations on the bar plots of the Egg and Olympus cases, they are most probably linked to ground-truth representativeness and reservoir complexity. As indexed by FOPT range and distribution, the Egg case is very simple and the picked ground-truths, relatively well-represented. On the other hand, the Olympus reservoir is much more complex, and picked groundtruths tended to be biased towards low or high oil productions relative to the prior ensemble. Indeed, ground-truths such as 40 and 35 showed FOPT that was close only to a handful realizations. As such, metrics on the Olympus case varied more, and on the Egg case, less.

6.2.2 Mixed effects of localization

Neither the Egg case nor the real field case demonstrated much sensitivity towards localization of the Kalman-like gain, but the Olympus case did. The reason behind this result is interpreted as combination of five factors: reservoir complexity (as indexed by FOPT), ground-truth representativeness, ensemble size, data dimensionality, feature selection, and localization tuning.

For the Egg case, this could be explained by the rationale that the reservoir is simple enough that the number of realizations constitute a rich enough sampling of the prior distribution. Importantly, this is also the case study with the smallest amount of features (table 4.6), and reduced dimensionality help with the estimation of covariances. Lastly, given the narrow range of this ensemble's FOPT and the roughly uniform sampling of ground-truths, it is fair to say no one ground-truth was left unrepresented, which explains the small standard deviations. In short, the Egg reservoir is too well-behaved for localization to be effective.

Conversely, the Olympus reservoir is much more complex while also having a much smaller ensemble and a lot more features, aggravating covariance estimation. Localization was conceived exactly for such situations, so it is not surprising it was effective here.

As for the real case, reservoir complexity and data dimensionality peak, but so does ensemble size. And unlike the Olympus case, the real field should be well-represented, at least FOPT-wise. Rather, the main reason for localization not being effective are more likely to be the adjustment of its parameters (anisotropy, critical lengths and time) and feature selection. Spatial localization, which was not performed, could have helped improve localization. Notably, feature selection by have been inadequate for this reservoir, as indicated by the high MSE and mismatch values. Cases like well I3 BHP in figure C.1, where many if not all reference points are at the outskirts of the prior occurred multiple times, massively impairing the search for an adequate posterior.

In that situation, methods that sequentially assimilate data points, such as the ensemble Kalman filter, might have an easier time matching data. Still, the most adequate thing to do seems to be reselecting the data features used in prediction.

6.2.3 Mild and small gating intensities versus standard DSI-ESMDA

It was commented (section 3.4) that gating intensity reverts the posterior to the prior. If the posterior given by the standard DSI-ESMDA is somewhere in that continuum, then it stands to reason that Gated DSI-ESMDA could approximate it simply by tuning the gating intensity.

Conversely, small gatings might actually produce worse posterior spreads, and exposed in the findings. Since Gated DSI-ESMDA does away with observation data resampling,

small gatings possibly underestimate uncertainty (i.e., likelihood covariance) at least when compared to standard DSI-ESMDA, producing posteriors that are not well-matched to data.

6.3 Evaluation of existing theories and models

Findings are generally in agreement with the theory established in [chapter 3](#), as discussed above. While Gated DSI-ESMDA provides a straightforward, effective mechanism for increasing posterior covariance, too small a gating intensity could produce worse spreads than the standard DSI-ESMDA as doing away with observation data resampling could underestimate uncertainty. This is remediated as the intensity increases, and mild intensities were shown to produce similar results to DSI-ESMDA. How large the intensity should be depends on how tolerable are the increase in MSE and data-mismatch that come as a trade-off.

7 Conclusion

7.1 Summary of findings

This work introduces Gated DSI-ESMDA, an extension of DSI-ESMDA with a nonlinear innovation gating mechanism that increases posterior variance the more intense the gating.

The proposed algorithm was evaluated on two synthetic reservoir ensembles, Egg and Olympus, and on a real field in the Brazilian post-salt. As expected, observation coverages of the posterior increased with the gating intensity for all case studies, albeit under different sensitivities. Results indicate that Gated DSI-ESMDA could do away with observation data resampling as long as gating intensity is not too small, since for mild intensities posteriors are similar to, or better than, the posteriors produced by the standard DSI-ESMDA, at least on the metrics considered.

Despite good results, Gated DSI-ESMDA is not beyond the limitations of its originator. In the real field case, it was exposed that both DSI-ESMDA and Gated DSI-ESMDA are susceptible to poor feature selection and localization tuning, as well as insufficient ground-truth representativeness, high data dimensionality, and reduced ensemble size.

7.2 Contribution to knowledge

This work's novelty lies in the nonlinear innovation gating mechanism for posterior variance increase, which is a direct adaptation of the work of [Fortaleza *et al.* \(2022\)](#).

7.3 Practical implications

By increasing or decreasing a single positive number, named the gating intensity, practitioners can straightforwardly adjust reservoir predictions. As discussed in [chapter 6](#), gating adjustment can be tuned as to match DSI-ESMDA outputs. Practitioners can find and use these values as baselines for data assimilation procedures, increasing or decreasing intensity as they see fit.

More generally, the gating mechanism in of itself can be easily adapted into model-space methods such as ESMDA, acting on outputs of the forward model in order to produce more tempered model parameter estimates, just as it does with predictions in data-space.

7.4 Limitations and future research

As an extension of DSI-ESMDA, Gated DSI-ESMDA shares the same benefits and limitations. Feature selection and localization optimization are thought to be two of the main culprits behind the lukewarm results in the real field case, and future research could address these problems.

7.5 Closing statement

The research questions asked in [section 1.4](#) are answered as follows.

Can a gating mechanism be introduced into DSI-ESMDA to selectively modulate the influence of observations based on their match quality?

Yes, and successfully so. However, how much match quality can be feasibly improved is dependent on parameters within and beyond the proposed approach, such as gating intensity tuning, localization tuning, and feature selection.

Does the proposed innovation gating reduce variance loss during assimilation and produce more robust uncertainty envelopes in the prediction phase without degrading the data match?

Yes. Posterior variance was shown to increase with gating intensity regardless of case study, but this does come at the cost of increased MSE and data-mismatch. Informally speaking, the posterior does “looser” for it, and how much variance should be tolerated change between applications.

How does the performance of Gated DSI-ESMDA compare to standard DSI-ESMDA across synthetic and real reservoir case studies?

For too small a gating intensity, the proposed approach tended to perform slightly worse, but mild gating produced similar results to DSI-ESMDA.

In short, while this work has plenty of room for improvement and exploration, it can be said to have been successful, and future efforts will be concentrated in improving it.

References

- AANONSEN, S. I.; NØVDAL, G.; OLIVER, D. S.; REYNOLDS, A. C.; VALLÈS, B. The ensemble kalman filter in reservoir engineering—a review. **SPE Journal**, v. 14, n. 03, p. 393–412, 09 2009. ISSN 1086-055x. DOI [10.2118/117274-pa](https://doi.org/10.2118/117274-pa). Cit. on pp. 27 and 28.
- ABBATE, E.; ANASTASI, P.; CURZIO, D. D.; FACCHI, G.; ROSSA, E. D. Ensemble data space inversion for fast co2 injection forecast evaluation. *European Association of Geoscientists & Engineers*, v. 2022, n. 1, p. 1–15, 2022. ISSN 2214-4609. DOI <https://doi.org/10.3997/2214-4609.202244022>. Cit. on pp. 27 and 32.
- BARROS, E. G. D.; SZKLARZ, S. P. Fast closed-loop well placement optimization with ensemble-based data-space inversion framework. *In: . [S.l.: s.n.], 2025. (SPE Reservoir Simulation Conference, SPE Reservoir Simulation Conference)*, p. D011s005r003. DOI [10.2118/223849-ms](https://doi.org/10.2118/223849-ms). Cit. on pp. 27 and 32.
- BERTINO, L.; EVENSEN, G.; WACKERNAGEL, H. Sequential data assimilation techniques in oceanography. **International Statistical Review / Revue Internationale de Statistique**, [Wiley, International Statistical Institute (ISI)], v. 71, n. 2, p. 223–241, 2003. ISSN 03067734, 17515823. Cit. on p. 28.
- BURGERS, G.; LEEUWEN, P. J. van; EVENSEN, G. Analysis scheme in the ensemble kalman filter. **Monthly Weather Review**, American Meteorological Society, Boston MA, USA, v. 126, n. 6, p. 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). Cit. on pp. 27 and 28.
- CHEN, Y.; OLIVER, D. S. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. **Mathematical Geosciences**, v. 44, n. 1, p. 1–26, 2012. DOI [10.1007/s11004-011-9376-z](https://doi.org/10.1007/s11004-011-9376-z). Cit. on pp. 27 and 29.
- CHEN, Y.; OLIVER, D. S. Levenberg-marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. **Computational Geosciences**, v. 17, n. 4, p. 689–703, 2013. DOI [10.1007/s10596-013-9351-5](https://doi.org/10.1007/s10596-013-9351-5). Cit. on pp. 27 and 29.
- CHEN, Y.; ZHANG, D. Data assimilation for transient flow in geologic formations via ensemble kalman filter. **Advances in Water Resources**, v. 29, n. 8, p. 1107–1122, 2006. ISSN 0309-1708. DOI <https://doi.org/10.1016/j.advwatres.2005.09.007>. Cit. on pp. 27, 28, and 29.
- de Beer, A.; POWER, A.; WONG, D.; DEKKERS, K.; GRAVATT, M.; BJARKASON, E. K.; O’SULLIVAN, J. P.; O’SULLIVAN, M. J.; MACLAREN, O. J.; NICHOLSON, R. Data space inversion for efficient predictions and uncertainty quantification for geothermal

