UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA
E INSTITUTO DE GEOCIÊNCIAS

GILSON MOURA SILVA NETO

SEISMIC DATA ASSIMILATION IN RESERVOIR MODELS: IMPROVING PRODUCTION FORECAST IN COMPLEX APPLICATIONS

ASSIMILAÇÃO DE DADOS SÍSMICOS EM MODELOS DE RESERVATÓRIOS: MELHORANDO A PREVISÃO DA PRODUÇÃO EM APLICAÇÕES COMPLEXAS

CAMPINAS
2021

GILSON MOURA SILVA NETO

SEISMIC DATA ASSIMILATION IN RESERVOIR MODELS: IMPROVING PRODUCTION FORECAST IN COMPLEX APPLICATIONS

ASSIMILAÇÃO DE DADOS SÍSMICOS EM MODELOS DE RESERVATÓRIOS: MELHORANDO A PREVISÃO DA PRODUÇÃO EM APLICAÇÕES COMPLEXAS

A thesis presented to the Mechanical Engineering Faculty and Geosciences Institute of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Petroleum Sciences and Engineering in the area of Reservoirs and Management.

Tese de doutorado apresentada à Faculdade de Engenharia Mecânica e Instituto de Geociências da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Ciências e Engenharia de Petróleo na área de Reservatórios e Gestão.

Orientador: Prof. Dr. Denis José Schiozer

Coorientadora: Dra. Alessandra Davólio Gomes

Este exemplar corresponde à versão final da Tese defendida pelo aluno Gilson Moura Silva Neto e orientada pelo Prof. Dr. Denis José Schiozer.

____

CAMPINAS

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

Si38s

Silva Neto, Gilson Moura, 1983-
Seismic data assimilation in reservoir models : improving production forecast in complex applications / Gilson Moura Silva Neto. – Campinas, SP : [s.n.], 2021.

Orientador: Denis José Schiozer.
Coorientador: Alessandra Davólio.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Mecânica.

1. Engenharia do petróleo. 2. Reservatórios (Simulação). 3. Filtragem de Kalman. 4. Prospecção sísmica. I. Schiozer, Denis José, 1963-. II. Davólio, Alessandra, 1980-. III. Universidade Estadual de Campinas. Faculdade de Engenharia Mecânica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Assimilação de dados sísmicos em modelos de reservatórios : melhorando a previsão da produção em aplicações complexas
Palavras-chave em inglês:
Petroleum engineering
Reservoirs (Simulation)
Kalman filtering
Seismic prospecting
Área de concentração: Reservatórios e Gestão
Titulação: Doutor em Ciências e Engenharia de Petróleo
Banca examinadora:
Denis José Schiozer [Orientador]
André Ricardo Fioravanti
Alexandre Anozé Emerick
Marcos Hexsel Grochau
Régis Kruel Romeu
Data de defesa: 02-06-2021
Programa de Pós-Graduação: Ciências e Engenharia de Petróleo

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-1218-3572
- Currículo Lattes do autor: http://lattes.cnpq.br/8409508075113234

UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA MECÂNICA

E INSTITUTO DE GEOCIÊNCIAS

Ph.D. THESIS

SEISMIC DATA ASSIMILATION IN RESERVOIR MODELS: IMPROVING PRODUCTION FORECAST IN COMPLEX APPLICATIONS

Author: Gilson Moura Silva Neto
Advisor: Prof. Dr. Denis José Schiozer
Co-advisor: Dra. Alessandra Davólio Gomes

The Examining Committee composed of the members below approved this thesis:
A Banca Examinadora composta pelos membros abaixo aprovou esta Tese:

Prof. Dr. Denis José Schiozer, Presidente
DE/FEM/UNICAMP

Prof. Dr. André Ricardo Fioravanti
DMC/FEM/UNICAMP

Dr. Alexandre Anozé Emerick
PETROBRAS

Dr. Marcos Hexsel Grochau
PETROBRAS

Dr. Régis Kruel Romeu
CONSULTOR

A Ata da Defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do aluno.

Campinas, 2 de junho de 2021

DEDICATION

To my family and friends.

ACKNOWLEDGMENTS

Firstly, I would like to express my gratitude to Prof. Denis José Schiozer, my advisor, for the precious guidance during this period. It is never easy to focus on one idea and keep open-minded, and you taught me how to balance those two vital attitudes to academic work.

I must express my sincere appreciation to Alessandra Davólio Gomes, my co-advisor and friend, for the constant support during all the work. You were always there to discuss the technical developments and also to help whenever I was insecure.

Geir Evensen, please accept my deepest thanks for the opportunity to expand my learning by working with you in Bergen. It was quite an experience and an honor to me. Your immense knowledge and willingness to help were invaluable to my development.

I would like to give special thanks to Victor de Souza Rios, Ricardo Vasconcellos Soares, Vinicius de Souza Rios, and Célio Maschio, my work partners and friends during this journey. Your contributions and personal aid were fundamental to overcome the difficulties. I extend my thanks to all my friends from CEPETRO, DEP, NORCE, and PETROBRAS. You really helped, not only technically, but also to adapt to the university, Campinas, and Bergen.

I sincerely appreciate the opportunity PETROBRAS provided me to do this course. Moreover, I am thankful to the Research Group in Reservoir Simulation and Management (UNISIM), Center of Petroleum Studies (CEPETRO), Department of Energy (DE-FEM-UNICAMP) for providing all kinds of support which enabled the realization of this work. I want to express special gratitude to the Norwegian Research Centre (NORCE) for providing everything I needed to work and feel welcome in Norway.

I sincerely appreciate the support from Energi Simulation and Libra Consortium. Furthermore, I am grateful to CMG and Schlumberger for software licenses.