- models. **Computers & Geosciences**, v. 197, p. 105882, 2025. ISSN 0098-3004. DOI <https://doi.org/10.1016/j.cageo.2025.105882>. Cit. on pp. 27 and 31.
- DONOHO, D. De-noising by soft-thresholding. **IEEE Transactions on Information Theory**, v. 41, n. 3, p. 613–627, 1995. DOI [10.1109/18.382009](https://doi.org/10.1109/18.382009). Cit. on p. 25.
- DONOHO, D. L.; JOHNSTONE, I. M. Ideal spatial adaptation by wavelet shrinkage. **Biometrika**, v. 81, n. 3, p. 425–455, 09 1994. DOI [10.1093/biomet/81.3.425](https://doi.org/10.1093/biomet/81.3.425). Cit. on p. 25.
- An Unconventional Rate Decline Approach for Tight and Fracture-Dominated Gas Wells**, Canadian Unconventional Resources and International Petroleum Conference of **SPE Canada Unconventional Resources Conference**, (SPE Canada Unconventional Resources Conference, Canadian Unconventional Resources and International Petroleum Conference). Spe-137748-ms p. DOI [10.2118/137748-ms](https://doi.org/10.2118/137748-ms). Cit. on p. 32.
- EMERICK, A. A. **Ensemble Data Assimilation Applied to Geological Reservoir Models**. 1. ed. Rio de Janeiro, RJ, Brazil: Universidade Petrobras, 2025. ISBN 978-65-88763-29-2. Cit. on pp. 21, 22, 25, 28, 30, 34, 38, 39, 44, 58, and 62.
- EMERICK, A. A.; REYNOLDS, A. C. History matching time-lapse seismic data using the ensemble kalman filter with multiple data assimilations. **Computational Geosciences**, v. 16, n. 3, p. 639–659, 2012. DOI [10.1007/s10596-012-9275-5](https://doi.org/10.1007/s10596-012-9275-5). Cit. on pp. 27 and 29.
- EMERICK, A. A.; REYNOLDS, A. C. Ensemble smoother with multiple data assimilation. **Computers & Geosciences**, v. 55, p. 3–15, 2013. ISSN 0098-3004. Ensemble Kalman filter for data assimilation. DOI <https://doi.org/10.1016/j.cageo.2012.03.011>. Cit. on pp. 20, 27, and 29.
- EMERICK, A. A.; REYNOLDS, A. C. History-matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation. In: . [S.l.: s.n.], 2013. (SPE Reservoir Simulation Conference, SPE Reservoir Simulation Symposium), p. Spe-163675–ms. DOI [10.2118/163675-ms](https://doi.org/10.2118/163675-ms). Cit. on pp. 27 and 29.
- ESSEN, G. M. van; ZANDVLIET, M. J.; HOF, P. M. J. van den; BOSGRA, O. H.; JANSEN, J. D. Robust Waterflooding Optimization of Multiple Geological Scenarios. **SPE Journal**, v. 14, n. 01, p. 202–210, 03 2009. DOI [10.2118/102913-pa](https://doi.org/10.2118/102913-pa). Cit. on p. 51.
- EVENSEN, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. **Journal of Geophysical Research: Oceans**, v. 99, n. C5, p. 10143–10162, 1994. DOI <https://doi.org/10.1029/94JC00572>. Cit. on pp. 20, 27, and 28.
- EVENSEN, G. Analysis of iterative ensemble smoothers for solving inverse problems. **Computational Geosciences**, v. 22, n. 3, p. 885–908, 2018. DOI [10.1007/s10596-018-9731-y](https://doi.org/10.1007/s10596-018-9731-y). Cit. on pp. 27 and 29.

- FILHO, J. O. d. A. L.; PAULA, T. T. d.; PASTRANA, M. A.; NOZ, D. M. M.; SANTOS, M. G.; FORTALEZA, E. L. F. Robust adaptive model aggregation for well production forecasting. **SPE Journal**, v. 30, n. 07, p. 3984–4006, 07 2025. ISSN 1086-055x. DOI [10.2118/226207-pa](https://doi.org/10.2118/226207-pa). Cit. on pp. 52 and 54.
- FONSECA, R.-M.; ROSSA, E.; EMERICK, A.; HANEA, R.; JANSEN, J.-D. Overview of the olympus field development optimization challenge. In: **ECMOR XVI - 16th European Conference on the Mathematics of Oil Recovery**. [S.l.: s.n.], 2018. v. 2018, p. 1–10. DOI [10.3997/2214-4609.201802246](https://doi.org/10.3997/2214-4609.201802246). Cit. on p. 53.
- FONSECA, R. M.; ROSSA, E. D.; EMERICK, A. A.; HANEA, R. G.; JANSEN, J. D. Introduction to the special issue: Overview of olympus optimization benchmark challenge. **Computational Geosciences**, v. 24, p. 1933–1941, 2020. DOI [10.1007/s10596-020-10003-4](https://doi.org/10.1007/s10596-020-10003-4). Cit. on pp. 25, 53, and 54.
- FORTALEZA, E. L.; GOMES, L. M.; de A. Limaverde Filho, J. O.; de Campos, M. C.; LONGHI, L. G. S.; de Lima, M. L.; TOGNETTI, E. S. Model-free adaptive filter to mitigate actuator wear. **ISA Transactions**, v. 129, p. 493–504, 2022. ISSN 0019-0578. DOI <https://doi.org/10.1016/j.isatra.2022.02.026>. Cit. on pp. 25, 45, and 68.
- GASPARI, G.; COHN, S. E. Construction of correlation functions in two and three dimensions. **Quarterly Journal of the Royal Meteorological Society**, v. 125, n. 554, p. 723–757, 1999. DOI <https://doi.org/10.1002/qj.49712555417>. Cit. on p. 44.
- GEIR, E. On the formulation of the ensemble history-matching problem. In: . [S.l.: s.n.], 2023. (SPE Reservoir Simulation Conference, SPE Reservoir Simulation Conference), p. D031s012r003. DOI [10.2118/212232-ms](https://doi.org/10.2118/212232-ms). Cit. on pp. 27 and 29.
- GRANA, D.; FIGUEIREDO, L. Passos de; AZEVEDO, L. Uncertainty quantification in bayesian inverse problems with model and data dimension reduction. **Geophysics**, v. 84, n. 6, p. M15–m24, Oct. 2019. ISSN 0016-8033. DOI [10.1190/geo2019-0222.1](https://doi.org/10.1190/geo2019-0222.1). Cit. on p. 30.
- HADAMARD, J. Sur les problèmes aux dérivées partielles et leur signification physique. **Princeton university bulletin**, Princeton University, p. 49–52, 1902. Cit. on p. 22.
- HE, X.; JIANG, S.; DURLOFSKY, L. J. Prediction of fault slip tendency in co2 storage using data-space inversion. **Computers & Geosciences**, v. 210, p. 106134, 2026. ISSN 0098-3004. DOI <https://doi.org/10.1016/j.cageo.2026.106134>. Cit. on pp. 27 and 32.
- HOUTEKAMER, P. L.; MITCHELL, H. L. Data assimilation using an ensemble kalman filter technique. **Monthly Weather Review**, American Meteorological Society, Boston MA, USA, v. 126, n. 3, p. 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). Cit. on pp. 27 and 28.

- HOUTEKAMER, P. L.; MITCHELL, H. L. Ensemble kalman filtering. **Quarterly Journal of the Royal Meteorological Society**, v. 131, n. 613, p. 3269–3289, 2005. DOI <https://doi.org/10.1256/qj.05.135>. Cit. on p. 28.
- HOUTEKAMER, P. L.; MITCHELL, H. L.; PELLERIN, G.; BUEHNER, M.; CHARRON, M.; SPACEK, L.; HANSEN, B. Atmospheric data assimilation with an ensemble kalman filter: Results with real observations. **Monthly Weather Review**, American Meteorological Society, Boston MA, USA, v. 133, n. 3, p. 604–620, 2005. DOI [10.1175/mwr-2864.1](https://doi.org/10.1175/mwr-2864.1). Cit. on p. 28.
- HOUTEKAMER, P. L.; ZHANG, F. Review of the ensemble kalman filter for atmospheric data assimilation. **Monthly Weather Review**, American Meteorological Society, Boston MA, USA, v. 144, n. 12, p. 4489–4532, 2016. DOI [10.1175/mwr-d-15-0440.1](https://doi.org/10.1175/mwr-d-15-0440.1). Cit. on p. 28.
- JANSEN, J. D.; FONSECA, R. M.; KAHROBAEI, S.; SIRAJ, M. M.; ESSEN, G. M. van; HOF, P. M. J. van den. The egg model - a geological ensemble for reservoir simulation. **Geoscience Data Journal**, v. 1, n. 2, p. 192–195, 2014. DOI [10.1002/gdj3.21](https://doi.org/10.1002/gdj3.21). Cit. on pp. 25 and 52.
- JIANG, S.; DURLOFSKY, L. J. Data-space inversion using a recurrent autoencoder for time-series parameterization. **Computational Geosciences**, v. 25, n. 1, p. 411–432, 2021. DOI [10.1007/s10596-020-10014-1](https://doi.org/10.1007/s10596-020-10014-1). Cit. on p. 31.
- JIANG, S.; DURLOFSKY, L. J. Treatment of model error in subsurface flow history matching using a data-space method. **Journal of Hydrology**, v. 603, p. 127063, 2021. ISSN 0022-1694. DOI <https://doi.org/10.1016/j.jhydrol.2021.127063>. Cit. on pp. 27 and 32.
- JIANG, S.; DURLOFSKY, L. J. History matching for geological carbon storage using data-space inversion with spatio-temporal data parameterization. **International Journal of Greenhouse Gas Control**, v. 134, p. 104124, 2024. ISSN 1750-5836. DOI <https://doi.org/10.1016/j.ijggc.2024.104124>. Cit. on pp. 27 and 32.
- JIANG, S.; HUI, M.-H.; DURLOFSKY, L. J. Data-space inversion with a recurrent autoencoder for naturally fractured systems. **Frontiers in Applied Mathematics and Statistics**, Volume 7 - 2021, 2021. ISSN 2297-4687. DOI [10.3389/fams.2021.686754](https://doi.org/10.3389/fams.2021.686754). Cit. on p. 31.
- JIANG, S.; SUN, W.; DURLOFSKY, L. J. A data-space inversion procedure for well control optimization and closed-loop reservoir management. **Computational Geosciences**, v. 24, n. 2, p. 361–379, 2020. DOI [10.1007/s10596-019-09853-4](https://doi.org/10.1007/s10596-019-09853-4). Cit. on pp. 27 and 31.
- KALMAN, R. E. A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, v. 82, n. 1, p. 35–45, 03 1960. ISSN 0021-9223. DOI [10.1115/1.3662552](https://doi.org/10.1115/1.3662552). Cit. on pp. 26 and 27.