I am very grateful to all my friends from Conquista, Salvador, Campinas, the Role-playing game group, the CS group, the potatoes literary society, and the “corrote” group. We are living in troubling times, and this was a challenging journey. Your presence, even if it were virtual, was vital to keep me happy and emotionally healthy. I express my special thanks to Amália, who taught us how to overcome some of life’s challenges and directly helped us when we were at a dead end. Thanks, Talita, for your continual support. You made our lives easier, even when we were in trouble.

Finally, I must express my gratitude to my family. I want you to know how much I value your constant support and encouragement. Soraya, my best friend and beloved partner, you comforted and helped me endure some of the most challenging days in my life. I know it was not easy, but your care and patience made us overcome everything. My son, Gilson “pequeno”, thank you for coming to be part of my life. I want you to know that, even in the hardest moments, to see that you are happy with us makes me happy too. I thank my father, Gilson, my mother, Elzane, and my sister, Clarissa, for the invaluable support. I miss living close to you, but you were always there for me. I extend my thanks to the Magalhães Family, specially Elisabete, my mother-in-law, Osir, my father-in-law, Verônica, Camila e André, for encouraging me whenever I needed it.

RESUMO

SILVA NETO, Gilson Moura, *Seismic Data Assimilation in Reservoir Models: Improving Production Forecast in Complex Applications*, Campinas, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, 2021. 259 p. Tese (Doutorado).

Apesar de ser considerada uma ferramenta essencial para projetos de produção de petróleo, a simulação numérica de reservatórios está sujeita a incertezas relacionadas à carência de informações sobre as propriedades do sistema em estudo e limitações típicas de modelos, que neste caso são críticas pelas atuais restrições de capacidade computacional. Os dados de vazão e pressão oriundos dos poços são usados para mitigar tais incertezas, mas eles são escassos espacialmente. Neste contexto, dados de sísmica com lapso de tempo tornam-se importantes no processo de calibração dos modelos, por proverem informações dinâmicas ricas espacialmente, sendo complementares aos dados de produção. Contudo, a assimilação de dados sísmicos para melhorar a capacidade preditiva de modelos de fluxo em reservatórios apresenta alguns desafios, dentre os quais podemos destacar dois. Primeiramente, a modelagem sísmica necessária para assimilação de dados comumente apresenta imperfeições capazes de afetar a comparação com os dados observados, levando a atualizações inconsistentes dos parâmetros. Além disso, os grandes conjuntos de dados provenientes das aquisições sísmicas elevam os requisitos computacionais, especialmente em casos com reservatórios de grandes dimensões.

Portanto, o objetivo deste trabalho é propor metodologias para assimilar dados sísmicos em cenários complexos, onde tais dificuldades estão presentes. O trabalho está estruturado em quatro estudos científicos. O primeiro estudo aborda uma metodologia para assimilar dados dinâmicos de poços e de sísmica com lapso de tempo na presença de erros de modelagem espacialmente correlacionados. O segundo estudo apresenta um método de assimilação baseado em conjuntos com análise local para grandes quantidades de dados sísmicos. No terceiro, é proposto um modelo de fluido para a simulação de dados sísmicos com variações composicionais, reduzindo os erros de modelagem em casos em que tais mudanças podem ocorrer no meio poroso. Estes dois últimos estudos viabilizam a aplicação final, no quarto estudo, em que é avaliada a assimilação de dados sísmicos num caso sintético com desafios semelhantes aos campos do pré-sal brasileiro.

Foi possível mitigar os efeitos dos erros de modelagem correlacionados espacialmente através de uma abordagem de condicionamento fraco dos modelos aos dados, através do acréscimo de parâmetros relacionados ao erro no processo de assimilação. Além disso, o suavizador por conjunto iterativo proposto com análise local reduziu as limitações decorrentes da assimilação de grandes conjuntos de dados oriundos da sísmica. O modelo composicional de fluidos para a modelagem petroelástica também contribuiu para a redução de erros de modelagem na assimilação de dados sísmicos, ao melhorar a descrição dos processos físicos relacionados a produção de fluidos voláteis com injeção miscível de gás no meio poroso. Finalmente, o método eficiente de assimilação de dados proposto, juntamente com o modelo de fluidos, se mostrou uma solução viável para um processo de assimilação de dados sísmicos em reservatórios que apresentam desafios semelhantes aos de campos de petróleo do pré-sal brasileiro.

Palavras-Chave: simulação de reservatórios; calibração de modelos; assimilação de dados baseada em conjuntos; métodos baseados em filtros de Kalman; suavizador por conjunto; análise local; erros de modelo; sísmica 4D; ajuste de histórico com sísmica; modelo de fluidos; modelo petroelástico.

ABSTRACT

Reservoir flow simulation is an essential tool for upstream projects. Nevertheless, uncertainties related to the lack of information regarding the reservoir properties and model limitations, critical due to the current computational capabilities, influence its results. The assimilation of fluid rates and pressure data from the wells can mitigate these uncertainties, but they are scarce in space. Therefore, time-lapse seismic data become essential because it can provide dynamical information distributed in space, complementing the production data. However, time-lapse seismic data assimilation has some significant challenges, among which we can highlight two. First, the forward models are imperfect, and their errors can lead to unphysical parameter updates. Furthermore, the large data sets provided by seismic increases the computational costs, particularly in applications related to big reservoirs.