- KEPPENNE, C. L.; RIENECKER, M. M. Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble kalman filter. **Journal of Marine Systems**, v. 40-41, p. 363–380, 2003. ISSN 0924-7963. The Use of Data Assimilation in Coupled Hydrodynamic, Ecological and Bio-geo-chemical Models of the Ocean. Selected papers from the 33rd International Liege Colloquium on Ocean Dynamics, held in Liege, Belgium on May 7-11th, 2001. DOI [https://doi.org/10.1016/S0924-7963\(03\)00025-3](https://doi.org/10.1016/S0924-7963(03)00025-3). Cit. on p. 28.
- KIM, Y.; MIN, B.; SUN, A.; REN, B.; JEONG, H. Learning-based pattern-data-driven forecast approach for predicting future well responses. **Journal of Petroleum Exploration and Production Technology**, v. 15, n. 2, p. 32, 2025. DOI [10.1007/s13202-025-01937-5](https://doi.org/10.1007/s13202-025-01937-5). Cit. on pp. 27 and 33.
- LEEUEWEN, P. J. van; EVENSEN, G. Data assimilation and inverse methods in terms of a probabilistic formulation. **Monthly Weather Review**, American Meteorological Society, Boston MA, USA, v. 124, n. 12, p. 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). Cit. on pp. 20, 27, and 29.
- LIMA, M. M.; EMERICK, A. A.; ORTIZ, C. E. P. Data-space inversion with ensemble smoother. **Computational Geosciences**, v. 24, n. 3, p. 1179–1200, Jun. 2020. ISSN 1573-1499. DOI [10.1007/s10596-020-09933-w](https://doi.org/10.1007/s10596-020-09933-w). Cit. on pp. 20, 27, 31, 43, and 44.
- LIN, H.; ZHOU, F.; XIAO, C.; YANG, X.; WANG, Y.; ZHANG, Y.; HOU, T. Data-driven inversion-free workflow of well performance forecast under uncertainty for fractured shale gas reservoirs. **Journal of Energy Resources Technology**, v. 145, n. 7, p. 072603, 02 2023. ISSN 0195-0738. DOI [10.1115/1.4055537](https://doi.org/10.1115/1.4055537). Cit. on p. 32.
- LIU, D.; RAO, X.; ZHAO, H.; XU, Y.-F.; GONG, R.-X. An improved data space inversion method to predict reservoir state fields via observed production data. **Petroleum Science**, v. 18, n. 4, p. 1127–1142, 2021. ISSN 1995-8226. DOI <https://doi.org/10.1016/j.petsci.2021.07.008>. Cit. on p. 31.
- LIU, G.; CHEN, Y.; ZHANG, D. Investigation of flow and transport processes at the made site using ensemble kalman filter. **Advances in Water Resources**, v. 31, n. 7, p. 975–986, 2008. ISSN 0309-1708. DOI <https://doi.org/10.1016/j.advwatres.2008.03.006>. Cit. on pp. 28 and 29.
- LIU, W.; XU, Y.; RAO, X.; LIU, D.; ZHAO, H. Reservoir closed-loop optimization method based on connection elements and data space inversion with variable controls. **Physics of Fluids**, v. 35, n. 11, p. 113322, Nov. 2023. ISSN 1070-6631. DOI [10.1063/5.0172378](https://doi.org/10.1063/5.0172378). Cit. on p. 31.
- LUO, X.; STORDAL, A. S.; LORENTZEN, R. J.; NÆVDAL, G. Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: Theory

- and applications. **SPE Journal**, v. 20, n. 05, p. 962–982, Oct. 2015. ISSN 1086-055x. DOI [10.2118/176023-pa](https://doi.org/10.2118/176023-pa). Cit. on pp. 27 and 29.
- MASCHIO, C.; MALEKI, M.; DAVOLIO, A.; SCHIOZER, D. J. Methodology to calibrate weight between production and 4d seismic in data assimilation to improve reservoir performance forecasting. **Geoenergy Science and Engineering**, v. 243, p. 213349, 2024. ISSN 2949-8910. DOI <https://doi.org/10.1016/j.geoen.2024.213349>. Cit. on pp. 27 and 29.
- MØYNER, O. Jutuldarcy.jl - a fully differentiable high-performance reservoir simulator based on automatic differentiation. **Computational Geosciences**, Springer, v. 29, n. 30, 2025. Cit. on p. 51.
- NÆVDAL, G.; MANNSETH, T.; VEFRING, E. H. Near-well reservoir monitoring through ensemble kalman filter. In: . [S.l.: s.n.], 2002. (SPE Improved Oil Recovery Conference, SPE/DOE Improved Oil Recovery Symposium), p. Spe-75235–ms. DOI [10.2118/75235-ms](https://doi.org/10.2118/75235-ms). Cit. on pp. 27 and 28.
- OLIVER, D. S.; CHEN, Y. Recent progress on reservoir history matching: a review. **Computational Geosciences**, v. 15, n. 1, p. 185–221, Jan. 2011. ISSN 1573-1499. DOI [10.1007/s10596-010-9194-2](https://doi.org/10.1007/s10596-010-9194-2). Cit. on pp. 27 and 28.
- OTSUKA, S.; KOTSUKI, S.; MIYOSHI, T. Nowcasting with data assimilation: A case of global satellite mapping of precipitation. **Weather and Forecasting**, American Meteorological Society, Boston MA, USA, v. 31, n. 5, p. 1409–1416, 2016. DOI [10.1175/waf-d-16-0039.1](https://doi.org/10.1175/waf-d-16-0039.1). Cit. on p. 28.
- RAANES, P. N.; STORDAL, A. S.; EVENSEN, G. Revising the stochastic iterative ensemble smoother. **Nonlinear Processes in Geophysics**, v. 26, n. 3, p. 325–338, 2019. DOI [10.5194/npg-26-325-2019](https://doi.org/10.5194/npg-26-325-2019). Cit. on pp. 27 and 29.
- RASMUSSEN, A. F.; SANDVE, T. H.; BAO, K.; LAUSER, A.; HOVE, J.; SKAFLESTAD, B.; KLÖFKORN, R.; BLATT, M.; RUSTAD, A. B.; SÆVAREID, O.; LIE, K.-A.; THUNE, A. The open porous media flow reservoir simulator. **Computers & Mathematics with Applications**, v. 81, p. 159–185, 2021. ISSN 0898-1221. Development and Application of Open-source Software for Problems with Numerical PDEs. DOI <https://doi.org/10.1016/j.camwa.2020.05.014>. Cit. on p. 35.
- REICHLE, R. H.; MCLAUGHLIN, D. B.; ENTEKHABI, D. Hydrologic data assimilation with the ensemble kalman filter. **Monthly Weather Review**, American Meteorological Society, Boston MA, USA, v. 130, n. 1, p. 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). Cit. on pp. 28 and 29.
- SATIJA, A.; CAERS, J. Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component