From this perspective, this work aims to propose methodologies to assimilate seismic data in cases where these difficulties are relevant. The work comprises four scientific studies. The first study proposes a methodology to assimilate time-lapse seismic data handling significant spatially-correlated errors related to the forward model and the observed data. The second study presents an ensemble-based data assimilation method with local analysis capable of dealing with big seismic data sets. In the third study, we propose a fluid model to simulate seismic data, reducing model errors when there are significant compositional changes in the porous media. These last two studies allowed for the last application, where we evaluate the seismic data assimilation in a synthetic case with similar challenges to a Brazilian pre-salt case.

The weak-constraint methodology enabled the mitigation of the spatially correlated model error effects by including error-related parameters in the data assimilation workflow. Moreover, the proposed iterative ensemble smoother with local analysis reduced the limitations related to the assimilation of time-lapse seismic big data sets. The compositional fluid model for petroelastic models contributed to the model error reduction by providing a physically consistent representation of the volatile fluid production with miscible gas injection. Finally, the efficient iterative ensemble smoother and the fluid modeling methodology were viable solutions to assimilate time-lapse seismic data from reservoirs with similar challenges to a Brazilian pre-salt field.

Keywords: reservoir simulation; model calibration; ensemble data assimilation; Kalman-filter-based methods; iterative ensemble smoothers; local analysis; model error; time-lapse seismic; 4D seismic; seismic history matching; fluid model; petroelastic model.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Reservoir simulation is an essential tool for decision-making in upstream projects. This type of simulation commonly aims to estimate future oil, gas, and water production. Nevertheless, uncertainties related to the lack of information used to build the flow models always influence these estimates. A straightforward approach to mitigate the influence of the uncertainties and improve the production forecasts is to assimilate data to improve the characterization of reservoir model parameters. The current goal is not to find a single model as the best match to the data. It is vital to represent the remaining uncertainties, considering all the information available, especially when these models are used in mid to long term decisions.

In reservoir engineering, it is common to refer to the data assimilation into reservoir flow models as history matching. The standard data used in this process is the well production data, including pressure, oil, gas, and water rates. Nevertheless, this information source is generally insufficient to mitigate the effects of the uncertainties in the forecasts. One of the limitations of this type of information is that it is scarce in space, although it is rich in time. Note that reservoir history matching is an ill-posed inverse problem.

Reservoir geoscientists and engineers have been applying seismic data as an essential source of information to build reservoir models. 3D seismic data commonly plays a crucial role in standard geological modeling workflows. Furthermore, it is possible to incorporate this type of data into the simulation models using data assimilation processes.

Another relevant source of information is time-lapse seismic (TLS) data, also called 4D seismic data. The initial applications of multiple time-lapsed 3D seismic acquisitions to obtain dynamic information regarding the porous media flow were qualitative. Nevertheless, recent studies apply TLS data quantitatively to calibrate model parameters. This data assimilation process is drawing increasing attention because it can provide dynamic information rich in space. Therefore, one can consider it complementary to the usual well data.

The ensemble-based methods have become a popular option to incorporate both well and time-lapse seismic data into reservoir models. Two characteristics of this type of method make for this choice. Firstly, these methods can handle complex nonlinear models with a vast number of parameters that reach magnitudes from 10^5 to 10^6 in reservoir applications. Moreover, they provide an ensemble of conditioned models representing a straightforward way to perform uncertainty quantification after the data assimilation. These are Bayesian methods, which aim to sample the parameters' posterior distribution, considering the prior information and the data to be assimilated. One represents the former by the prior ensemble of models.

The problem of assimilating time-lapse seismic data using ensemble-based methods has particular challenges. First, TLS data inclusion in the assimilation process significantly increases the data sets, leading to higher computational requirements. Secondly, correlated errors can significantly affect the comparison of the observed TLS data with the simulated response. These errors are hard to define in practical applications, and some of their causes are missing or overly simplified physics, resolution issues, neglected uncertainties, among others. Note that conventional ensemble-based data assimilation workflows neglect model errors, which may cause exaggerated uncertainty reduction and unphysical changes of the parameters.

Some of these difficulties are more severe in situations like Brazilian pre-salt reservoirs, where one deals with complex geological models with a large number of parameters, big data sets, and volatile fluid containing significant amounts of carbon dioxide. The reservoir sizes influence the dimension of the data assimilation problem in terms of the number of parameters. Besides, the application of permanent seismic monitoring and the models' dimensions tend to increase the computational requirements due to the number of seismic data points to assimilate. Finally, the physics complexity, including seismic wave propagation and volatile fluid flow with miscible gas injection, increases the chance of significant model errors.

This thesis focuses on TLS data assimilation using ensemble-based methods dealing with the two mentioned challenges, correlated errors related to the forward model, and assimilation of big data sets into big models. It comprises seven chapters, the introduction, four chapters with scientific studies, the conclusion, and future researches. Moreover, it includes two appendixes, one with an additional study and one with the published manuscripts' license agreements.

The first scientific study presents a methodology to assimilate production and TLS data in a complete history-matching problem, considering the influence of spatially correlated model errors. In this study's applications, these errors arise from pressure sensitivity mismatch and unmodeled seismic resolution losses. The second study presents an iterative ensemble smoother (IES) with local analysis to assimilate TLS data, which can handle big data sets and big models. The third study improves the fluid representation in the seismic forward model, using the same equation of state (EOS) as the flow simulation. In this way, it is possible to reduce model errors when the fluids' compositions change in the porous media. The last work is an application of production and seismic data assimilation using a data set and model with similar features of a Brazilian pre-salt reservoir. In this case study, the new fluid model and the new data assimilation method play a crucial role in reducing a complex model's uncertainties using well and TLS data.