- space. **Advances in Water Resources**, v. 77, p. 69–81, 2015. ISSN 0309-1708. DOI <https://doi.org/10.1016/j.advwatres.2015.01.002>. Cit. on pp. 27 and 30.
- SATIJA, A.; SCHEIDT, C.; LI, L.; CAERS, J. Direct forecasting of reservoir performance using production data without history matching. **Computational Geosciences**, v. 21, n. 2, p. 315–333, 2017. DOI [10.1007/s10596-017-9614-7](https://doi.org/10.1007/s10596-017-9614-7). Cit. on pp. 27 and 30.
- SCHEIDT, C.; RENARD, P.; CAERS, J. Prediction-focused subsurface modeling: Investigating the need for accuracy in flow-based inverse modeling. **Mathematical Geosciences**, v. 47, n. 2, p. 173–191, 2015. DOI [10.1007/s11004-014-9521-6](https://doi.org/10.1007/s11004-014-9521-6). Cit. on pp. 27 and 30.
- Silva Neto, G. M.; DAVOLIO, A.; SCHIOZER, D. J. Assimilating time-lapse seismic data in the presence of significant spatially correlated model errors. **Journal of Petroleum Science and Engineering**, v. 207, p. 109127, 2021. ISSN 0920-4105. DOI <https://doi.org/10.1016/j.petrol.2021.109127>. Cit. on pp. 27 and 30.
- SINGH, G.; HANEA, R.; SADIGOV, S.; MÜLLER, T.; BARROS, E.; SZKLARZ, S. Application of ensemble-based forecast calibration with data-space inversion for production optimization. In: . [S.l.: s.n.], 2025. (SPE Reservoir Simulation Conference, SPE Reservoir Simulation Conference), p. D012s001r004. DOI [10.2118/223878-ms](https://doi.org/10.2118/223878-ms). Cit. on pp. 27 and 32.
- SKJERVHEIM, J.-A.; EVENSEN, G.; HOVE, J.; VABØ, J. G. An ensemble smoother for assisted history matching. In: **SPE Reservoir Simulation Symposium**. The Woodlands, TX, USA: Spe, 2011. p. Spe-141929–ms. DOI [10.2118/141929-ms](https://doi.org/10.2118/141929-ms). Cit. on pp. 20, 27, and 29.
- SUN, L.; SEIDOU, O.; NISTOR, I.; LIU, K. Review of the kalman-type hydrological data assimilation. **Hydrological Sciences Journal**, Taylor & Francis, v. 61, n. 13, p. 2348–2366, 2016. DOI [10.1080/02626667.2015.1127376](https://doi.org/10.1080/02626667.2015.1127376). Cit. on pp. 27, 28, and 29.
- SUN, W.; DURLOFSKY, L. J. A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. **Mathematical Geosciences**, v. 49, n. 6, p. 679–715, 2017. DOI [10.1007/s11004-016-9672-8](https://doi.org/10.1007/s11004-016-9672-8). Cit. on p. 20.
- SUN, W.; DURLOFSKY, L. J. A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. **Mathematical Geosciences**, v. 49, n. 6, p. 679–715, 2017. DOI [10.1007/s11004-016-9672-8](https://doi.org/10.1007/s11004-016-9672-8). Cit. on pp. 30 and 31.
- SUN, W.; DURLOFSKY, L. J. Data-space approaches for uncertainty quantification of co2 plume location in geological carbon storage. **Advances in Water Resources**, v. 123, p. 234–255, 2019. ISSN 0309-1708. DOI <https://doi.org/10.1016/j.advwatres.2018.10.028>. Cit. on pp. 27 and 31.
- SUN, W.; HUI, M.-H.; DURLOFSKY, L. J. Production forecasting and uncertainty quantification for naturally fractured reservoirs using a new data-space inversion procedure.

- Computational Geosciences**, v. 21, n. 5, p. 1443–1458, 2017. DOI [10.1007/s10596-017-9633-4](https://doi.org/10.1007/s10596-017-9633-4). Cit. on pp. 27 and 31.
- SZKLARZ, S.; BARROS, E. Accelerating ensemble-based well control optimization with es-mda data-space inversion framework. *European Association of Geoscientists & Engineers*, v. 2022, n. 1, p. 1–11, 2022. ISSN 2214-4609. DOI <https://doi.org/10.3997/2214-4609.202244075>. Cit. on p. 32.
- SZUNYOGH, I.; KOSTELICH, E. J.; GYARMATI, G.; PATIL, D. J.; HUNT, B. R.; KALNAY, E.; OTT, E.; YORKE, J. A. Assessing a local ensemble kalman filter: perfect model experiments with the national centers for environmental prediction global model. **Tellus B: Chemical and Physical Meteorology**, Jan. 2005. DOI [10.3402/tellusa.v57i4.14721](https://doi.org/10.3402/tellusa.v57i4.14721). Cit. on p. 28.
- TAO, Z.; LIU, P.; ZHANG, K.; FENG, G.; ZHANG, J.; ZHANG, L.; YANG, Y.; YAO, J. Estimating groundwater pollution source parameters and hydraulic conductivity fields through deep learning-aided surrogate modeling. **Hydrogeology Journal**, v. 33, n. 8, p. 2145–2161, 2025. DOI [10.1007/s10040-025-02975-3](https://doi.org/10.1007/s10040-025-02975-3). Cit. on pp. 26, 27, and 30.
- WHITAKER, J. S.; HAMILL, T. M.; WEI, X.; SONG, Y.; TOTH, Z. Ensemble data assimilation with the ncep global forecast system. **Monthly Weather Review**, American Meteorological Society, Boston MA, USA, v. 136, n. 2, p. 463–482, 2008. DOI [10.1175/2007mwr2018.1](https://doi.org/10.1175/2007mwr2018.1). Cit. on pp. 27 and 28.
- XIAO, C.; WANG, G.; ZHANG, Y.; DENG, Y. Machine-learning-based well production prediction under geological and hydraulic fracture parameters uncertainty for unconventional shale gas reservoirs. **Journal of Natural Gas Science and Engineering**, v. 106, p. 104762, 2022. ISSN 1875-5100. DOI <https://doi.org/10.1016/j.jngse.2022.104762>. Cit. on pp. 27 and 33.
- XIAO, C.; ZHANG, S.; HU, Y.; GU, X.; MA, X.; ZHOU, T.; JIN, J. Robust optimization of geoenergy production using data-driven deep recurrent auto-encoder and fully-connected neural network proxy. **Expert Systems with Applications**, v. 242, p. 122797, 2024. ISSN 0957-4174. DOI <https://doi.org/10.1016/j.eswa.2023.122797>. Cit. on p. 26.
- XIAO, C.; ZHANG, S.; MA, X.; ZHOU, T.; HOU, T.; CHEN, F. Data-driven model predictive control for closed-loop refracturing design and optimization in naturally fractured shale gas reservoir under geological uncertainty. **Computers & Chemical Engineering**, v. 169, p. 108096, 2023. ISSN 0098-1354. DOI <https://doi.org/10.1016/j.compchemeng.2022.108096>. Cit. on p. 32.
- XIAO, C.; ZHANG, S.; MA, X.; ZHOU, T.; HOU, T.; CHEN, F. Deep-learning-generalized data-space inversion and uncertainty quantification framework for accelerating geological

- co2 plume migration monitoring. **Geoenergy Science and Engineering**, v. 224, p. 211627, 2023. ISSN 2949-8910. DOI <https://doi.org/10.1016/j.geoen.2023.211627>. Cit. on pp. 27 and 32.
- XIAO, C.; ZHANG, S.; MA, X.; ZHOU, T.; HOU, T.; CHEN, F. Robust production forecast and uncertainty quantification for waterflooding reservoir using hybrid recurrent auto-encoder and long short-term memory neural network. **Geoenergy Science and Engineering**, v. 230, p. 212252, 2023. ISSN 2949-8910. DOI <https://doi.org/10.1016/j.geoen.2023.212252>. Cit. on p. 33.
- ZANDVLIET, M. J.; BOSGRA, O. H.; JANSEN, J. D.; HOF, P. M. J. van den; KRAAIJEVANGER, J. F. B. M. Bang-bang control and singular arcs in reservoir flooding. **Journal of Petroleum Science and Engineering**, v. 58, n. 1, p. 186–200, 2007. ISSN 0920-4105. DOI [10.1016/j.petrol.2006.12.008](https://doi.org/10.1016/j.petrol.2006.12.008). Cit. on p. 51.
- ZHAN, W.; LIU, D.; ZHAO, H.; ZHOU, Y.; RAO, X. An integrated workflow of history matching and production prediction for fractured horizontal wells. **Physics of Fluids**, v. 36, n. 11, p. 113379, Nov. 2024. ISSN 1070-6631. DOI [10.1063/5.0244253](https://doi.org/10.1063/5.0244253). Cit. on p. 31.
- ZHANG, J.; LIN, G.; LI, W.; WU, L.; ZENG, L. An iterative local updating ensemble smoother for estimation and uncertainty assessment of hydrologic model parameters with multi-modal distributions. **Water Resources Research**, v. 54, n. 3, p. 1716–1733, 2018. DOI <https://doi.org/10.1002/2017WR020906>. Cit. on pp. 27 and 29.

Appendices

Appendix A – Academic productions

The master's program in Mechanical Engineering at University of Brasília requires the completion of a minimum of 360 hours of coursework (24 credits) and at least one publication in a national or international conference/journal. This appendix exhibits the academic productions resulting from the research work developed by the author during the course of the master's degree.

A.1 Disciplines completed

The author has completed the following disciplines, which are summarized in [table A.1](#).

Table A.1 – Summary of the disciplines completed by the author during their master's degree.