The studies mentioned above relates to the general TLS data assimilation workflow depicted in Figure 1.1. The first step in this workflow is the reservoir flow simulation, using an ensemble of reservoir models. This simulation provides the TLS fluid model inputs in step two, which computes the fluid properties, such as density and bulk modulus. The petroelastic model in the third step provides the simulated TLS data. Compared to the observed data, the simulated production and TLS data are the IES inputs, enabling the update of the model and error parameters. The workflow considered error modeling for the TLS seismic, although one could also consider it for the well data simulation or neglect model error in all simulations if they are not relevant.

This thesis' first study relates to the third step of Figure 1.1, which includes error-related parameters in the data assimilation workflow to handle significant spatially-correlated errors. The second work refers to the fourth step, as it proposes an iterative ensemble smoother with local analysis for time-lapse seismic data assimilation. The third study connects to the second step, proposing a compositional fluid model to pertain to the petroelastic model. The fourth work uses this general workflow to assimilate TLS data in a case with Brazilian pre-salt characteristics, using the last two methods. In this last application, it was not necessary to include error-related parameters since quasi-ideal TLS data was available.

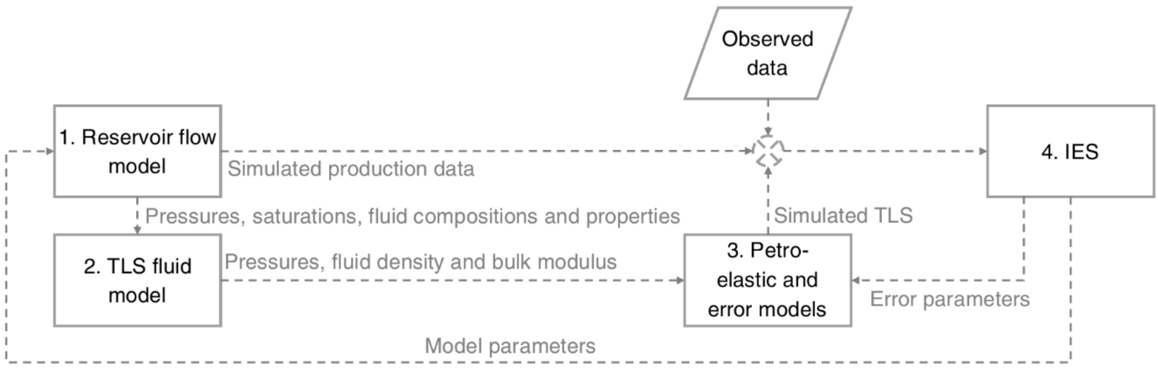


Figure 1.1: General time-lapse seismic data assimilation workflow.

This thesis also includes a study regarding 3D data assimilation considering forward-model-related errors as Appendix A. Although not directly pertaining to the workflow in Figure 1.1, this study could be necessary for some applications as an initial model improvement to assure that they represent the 3D seismic data before they compute time-lapse seismic simulations.

1.1 Motivation

The motivation for this work comes from some established premises. First, reservoir flow models are vital for the decision process in upstream projects, but uncertainties always influence the forecast quality. Secondly, there is a lack of data to mitigate the impact of these uncertainties through data assimilation, which makes for the importance of using time-lapse seismic data quantitatively to calibrate the reservoir parameters. This type of data is rich in space and is complementary to the information from the wells.

Nevertheless, it is possible to identify two aspects of TLS data assimilation that complicate the applications or jeopardize the results. The first aspect is that the increase of the computational requirements due to big data sets may limit some applications using straightforward ensemble-based methods. The second is that correlated model-related errors in the comparison of simulated and observed data, if neglected, may impair the data assimilation results. Therefore, it is crucial to propose methodologies that handle these complex situations, present in many fields worldwide, especially in Brazilian pre-salt.

1.2 Objective

The main objective of this work is to propose methodologies to assimilate TLS data in reservoir models in realistic complex cases, considering forward model imperfections and the assimilation of big data sets. The final goal is to apply the methodologies in a case that mimics some complexities of a pre-salt reservoir, including compositional fluid changes by the miscible gas injection. As with any general data assimilation workflow, the ultimate objective is to develop methodologies to improve oil production forecasts using TLS data, even in complex applications.

1.3 Work description

This work comprises four scientific studies and one additional study as an appendix. This subsection summarizes each of them, highlighting their main contributions and their relations with the thesis goals. We present the studies' complete texts in the following chapters of this document and Appendix A. Appendix B contains the license agreements of the published material.

1.3.1 Assimilating time-lapse seismic data in the presence of significant spatially correlated model errors (published work)

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, p. 109127, Jun. 2021. https://doi.org/10.1016/j.petrol.2021.109127.

This study aims to improve the production forecast using time-lapse seismic data containing distortions related to correlated errors when comparing the observed and simulated data. In practical applications, model imperfections may cause correlated errors, which are neglected in straightforward data assimilation workflows. For instance, unmodeled processes such as resolution loss, scale differences, seismic wave propagation, and seismic inversion algorithms may significantly affect the data. Furthermore, the petroelastic model can also have relevant imperfections.

We apply a weak-constraint formulation to the TLS data assimilation process to mitigate the impact of these errors. In this type of formulation, we include error-related variables as parameters in the calibration process. In this study, we also include well production data to update the reservoir parameters, and we test the methodology using the UNISIM-I-H benchmark.

This research focuses on the third step of Figure 1.1, as it mitigates the influence of TLS model-related errors by including error-related parameters in the data assimilation workflow. Furthermore, it relates to the main objective to use TLS data to improve reservoir characterization and production forecast, even when we have imperfect forward models.

This work improved the reservoir characterization and production forecast using relatively low-resolution TLS data and models with pressure sensitivity mismatch, mimicking practical challenges. In a synthetic 2D application, the proposed methodology reduced the bias and porosity mismatch, using a petroelastic model with a pressure sensitivity deviation, a common problem in real applications. Furthermore, in a synthetic realistic field application, the ensemble updated using the same methodology could provide better forecasts than when one assimilates the TLS data neglecting the model errors. Furthermore, the data assimilation methodology with TLS data improved the forecasts compared to the test where only well data was available.

1.3.2 Subspace Ensemble Randomized Maximum Likelihood with local analysis for time-lapse seismic data assimilation (published work)

SILVA NETO, G. M.; SOARES, R. V.; EVENSEN, G.; DAVOLIO, A.; SCHIOZER, D. J. Subspace Ensemble Randomized Maximum Likelihood with Local Analysis for Time-Lapse-Seismic-Data Assimilation. SPE Journal, p. 1–21, 1 Feb. 2021. https://doi.org/10.2118/205029-PA.

In the initial applications of TLS data assimilation and the 3D case study in Appendix A, we dealt with some difficulties related to the massive amounts of data that the seismic includes in the workflows. Furthermore, in this thesis's final application, we analyze the TLS data assimilation in a synthetic case representing similar challenges to a Brazilian pre-salt reservoir. Therefore, there are higher computational costs due to the models' size and the data sets, both in the forward model simulation and the data assimilation algorithm. This context motivated this second study to develop an ensemble-based method that handles big models and big data sets in TLS assimilation. Therefore, it relates to the fourth step of Figure 1.1 and the goal to use TLS big data sets to improve the reservoir models.

The study proposes a local-analysis scheme using an efficient implementation of the Subspace Ensemble Randomized Maximum Likelihood (SEnRML) method. In this implementation, the computations scale linearly with the dimension of the data set. Furthermore, our results with the local analysis scheme show that it could provide similar results to the Ensemble Smoother with Multiple Data Assimilations, a popular choice for reservoir history matching. We tested both distance-based and correlation-based local analysis. The latter has the advantages of applying to nonlocal data and parameters and not requiring tuning localization lengths. We used a simple 2D model and the simulation models of the UNISIM-I-H benchmark in our tests.

1.3.3 Improving fluid modeling representation for seismic data assimilation in compositional reservoir simulation (published work)

SILVA NETO, G. M.; RIOS, V. de S.; DAVOLIO, A.; SCHIOZER, D. J. Improving fluid modeling representation for seismic data assimilation in compositional reservoir simulation. Journal of Petroleum Science and Engineering, vol. 194, p. 107446, Nov. 2020. https://doi.org/10.1016/j.petrol.2020.107446.

In a time-lapse seismic data assimilation workflow, it is necessary to provide fluid models with two distinct objectives. The first one is to describe the pressure, volume, and temperature behavior for the fluid flow in the porous media, the first step of Figure 1.1. The second one is to describe the density and bulk modulus for the petroelastic simulation in the workflow's second step. The first objective is well established for typical applications, including situations where the phases' compositions change in the reservoir (PEDERSEN; CHRISTENSEN; SHAIKH, 2015). Moreover, there are well-known correlations suitable for petroelastic models for black-oils (BATZLE; WANG, 1992). Nevertheless, this thesis considers the time-lapse seismic data assimilation in a reservoir model with volatile fluid with varying composition. Therefore, this work proposes a methodology to model the fluid to represent the flow in the porous media and the acoustic impedance variations, which we use as seismic data. We propose to use a cubic equation of state calibrated using pressure, volume, and temperature (PVT) data to characterize the fluid for both the flow and the petroelastic models. Our model considers volatile hydrocarbons with a significant amount of CO_2, typical characteristics in Brazilian pre-salt reservoirs. Moreover, it enables a physically consistent simulation of the seismic signal of miscible gas injection and the water-alternating-gas (WAG) process.

From the TLS data assimilation perspective, this study proposes a new way of representing the fluid elastic properties variations, referring to the second step of Figure 1.1. In this regard, it tends to reduce model errors related to the classical fluid correlations commonly applied in petroelastic models. Our results indicate that an overly simplified fluid model could influence the data assimilation results by changing the reservoir's simulated acoustic impedance variations. Furthermore, we show that it is possible to improve the fluid model, for both flow and seismic simulation, by incorporating the experimental speed of sound data without impairing the PVT match. Up to the present, this is not a common practice in the industry.

1.3.4 Assimilating well and time-lapse seismic data in a challenging pre-salt-like case using an iterative ensemble smoother for big data sets (to be submitted)

SILVA NETO, G.M., RIOS, V.S., MASCHIO, C., DAVOLIO, A. and SCHIOZER, D.J. 2021. Assimilating well and time-lapse seismic data in a challenging pre-salt-like case using an iterative ensemble smoother for big data sets.

This synthetic case study aims to assimilate time-lapse seismic data considering some of the challenges of Brazilian pre-salt reservoirs. We consider the effects of volatile fluids with significant amounts of CO_2 and miscible gas injection. We also consider the simultaneous assimilation of multiple seismic monitors, which add to the data sets' size. Therefore, the fluid modeling methodology developed in the third study and the data assimilation method proposed in study four play a crucial role in this application.

This work integrates previous methods developed in this thesis in a realistic problem, testing their interaction in the workflow presented in Figure 1.1. It is important to note that our data set did not evidence significant model-related correlated errors that would motivate the application of a weak-constraint formulation (section 1.3.1). Moreover, the application does not include a giant reservoir model, due to time and infrastructure restrictions to run the flow simulations. We used sector 1 of the benchmark case called UNISIM-III, built with data and information from pre-salt reservoirs.