Start	End	Code	Name	Hours	Credits
10/2024	05/2026	PCMEC0073	SEMINÁRIO 1 EM CIÊNCIAS MECÂNICAS	60h	4
10/2024	02/2025	PCMEC0076	PROCESSAMENTO DE SINAIS E SISTEMAS MECÂNICOS LINEARES	60h	4
10/2024	02/2025	PCMEC2158	MÉTODOS NUMÉRICOS EM CIÊNCIAS MECÂNICAS	60h	4
10/2024	02/2025	PCMEC0091	TÉCNICAS NÃO LINEARES EM CONTROLE*	60h	4
03/2025	08/2025	PCMEC0147	VIBRAÇÕES MECÂNICAS	60h	4
03/2025	07/2025	PCMEC0192	TERMODINÂMICA CLÁSSICA	60h	4
08/2025	12/2025	PCMEC0180	ESTÁGIO DE DOCÊNCIA EM CIÊNCIAS MECÂNICAS 1	60h	4
08/2025	12/2025	PPGEE2249	APRENDIZADO DE MÁQUINA	60h	4
03/2026	07/2025	PPGEE2166	PROCESSOS ESTOCÁSTICOS	60h	4
Total (minimum: 360h / 24 credits)				540h	36

Source: Self-elaboration.

*: This discipline was completed during the author's undergraduate studies, but is a graduate-level discipline.

A.2 Publications

Papers the author has published during their master's degree are summarized in [table A.2](#).

Table A.2 – Summary of the author’s publications during their master’s degree.

Type	Title	Event	Submission
National conference paper	Custos de Operação e Manutenção Associados ao Serviço de Flexibilidade das Hidrelétricas	XI Simpósio Brasileiro de Sistemas Elétricos (SBSE)	March 17th, 2025
National conference paper	Error-Based Active Disturbance Rejection Control for Parkinson’s Disease Tremor Suppression in Wrist Considering Upper Limb Dynamics	XXVI Congresso Brasileiro de Automática (CBA)	May 2nd, 2026
International conference paper	Nonlinear Innovation Gating for Variance Preservation in DSI-ESMDA Forecasting	20th European Conference on the Mathematics of Geological Reservoirs (ECMOR)	July 11th, 2026
Total (minimum: 1)			3

Source: Self-elaboration.

A.3 Software

During their master’s degree, the author participated in a research project with Axia Energia (then Eletrobras Furnas), which produced a software for the estimation of the opportunity cost of hidreletric flexibility. Details of the software are provided in [table A.3](#).

Table A.3 – Summary of the software developed by the author during their master’s degree.

Title	Cálculo do Custo de Oportunidade de Usinas Hidrelétricas
Summary	Library for estimating the opportunity cost associated with the flexibility of hydroelectric plants, through deterministic models, to be used as a decision support tool for electric grid operators.
Company concerns being addressed	Renewable energies, predictability, cost assessment
Creation date	20/10/2025
Hash algorithm	SHA-512
Hash	3a24c572e979a5f8897f4787a33fac173d7fc3d5a197bb79d640fc26423744d640f1eec72e082973b85af7c5f555bc20294b47a5e05479d2f06cf9ad5e5c3da5
Language	Python
Application fields	Energy politics, Technology and energy (energy generation)
Kind of program	Library, support tool, simulation and modeling, techno-scientific applications
Authors	Thiago Tomás de Paula , Yan Tavares de Oliveira, Matheus Bawden Silverio de Castro, Matheus Teixeira de Sousa, Eduardo Bezerra Rufino Ferreira Paiva, José Oniram de Aquino Limaverde Filho, William Humberto Cuéllar Sánchez, Thiago Felício de Souza, Eugênio Libório Feitosa Fortaleza

Source: Self-elaboration.

Appendix B – Proof of the simplified Kalman gain formula

This appendix presents a proof for the equivalence between [equation \(3.37\)](#) and [equation \(3.38\)](#). It suffices to use the below identity substituting $\mathbf{U} = (\mathbf{C}_p^0)^{-1}$, $\mathbf{V} = (\mathbf{C}_e^\ell)^{-1}$, $\mathbf{W} = \mathbf{H}$.

Theorem B.1. For $\mathbf{U} \in \mathbb{R}^{n \times n}$ and $\mathbf{V} \in \mathbb{R}^{m \times m}$ invertible matrices, and $\mathbf{W} \in \mathbb{R}^{m \times n}$,

$$(\mathbf{U} + \mathbf{W}^\top \mathbf{V} \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{V} = \mathbf{U}^{-1} \mathbf{W}^\top (\mathbf{V}^{-1} + \mathbf{W} \mathbf{U}^{-1} \mathbf{W}^\top)^{-1}. \quad (\text{B.1})$$

Proof. By the property of matrix inversion $(\mathbf{X}\mathbf{Y})^{-1} = \mathbf{Y}^{-1}\mathbf{X}^{-1}$, with both $\mathbf{X}$ and $\mathbf{Y}$ invertible,

$$(\mathbf{U} + \mathbf{W}^\top \mathbf{V} \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{V} = [(\mathbf{I} + \mathbf{W}^\top \mathbf{V} \mathbf{W} \mathbf{U}^{-1}) \mathbf{U}]^{-1} \mathbf{W}^\top \mathbf{V} \quad (\text{B.2})$$

$$= \mathbf{U}^{-1} (\mathbf{I} + \mathbf{W}^\top \mathbf{V} \mathbf{W} \mathbf{U}^{-1})^{-1} \mathbf{W}^\top \mathbf{V}. \quad (\text{B.3})$$

Set $\mathbf{Z}_1 = \mathbf{W}^\top$, $\mathbf{Z}_2 = \mathbf{V} \mathbf{W} \mathbf{U}^{-1}$. Then

$$\mathbf{Z}_1 + \mathbf{Z}_1 \mathbf{Z}_2 \mathbf{Z}_1 = \mathbf{Z}_1 (\mathbf{I} + \mathbf{Z}_2 \mathbf{Z}_1) = (\mathbf{I} + \mathbf{Z}_1 \mathbf{Z}_2) \mathbf{Z}_1 \quad (\text{B.4})$$

$$(\mathbf{I} + \mathbf{Z}_1 \mathbf{Z}_2)^{-1} \mathbf{Z}_1 = \mathbf{Z}_1 (\mathbf{I} + \mathbf{Z}_2 \mathbf{Z}_1)^{-1} \quad (\text{B.5})$$

or

$$(\mathbf{I} + \mathbf{W}^\top \mathbf{V} \mathbf{W} \mathbf{U}^{-1})^{-1} \mathbf{W}^\top = \mathbf{W}^\top (\mathbf{I} + \mathbf{V} \mathbf{W} \mathbf{U}^{-1} \mathbf{W}^\top)^{-1}. \quad (\text{B.6})$$

Substituting [equation \(B.6\)](#) in [equation \(B.3\)](#) and using the property of the inverse again,

$$(\mathbf{U} + \mathbf{W}^\top \mathbf{V} \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{V} = \mathbf{U}^{-1} (\mathbf{I} + \mathbf{W}^\top \mathbf{V} \mathbf{W} \mathbf{U}^{-1})^{-1} \mathbf{W}^\top \mathbf{V}. \quad (\text{B.7})$$

$$(\mathbf{U} + \mathbf{W}^\top \mathbf{V} \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{V} = \mathbf{U}^{-1} \mathbf{W}^\top (\mathbf{I} + \mathbf{V} \mathbf{W} \mathbf{U}^{-1} \mathbf{W}^\top)^{-1} \mathbf{V} \quad (\text{B.8})$$

$$= \mathbf{U}^{-1} \mathbf{W}^\top (\mathbf{I} + \mathbf{V} \mathbf{W} \mathbf{U}^{-1} \mathbf{W}^\top)^{-1} (\mathbf{V}^{-1})^{-1} \quad (\text{B.9})$$

$$= \mathbf{U}^{-1} \mathbf{W}^\top (\mathbf{V}^{-1} + \mathbf{W} \mathbf{U}^{-1} \mathbf{W}^\top)^{-1}. \quad (\text{B.10})$$

□

Appendix C – Detailed results

This appendix expounds the results shown in [chapter 5](#) by presenting tables and figures whose placement within the main body of the manuscript would have been unwieldy.

For the synthetic Egg ([section C.1](#)) and Olympus ([section C.2](#)) ensembles, the observation coverage (OC), mean squared error (MSE), and data mismatch of the posterior distributions are tabled for each experiment conducted.

Since the OC, MSE and mismatch of the real field have already been presented in the main body of the manuscript, they are not repeated here. Instead, [section C.3](#) presents the plots of the posterior ensemble generated by running the proposed Gated DSI-ESMDA algorithm with a gating intensity of $\lambda = 4$. As [tables 5.3](#) and [5.4](#) indicate, this level of gating produces results similar to those using the standard DSI-ESMDA, albeit with a slightly better coverage in exchange of a slightly increased mismatch.

C.1 Egg ensemble

Table C.1 – DSI-ESMDA.

Gt.	Loc.	OC		MSE	Data mismatch	
		His.	Val.		Avg.	Std. dev.
10	No	0.34	1.00	0.96	0.50	5.1×10^{-3}
	Yes	0.48	1.00	0.96	0.55	0.03
20	No	0.36	0.98	0.93	0.49	4.8×10^{-3}
	Yes	0.52	1.00	0.93	0.53	0.02
30	No	0.36	1.00	0.94	0.49	4.6×10^{-3}
	Yes	0.53	0.90	0.94	0.57	0.03
40	No	0.35	1.00	0.96	0.50	5.0×10^{-3}
	Yes	0.52	1.00	0.96	0.55	0.02
50	No	0.35	0.90	0.94	0.50	5.2×10^{-3}
	Yes	0.53	1.00	0.94	0.55	0.03
60	No	0.36	0.94	0.94	0.49	4.7×10^{-3}
	Yes	0.53	0.98	0.94	0.55	0.03
70	No	0.34	1.00	0.95	0.50	5.6×10^{-3}
	Yes	0.53	1.00	0.95	0.55	0.02
80	No	0.36	0.90	0.94	0.50	5.3×10^{-3}
	Yes	0.56	1.00	0.94	0.58	0.04
90	No	0.36	0.96	0.95	0.50	4.9×10^{-3}
	Yes	0.55	1.00	0.95	0.59	0.05
100	No	0.35	1.00	0.93	0.49	4.7×10^{-3}
	Yes	0.52	1.00	0.93	0.53	0.02

Source: Self-elaboration.