Our results show that the data assimilation workflow, including the fluid representation using an EOS and the SEnRML with local analysis, is a viable solution to assimilate well and TLS data in a challenging pre-salt-like case. Moreover, the TLS data provided important information to improve the production forecast in this application. Furthermore, this work adds to the discussion of the application of ensemble-based methods in highly nonlinear problems due to heterogeneous reservoirs with complex physics.

1.3.5 3D seismic data assimilation to reduce uncertainties in reservoir simulation considering model errors (published work) – Appendix A

SILVA NETO, G. M.; DAVOLIO, A.; SCHIOZER, D. J. 3D seismic data assimilation to reduce uncertainties in reservoir simulation considering model errors. Journal of Petroleum Science and Engineering, vol. 189, p. 106967, 1 Jun. 2020. https://doi.org/10.1016/j.petrol.2020.106967.

This work proposes a methodology to assimilate 3D seismic data considering model errors due to missing physics. Although other phenomena may also affect the observed data, in this study, the unmodeled loss of resolution related to the seismic wave propagation causes the distortion that influences the data assimilation results. We focus on two approaches to handle the limitations of our forward model. First, we update the covariance matrix of total observation errors following an iterative approach. Furthermore, we introduce an analytical function that acts as a proxy to the systematic errors in comparing the observed and the simulated data. By applying the proposed methodology, we improved the volume characterization and the production forecast using 3D seismic data with a relatively low resolution. We used the benchmark UNISIM-I-H in our tests.

This study focuses on 3D seismic data, commonly an information source during the geological modeling process. Thereby, it is not directly related to the workflow in Figure 1.1, and we include it as an appendix in this thesis for two reasons. Firstly, it is important to condition the model to the 3D data before assimilating TLS data when the former is not part of the geological modeling workflow since 3D and time-lapse seismic data play different roles in the parameter calibration. While the former indicates general geological features and is closely related to the 3D volume, the latter represents the porous media's regional dynamical features. Secondly, a similar methodology proposed to 3D seismic data assimilation can help TLS data assimilation in some applications.

2 ASSIMILATING TIME-LAPSE SEISMIC DATA IN THE PRESENCE OF SIGNIFICANT SPATIALLY CORRELATED MODEL ERRORS

Authors:

Gilson M. Silva Neto

Alessandra Davolio

Denis J. Schiozer

2.1 Abstract

Time-lapse seismic data is becoming a common information source in reservoir model calibration workflows to improve production forecasts. The standard process compares a forward model's results with the observed data to update the model's parameters from the existing deviations.

Ensemble-based methods are popular choices for this process. However, the so-called forward model is always a simplification of the real phenomena. These simplifications may significatively influence the relation between simulated and observed data and possibly yield inconsistent updates of the parameters and uncertainty underestimation. In the conventional approach for this problem, the so-called strong-constraint formulation neglects the model's limitations, causing unphysical updates of the parameters to reduce the distance between simulated and observed data.

In this work, we propose a methodology to apply a weak-constraint formulation to the time-lapse seismic data assimilation to mitigate the above problem. We consider the forward model error with an additive term and update it during the data assimilation workflow. By adopting this approach, we reduce the impact of the model errors in the calibrated parameters. Also, the proposed methodology handles model bias as a type of general model error. The inclusion of the additive term weakens the updates of the model due to the time-lapse seismic data.

We show that this procedure significantly benefits the data assimilation results when there are substantial spatially correlated model errors, and it has a minor impact when

applied to a low-error case. We apply the proposed method to assimilate time-lapse seismic data using the Ensemble Smoother with Multiple Data Assimilations in a simple 2D case and in a realistic benchmark synthetic case based on a real offshore reservoir. In the former case, we consider model error related to the pressure sensitivity in the petroelastic model. In the latter, we consider a realistic synthetic time-lapse seismic and the correlated errors result from seismic modeling and inversion. Moreover, in the latter, we also assimilate well data.

The results indicate that our methodology improved the reservoir characterization and the production forecast using relatively low-resolution time-lapse seismic data.

Abbreviations:

ESMDA – Ensemble Smoother with Multiple Data Assimilations

NQDS – Normalized Quadratic Deviation with Sign

PEM – Petroelastic Model

TLS – Time-Lapse Seismic

VOIP – Volume of Oil in Place

Keywords:

Model calibration; ensemble data assimilation; iterative ensemble smoothers; history matching; model errors; ESMDA

2.2 Introduction

The application of time-lapse seismic (TLS) data to calibrate reservoir models using ensemble-based methods is becoming more frequent in petroleum literature. Some publications already report real field applications (EMERICK, Alexandre A., 2016; EMERICK, Alexandre A.; REYNOLDS, 2013b; FAHIMUDDIN; AANONSEN; SKJERVHEIM, 2010a; SKJERVHEIM, Jan-Arild *et al.*, 2007). More recently, Fossum and Lorentzen (2019) demonstrated the capability of iterative ensemble smoothers to assimilate large amounts of seismic and production data using different noise handling and localization approaches. Furthermore, Wojnar et al. (2020) proposed an ensemble-based time-lapse seismic data assimilation workflow and applied it to a mature undersaturated oil field.

In typical TLS data assimilation applications, the forward model includes a reservoir flow model and a petroelastic model (PEM), as in some examples from the literature (EMERICK, Alexandre A., 2016; SKJERVHEIM, Jan-arild; EVENSEN, 2011; TAHA *et al.*, 2019). Both models have parameters that one cannot obtain exactly. Furthermore, both models simplify reality, and unmodelled or overly simplified processes may influence the data. Some examples of these processes are scale differences, seismic wave propagation effects, and inversion algorithms (SILVA NETO, Gilson M.; DAVOLIO; SCHIOZER, 2020). Another source of model errors is an inconsistent prior ensemble, lacking critical flow-related elements, such as fractures (MA; JAFARPOUR; QIN, 2019). Therefore, model errors of different sources may significantly affect the TLS data assimilation. It is essential to mention that some applications compare seismic data in the amplitude domain by including simplified seismic modeling in the forward model (FAHIMUDDIN; AANONSEN; SKJERVHEIM, 2010b; LEEUWENBURGH; BROUWER; TRANI, 2011), another approximated model of the actual process. Furthermore, some studies represent the TLS signal using binary images (CHASSAGNE *et al.*, 2016; DAVOLIO; SCHIOZER, 2018; OBIDEGWU; CHASSAGNE; MACBETH, 2017), whose integration with iterative ensemble smoothers is not straightforward.

The history-matching workflows, using ensemble-based methods, usually neglect the model-related error (DONG; GU; OLIVER, 2006; EMERICK, Alexandre A., 2016; FAHIMUDDIN; AANONSEN; SKJERVHEIM, 2010b; LORENTZEN *et al.*, 2019; SKJERVHEIM, Jan-Arild *et al.*, 2007; YIN; FENG; MACBETH, 2019). This approach may lead to unreliable results from TLS data assimilation because the ultimate goal is to use the data to improve the production forecast. Differences between the observed and simulated data may originate from the model imperfections, and they may cause unphysical model updates (EVENSEN, Geir, 2019; OLIVER, D.S.; ALFONZO, 2018a). Moreover, the parameters' unphysical changes may significantly impact the well rate forecasts and the decision process based on the reservoir models.

Accounting for model errors in ensemble-based data assimilation workflows is a relatively recent subject in petroleum-related literature. Oliver and Alfonzo (2018a) proposed a methodology to consider the model error in data assimilation in an iterative approach. One of the critical steps in their methodology was to update the covariance matrix of the total observation error using the last data assimilation residuals. Alfonzo and Oliver (2020) applied this iterative methodology to time-lapse seismic data assimilation in a field-scale problem. Although their methodology reduced the impact of biased observations, they did not include bias treatment directly in the data assimilation. Silva Neto et al. (2020) applied a similar iterative approach to 3D seismic data assimilation using another realistic synthetic case. In addition to updating the error covariance matrix, they changed the forward model to reduce bias. Lu and Chen (2020) applied a related approach to the assimilation of well data from the Norne field and discussed how to consider model error in the production forecast. They assumed that the model error has zero mean but mentioned the future investigation of cases where this assumption does not hold.

Another known technique to reduce spurious updates of parameters in the presence of significant model errors is to inflate the covariance matrix of measurement errors (LIMA; EMERICK; ORTIZ, 2020; SUN; VINK; GAO, 2017; VINK; GAO; CHEN, 2015). When the distortions related to model imperfections or data issues occur at specific regions, it is possible to use confidence maps to discard part of the data in the assimilation workflow (DOS SANTOS *et al.*, 2018; EMERICK, Alexandre A.; REYNOLDS, 2013b). Nevertheless, model errors tend to have distinctive characteristics, with a high degree of spatial or temporal correlations, and these approaches may be insufficient to mitigate its effects (DOHERTY; WELTER, 2010). In more recent work, Akter et al. (2021) addressed parameter estimation in the presence of model uncertainty, using ensemble Kalman filter and perturbing the model inputs and outputs with Gaussian noise whose amplitude was a tuning parameter.

Rammay et al. (2019) handled model error in a history matching workflow by parametrizing an error model using principal component analysis and including these parameters as part of the data assimilation problem. Their results show that the data assimilation workflow considering model errors reduced the parameter bias and improved the model's reliability. Nevertheless, to define the prior statistics of the model error, their methodology requires high-fidelity models, which have negligible model errors. Other authors relied on a high-fidelity model to improve the calibration of the parameters in the presence of model errors caused by coarse grid forward models (STEPHEN, K. D., 2007) and simplified streamline reservoir simulation (STEPHEN, Karl D.; SHAMS; MACBETH, 2009). The necessity of models with minor model errors limits the applicability to situations where the sources of model errors are known, and one can alleviate them.

Recently, Rammay et al. (2021) proposed a modified form of the Ensemble Smoother with Multiple Data Assimilations (ESMDA) (EMERICK, Alexandre A.; REYNOLDS, 2013a) algorithm to handle model errors. They split the differences between the observed and simulated data into two parts. One was related to the parameter updates, and the other originated from model errors. They proposed an approximated split parameter, updated during the data assimilation, to set the two portions, and assumed that this split parameter was directly proportional to the residuals' mean. The results indicate that this approach reduced parameter bias and uncertainty underestimation due to imperfect models. However, they only tested the method using toy problems and a simple 2D reservoir model without TLS data.

Luo et al. (2021) proposed the consideration of model errors in ensemble-based seismic data assimilation, treating them as a data-driven functional approximation problem (LUO, 2019). This problem was solved using a machine learning method integrated with an iterative ensemble smoother. The main advantages of their approach are that it considers the relations between model errors and other variables and does not need to assume the model error statistics. Although their results seem promising, additional investigation defining the residual model in different cases is still required. Their approach is related to the current work because both cases calibrate the model error term during the data assimilation. Nevertheless, we do not address the machine learning method integration to the TLS data assimilation in this study.

Evensen (2019) presented the theoretical foundation to consider model errors using iterative ensemble smoothers. He adopted a weak-constraint formulation, where one introduces a model error parameter and calibrates it during the data assimilation. He mentioned that one of the most challenging steps of the methodology is to define the prior statistics of the error parameter. Nevertheless, this task is similar to defining any uncertain parameter in the model, and this difficulty should not hinder accounting for this type of error. Later, Evensen (2020) applied this idea to production data history matching. We follow his formulation and propose a practical methodology to apply it to TLS data assimilation in this work.