Table C.2 – Gated DSI-ESMDA ($\lambda = 1$).

	OC		MSE	Data mismatch	
	His.	Val.		Avg.	Std. dev.
	0.22	0.98	0.97	0.49	3.3×10^{-3}
	0.46	1.00	0.97	0.55	0.03
	0.25	0.81	0.93	0.48	2.8×10^{-3}
	0.48	1.00	0.93	0.52	0.02
	0.24	0.95	0.94	0.48	2.7×10^{-3}
	0.48	0.90	0.95	0.56	0.03
	0.23	0.86	0.96	0.49	3.1×10^{-3}
	0.48	1.00	0.97	0.55	0.02
	0.24	0.86	0.94	0.48	2.9×10^{-3}
	0.49	1.00	0.95	0.55	0.02
	0.25	0.86	0.94	0.48	2.6×10^{-3}
	0.49	0.97	0.95	0.54	0.03
	0.23	0.91	0.96	0.49	3.4×10^{-3}
	0.48	1.00	0.96	0.54	0.02
	0.23	0.82	0.95	0.48	2.9×10^{-3}
	0.51	1.00	0.95	0.57	0.04
	0.22	0.65	0.96	0.49	2.9×10^{-3}
	0.51	0.99	0.96	0.58	0.05
	0.24	0.92	0.94	0.48	2.7×10^{-3}
	0.47	1.00	0.94	0.53	0.02

Source: Self-elaboration.

Table C.3 – Gated DSI-ESMDA ($\lambda = 2$).

Gt.	Loc.	OC		MSE	Data mismatch	
		His.	Val.		Avg.	Std. dev.
10	No	0.41	1.00	0.97	0.52	0.01
	Yes	0.57	1.00	0.97	0.60	0.02
20	No	0.43	1.00	0.94	0.51	0.01
	Yes	0.58	1.00	0.94	0.58	0.02
30	No	0.39	0.99	0.95	0.51	0.01
	Yes	0.55	0.90	0.95	0.61	0.03
40	No	0.42	1.00	0.97	0.53	0.01
	Yes	0.57	1.00	0.97	0.60	0.02
50	No	0.43	0.90	0.95	0.51	0.01
	Yes	0.59	1.00	0.95	0.60	0.02
60	No	0.41	0.95	0.94	0.51	0.01
	Yes	0.57	0.96	0.95	0.60	0.02
70	No	0.40	1.00	0.96	0.52	0.01
	Yes	0.57	1.00	0.97	0.60	0.02
80	No	0.40	0.93	0.95	0.50	0.01
	Yes	0.59	1.00	0.95	0.60	0.03
90	No	0.38	0.94	0.96	0.51	9.3×10^{-3}
	Yes	0.56	0.99	0.97	0.61	0.04
100	No	0.43	1.00	0.94	0.51	9.3×10^{-3}
	Yes	0.58	1.00	0.94	0.58	0.02

Source: Self-elaboration.

Table C.4 – Gated DSI-ESMDA ($\lambda = 4$).

	Loc.	OC		MSE	Data mismatch	
		His.	Val.		Avg.	Std. dev.
	No	0.73	1.00	0.99	0.79	0.06
	Yes	0.80	1.00	0.99	0.93	0.06
	No	0.74	1.00	0.95	0.79	0.05
	Yes	0.81	1.00	0.95	0.92	0.08
	No	0.68	0.97	0.96	0.75	0.05
	Yes	0.76	0.89	0.97	0.92	0.06
	No	0.72	1.00	0.98	0.79	0.05
	Yes	0.80	1.00	0.98	0.92	0.06
	No	0.73	1.00	0.95	0.77	0.05
	Yes	0.79	1.00	0.96	0.93	0.07
	No	0.72	0.97	0.95	0.78	0.05
	Yes	0.80	0.96	0.96	0.93	0.06
	No	0.71	1.00	0.97	0.79	0.06
	Yes	0.78	1.00	0.98	0.93	0.08
	No	0.71	0.95	0.96	0.70	0.08
	Yes	0.79	1.00	0.97	0.85	0.07
	No	0.66	0.98	0.98	0.68	0.07
	Yes	0.74	1.00	1.00	0.82	0.07
	No	0.74	1.00	0.95	0.78	0.06
	Yes	0.81	1.00	0.95	0.91	0.09

Source: Self-elaboration.

Table C.5 – Gated DSI-ESMDA ($\lambda = 8$).

Gt.	Loc.	OC		MSE	Data mismatch	
		His.	Val.		Avg.	Std. dev.
10	No	0.89	1.00	1.02	1.93	0.41
	Yes	0.92	1.00	1.01	2.16	0.51
20	No	0.90	1.00	0.96	1.80	0.40
	Yes	0.93	1.00	0.96	1.99	0.47
30	No	0.84	0.93	0.99	1.82	0.30
	Yes	0.87	0.90	0.99	2.10	0.36
40	No	0.89	1.00	1.01	1.79	0.38
	Yes	0.93	1.00	1.01	1.98	0.44
50	No	0.87	1.00	0.98	1.85	0.38
	Yes	0.90	1.00	0.98	2.09	0.45
60	No	0.88	0.97	0.97	1.82	0.32
	Yes	0.92	0.96	0.97	2.04	0.38
70	No	0.87	1.00	1.00	1.84	0.42
	Yes	0.90	1.00	1.00	2.05	0.49
80	No	0.87	0.97	0.99	1.86	0.30
	Yes	0.90	1.00	0.99	2.15	0.30
90	No	0.82	0.99	1.02	1.74	0.32
	Yes	0.86	0.99	1.05	1.98	0.33
100	No	0.90	1.00	0.96	1.76	0.41
	Yes	0.92	1.00	0.96	1.95	0.48

Source: Self-elaboration.

Table C.6 – Gated DSI-ESMDA ($\lambda = 16$).

	Loc.	OC		MSE	Data mismatch	
		His.	Val.		Avg.	Std. dev.
	No	0.92	1.00	1.03	4.10	1.54
	Yes	0.93	1.00	1.03	4.32	1.66
	No	0.92	1.00	0.97	3.40	1.14
	Yes	0.94	1.00	0.97	3.57	1.22
	No	0.88	0.92	1.00	4.38	1.39
	Yes	0.89	0.91	0.99	4.73	1.50
	No	0.91	1.00	1.02	3.62	1.35
	Yes	0.94	1.00	1.02	3.82	1.44
	No	0.91	1.00	1.00	3.96	1.30
	Yes	0.92	1.00	0.99	4.21	1.39
	No	0.91	0.98	0.98	3.79	1.26
	Yes	0.92	0.98	0.98	4.02	1.37
	No	0.90	1.00	1.02	3.78	1.32
	Yes	0.92	1.00	1.02	3.98	1.41
	No	0.91	1.00	1.01	5.73	1.69
	Yes	0.92	1.00	1.00	6.18	1.90
	No	0.85	0.99	1.05	6.07	1.16
	Yes	0.87	0.99	1.05	6.53	1.28
	No	0.92	1.00	0.98	3.39	1.14
	Yes	0.93	1.00	0.97	3.56	1.22

Source: Self-elaboration.

C.2 Olympus ensemble

Table C.7 – DSI-ESMDA.

Gt.	Loc.	OC		MSE	Data mismatch	
		His.	Val.		Avg.	Std. dev.
5	No	0.28	0.26	2.70	1.42	0.01
	Yes	0.66	0.76	2.29	17.80	18.21
10	No	0.40	0.59	1.83	0.98	0.02
	Yes	0.59	0.85	1.56	1.30	0.29
15	No	0.41	0.26	2.05	1.10	0.02
	Yes	0.70	0.89	1.75	1.49	0.33
20	No	0.42	0.63	1.11	0.62	0.02
	Yes	0.57	0.86	1.09	0.72	0.06
25	No	0.36	0.27	1.64	0.89	0.02
	Yes	0.71	0.91	1.41	5.49	4.78
30	No	0.43	0.78	1.15	0.63	0.01
	Yes	0.60	0.87	1.10	0.75	0.09
35	No	0.42	0.45	1.49	0.81	0.01
	Yes	0.59	0.92	1.34	1.05	0.18
40	No	0.45	0.56	1.41	0.77	0.01
	Yes	0.72	0.88	1.24	1.76	0.74
45	No	0.39	0.34	2.07	1.10	0.01
	Yes	0.64	0.93	1.82	1.99	0.73
50	No	0.45	0.61	1.18	0.65	0.01
	Yes	0.70	0.91	1.12	1.05	0.23

Source: Self-elaboration.

Table C.8 – Gated DSI-ESMDA ($\lambda = 1$).

	Gt.	Loc.	OC		MSE	Data mismatch	
			His.	Val.		Avg.	Std. dev.
	5	No	0.06	0.08	2.71	1.36	2.8×10^{-3}
		Yes	0.64	0.75	2.24	17.76	18.21
	10	No	0.18	0.26	1.84	0.93	6.0×10^{-3}
		Yes	0.53	0.85	1.60	1.26	0.30
	15	No	0.16	0.12	2.06	1.05	8.7×10^{-3}
		Yes	0.64	0.88	1.74	1.45	0.34
	20	No	0.20	0.43	1.12	0.58	8.0×10^{-3}
		Yes	0.49	0.84	1.12	0.69	0.04
	25	No	0.11	0.08	1.65	0.83	2.5×10^{-3}
		Yes	0.68	0.91	1.38	5.42	4.77
	30	No	0.21	0.58	1.15	0.59	7.1×10^{-3}
		Yes	0.46	0.87	1.11	0.72	0.08
	35	No	0.18	0.21	1.50	0.76	6.0×10^{-3}
		Yes	0.49	0.91	1.37	1.01	0.18
	40	No	0.15	0.22	1.41	0.71	3.3×10^{-3}
		Yes	0.67	0.88	1.24	1.70	0.72
	45	No	0.13	0.15	2.07	1.05	4.6×10^{-3}
		Yes	0.57	0.92	1.82	1.94	0.72
	50	No	0.19	0.25	1.19	0.60	3.8×10^{-3}
		Yes	0.64	0.91	1.13	1.00	0.23

Source: Self-elaboration.

Table C.9 – Gated DSI-ESMDA ($\lambda = 2$).

Gt.	Loc.	OC		MSE	Data mismatch	
		His.	Val.		Avg.	Std. dev.
5	No	0.16	0.14	2.71	1.37	5.7×10^{-3}
	Yes	0.66	0.76	2.26	17.80	18.21
10	No	0.29	0.47	1.84	0.96	0.01
	Yes	0.57	0.87	1.62	1.31	0.30
15	No	0.30	0.21	2.07	1.07	0.01
	Yes	0.67	0.89	1.75	1.48	0.34
20	No	0.32	0.55	1.14	0.61	0.01
	Yes	0.54	0.86	1.15	0.74	0.04
25	No	0.25	0.18	1.65	0.84	6.2×10^{-3}
	Yes	0.69	0.91	1.39	5.45	4.76
30	No	0.34	0.71	1.17	0.63	0.01
	Yes	0.55	0.91	1.14	0.77	0.08
35	No	0.30	0.41	1.52	0.79	0.01
	Yes	0.53	0.92	1.40	1.06	0.18
40	No	0.29	0.34	1.41	0.73	7.8×10^{-3}
	Yes	0.70	0.89	1.25	1.74	0.72
45	No	0.22	0.26	2.08	1.06	8.5×10^{-3}
	Yes	0.59	0.93	1.83	1.97	0.72
50	No	0.33	0.47	1.19	0.63	7.8×10^{-3}
	Yes	0.67	0.91	1.13	1.05	0.23

Source: Self-elaboration.

Table C.10 – Gated DSI-ESMDA ($\lambda = 4$).

	Gt.	Loc.	OC		MSE	Data mismatch	
			His.	Val.		Avg.	Std. dev.
	5	No	0.37	0.38	2.73	1.50	0.04
		Yes	0.69	0.76	2.30	18.01	18.20
	10	No	0.58	0.73	1.89	1.14	0.04
		Yes	0.68	0.90	1.68	1.58	0.31
	15	No	0.56	0.51	2.09	1.24	0.05
		Yes	0.76	0.92	1.78	1.71	0.32
	20	No	0.64	0.72	1.21	0.84	0.05
		Yes	0.72	0.86	1.26	1.04	0.06
	25	No	0.56	0.56	1.67	0.97	0.04
		Yes	0.76	0.97	1.40	5.67	4.74
	30	No	0.61	0.88	1.26	0.84	0.05
		Yes	0.72	0.91	1.26	1.06	0.09
	35	No	0.56	0.70	1.61	1.00	0.05
		Yes	0.68	0.92	1.46	1.35	0.18
	40	No	0.61	0.69	1.43	0.87	0.05
		Yes	0.80	0.93	1.29	1.94	0.70
	45	No	0.51	0.53	2.09	1.20	0.05
		Yes	0.72	0.95	1.86	2.17	0.69
	50	No	0.66	0.83	1.21	0.79	0.05
		Yes	0.79	0.92	1.15	1.28	0.22

Source: Self-elaboration.

Table C.11 – Gated DSI-ESMDA ($\lambda = 8$).

Gt.	Loc.	OC		MSE	Data mismatch	
		His.	Val.		Avg.	Std. dev.
5	No	0.72	0.78	2.75	2.18	0.21
	Yes	0.78	0.83	2.41	18.92	18.17
10	No	0.76	0.87	2.02	2.07	0.16
	Yes	0.81	0.90	1.78	2.74	0.36
15	No	0.81	0.70	2.16	2.13	0.23
	Yes	0.88	0.91	1.84	2.74	0.32
20	No	0.82	0.83	1.40	1.94	0.17
	Yes	0.82	0.86	1.46	2.30	0.22
25	No	0.80	0.85	1.71	1.77	0.24
	Yes	0.86	0.97	1.45	6.72	4.66
30	No	0.80	0.94	1.41	1.88	0.16
	Yes	0.85	0.92	1.36	2.27	0.23
35	No	0.78	0.88	1.72	2.03	0.21
	Yes	0.84	0.93	1.57	2.55	0.30
40	No	0.87	0.98	1.48	1.64	0.23
	Yes	0.89	0.95	1.37	2.90	0.62
45	No	0.78	0.86	2.16	1.98	0.23
	Yes	0.85	0.97	2.03	3.16	0.62
50	No	0.89	0.93	1.24	1.64	0.21
	Yes	0.90	0.95	1.22	2.31	0.26

Source: Self-elaboration.

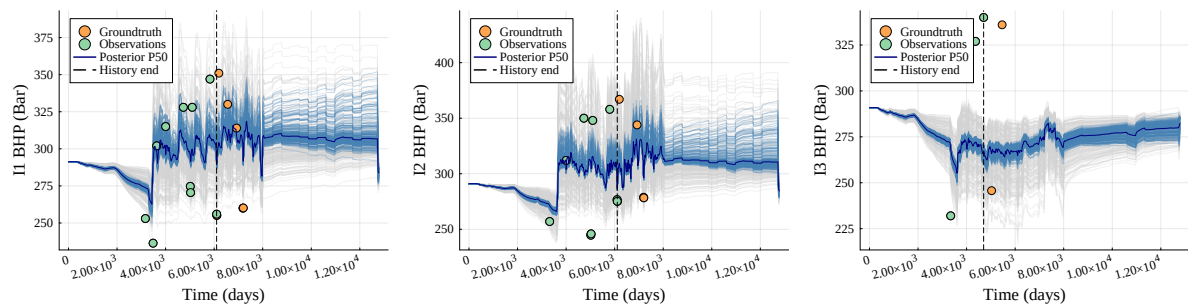
Table C.12 – Gated DSI-ESMDA ($\lambda = 16$).

		OC		MSE	Data mismatch	
		His.	Val.		Avg.	Std. dev.
		0.87	0.96	2.85	4.87	0.61
		0.85	0.88	2.52	22.13	18.29
		0.85	0.90	2.04	5.01	0.76
		0.87	0.92	1.83	5.94	0.86
		0.89	0.86	2.21	5.18	0.71
		0.91	0.91	1.85	6.06	0.76
		0.82	0.90	1.50	4.99	0.80
		0.83	0.84	1.47	5.45	0.88
		0.89	0.97	1.74	4.60	0.87
		0.89	1.00	1.52	10.02	4.49
		0.86	0.94	1.39	4.75	0.77
		0.88	0.93	1.36	5.27	0.93
		0.85	0.93	1.75	4.89	0.91
		0.88	0.93	1.66	5.53	1.00
		0.92	1.00	1.64	4.60	0.73
		0.91	0.94	1.56	6.25	0.84
		0.90	0.91	2.22	5.23	0.69
		0.91	0.97	2.17	6.88	0.91
		0.93	0.95	1.28	4.64	0.76
		0.94	0.95	1.38	5.62	0.84

Source: Self-elaboration.

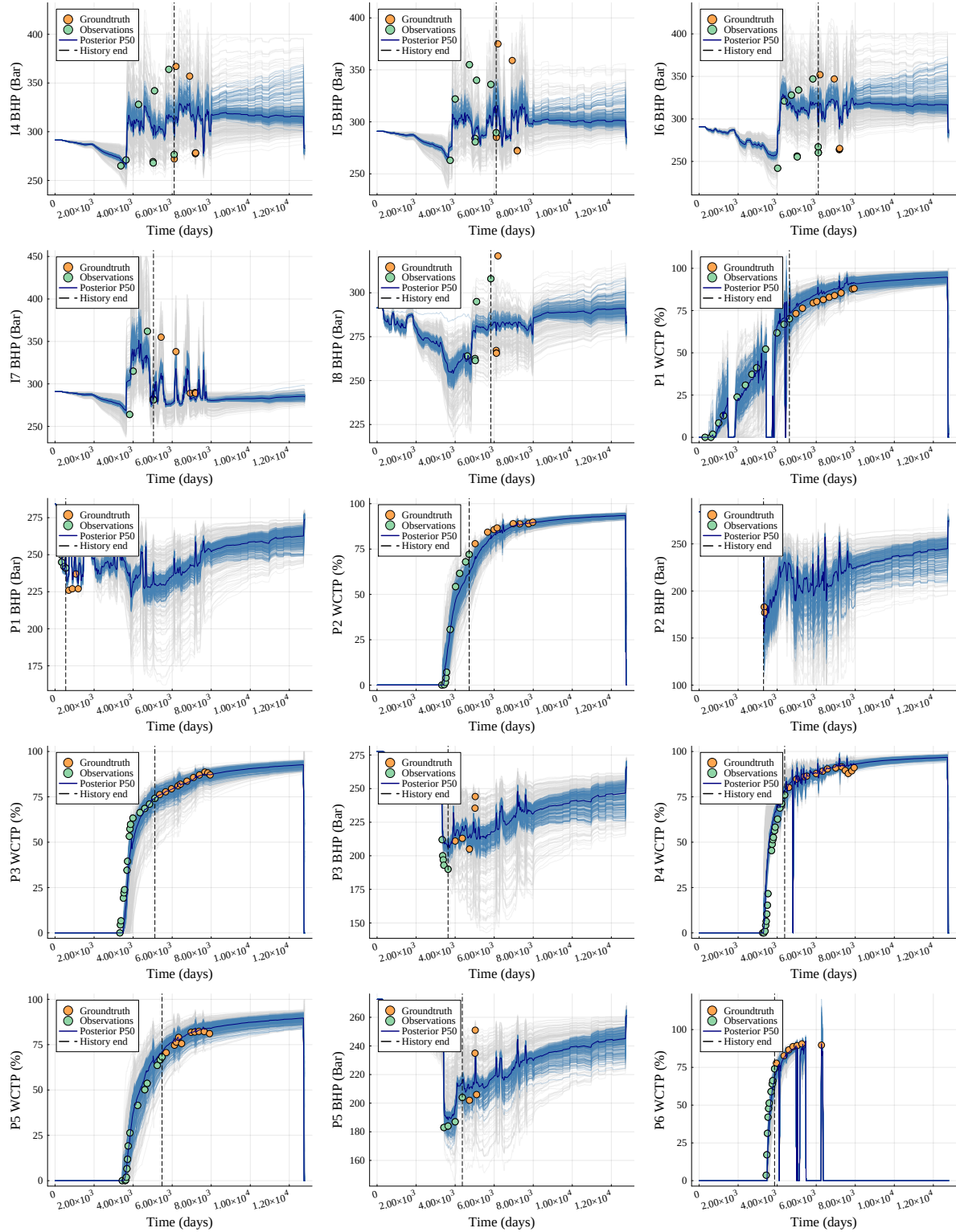
C.3 Real field

In the plots that follow, the gray curves denote the real field prior distribution; the blue curves denote the real field posterior distribution; the dark blue line denotes the median (50th percentile) of the posterior distribution; the dashed vertical line indicates the end of the history period; the teal circles denote real production data within the history period; the yellow circles denote real production data within the validation period.

Figure C.1 – Prior and posterior ensembles obtained with $\lambda = 4$ gating in the real field.

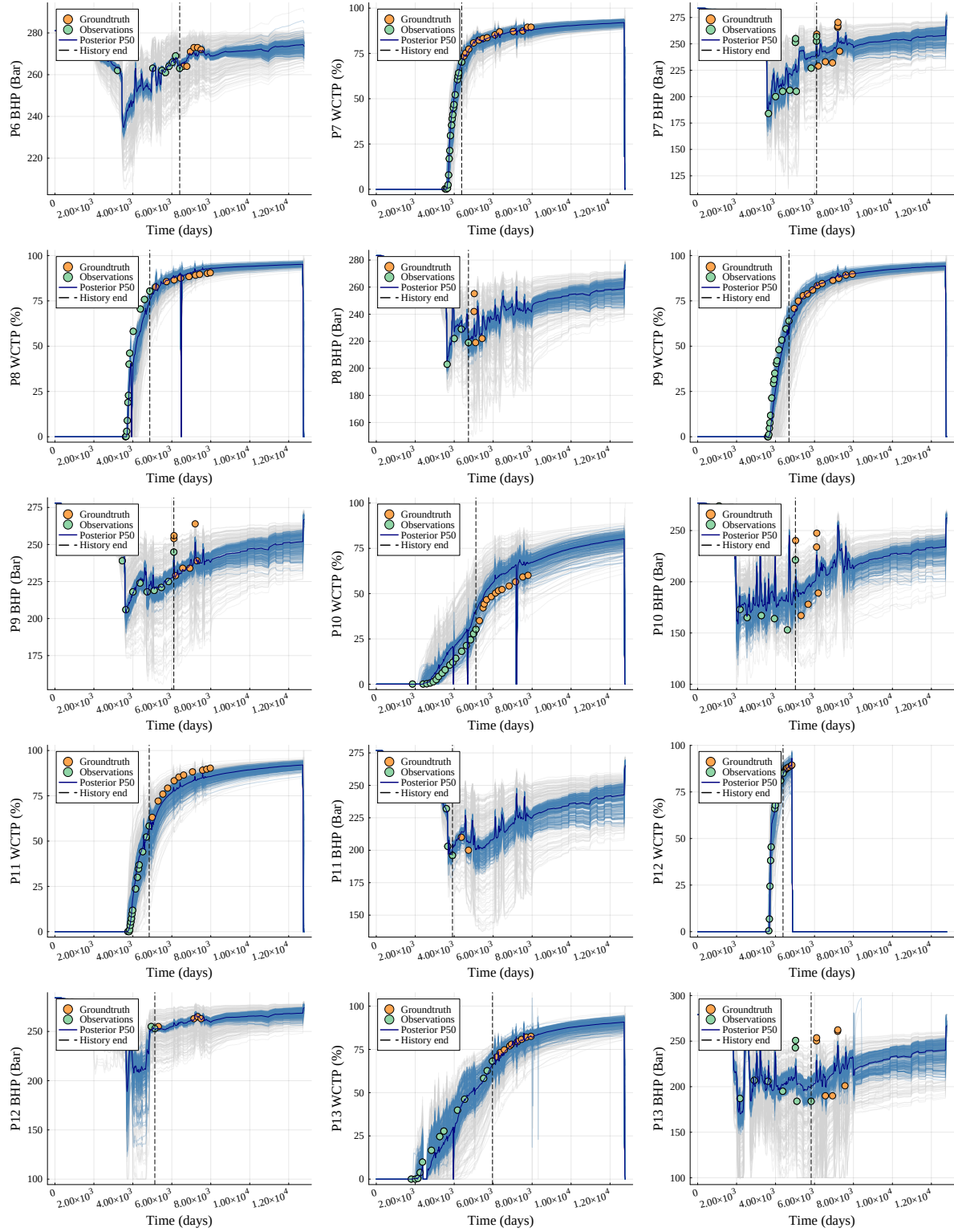
Source: Self-elaboration.

Figure C.2 – Continuation of figure C.1.



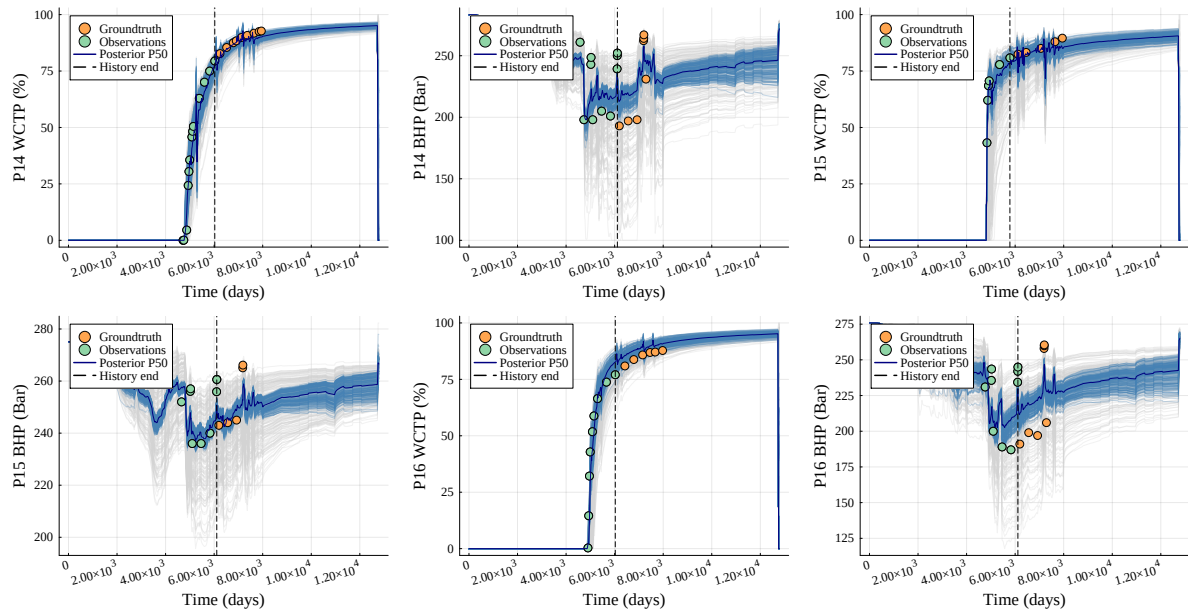
Source: Self-elaboration.

Figure C.3 – Continuation of figure C.1.



Source: Self-elaboration.

Figure C.4 – End of figure C.1.



Source: Self-elaboration.