In this work, we propose a methodology to alleviate model error impacts on TLS data assimilation results. We present a practical procedure to apply a weak-constraint formulation on ESMDA to assimilate TLS data. This formulation differs from the classical strong-constraint problem because, in the weak-constraint formulation, one assumes that the model is imperfect by augmenting the parameter matrices to include error variables in the problem, attenuating the changes in the original parameters. This approach mitigates unphysical parameter updates due to inconsistencies in the relation between the simulated and observed data caused by model errors. The current methodology also handles model bias, which we consider part of the model error. We apply our methodology to two synthetic cases: a simple 2D example and a realistic production and TLS data assimilation case. The latter comprises the benchmark UNISIM-I-H (MASCHIO, C. *et al.*, 2013), which resembles a real offshore reservoir in the Campos Basin, Brazil.

The specific objectives of this study are:

- Show the possible impacts of correlated model-related errors in TLS data assimilation.
- Propose a practical methodology to mitigate the effects of this type of error.
- Show that our methodology does not impair the data assimilation in a low-error situation.
- Show that it is possible to improve the production forecasts using TLS data even with imperfect forward models.

2.3 Ensemble Smoother with Multiple Data Assimilations

We apply the ESMDA method (EMERICK, Alexandre A.; REYNOLDS, 2013a) to update the grid and scalar reservoir parameters using production and time-lapse seismic data. Like other iterative ensemble smoothers, the choice of ESMDA enables us to update a large number of parameters, considering both the prior geological information and the acquired production and TLS data simultaneously. Furthermore, the data assimilation workflow provides an ensemble of calibrated models, which helps a posterior uncertainty analysis. Since we include seismic as an information source, we used the ESMDA algorithm implementation for a large number of measurements from Emerick (2016). In this section, we briefly explain the method, highlighting the main features related to our methodology.

We assume that there is a forward model $\boldsymbol{g}$ used to simulate the data $\boldsymbol{d}^{sim} \in \Re^{m\times 1}$ with a vector of parameters $\boldsymbol{x} \in \Re^{n\times 1}$ containing reservoir properties. The model parameters are uncertain, and one can represent the prior uncertainties through an ensemble of N realizations of the parameter vector, obtaining the matrix $\boldsymbol{X}^f = \left(\boldsymbol{x}_1^f, \boldsymbol{x}_2^f, \boldsymbol{x}_3^f, \dots, \boldsymbol{x}_N^f\right) \in \Re^{n\times N}$. If we simulate $\boldsymbol{g}$ using each column of $\boldsymbol{X}^f$, we obtain the simulated data matrix $\left(\boldsymbol{D}^{sim}\right)^f \in \Re^{m\times N}$.

Assuming that the model is imperfect, the observed data relates to the model results as

$$\boldsymbol{d}^{\boldsymbol{obs}} - \boldsymbol{\varepsilon}_d = \boldsymbol{g}(\boldsymbol{x}^{\boldsymbol{true}}) - \boldsymbol{\varepsilon}_g, \tag{2.1}$$

where $\boldsymbol{d}^{obs} \in \Re^{m\times 1}$ is the observed data, $\boldsymbol{\varepsilon_d} \in \Re^{m\times 1}$ is the measurement noise, $\boldsymbol{x}^{true}$ is the so-called vector of the "true" parameters, and $\boldsymbol{\varepsilon_g} \in \Re^{m\times 1}$ is the model error (model-related error). The original ESMDA formulation that we present in this section neglects the term $\boldsymbol{\varepsilon_g}$. Later, we circumvent this problem for TLS data assimilation introducing the additive model-error parameter in our methodology.

ESMDA is a Bayesian method that aims at sampling the posterior distribution of the parameters, considering the data and the prior information. One can interpret it as a minimum ensemble solution of the cost function

$$\mathcal{J}(\boldsymbol{x}_j) = \left(\boldsymbol{x}_j - \boldsymbol{x}_j^f\right)^T \boldsymbol{C}_{xx}^{-1}\left(\boldsymbol{x}_j - \boldsymbol{x}_j^f\right) + \left(\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j\right)^T \boldsymbol{C}_{dd}^{-1}\left(\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j\right), \tag{2.2}$$

where $\boldsymbol{C}_{xx} \in \Re^{n\times n}$ is the parameter covariance matrix, $\boldsymbol{d}_j$ is a vector of perturbed observed data, whose mean is $\boldsymbol{d}^{obs}$ and covariance matrix is $\boldsymbol{C}_{dd} \in \Re^{m\times m}$ (EVENSEN, Geir, 2018). The first term on the right side of equation (2.2) relates to the distance from the prior estimate, while the second term relates to the data mismatch, evidencing the Bayesian formulation.

In the ESMDA algorithm, one updates the parameters at each i iteration, or multiple assimilations, using the analysis equation

$$\boldsymbol{X}^{i+1} = \boldsymbol{X}^i + \boldsymbol{\rho} \circ \boldsymbol{K}^i \left(\boldsymbol{D}^i - \left(\boldsymbol{D}^{sim}\right)^i\right), \tag{2.3}$$

where $\boldsymbol{\rho}$ is the localization matrix and $\boldsymbol{K}^i$ is the Kalman gain defined as

$$\boldsymbol{K}^i = \boldsymbol{C}_{xy}^i\left(\boldsymbol{C}_{yy}^i + \alpha^i \boldsymbol{C}_{dd}\right)^{-1}. \tag{2.4}$$

Note that, in the first update, we use the results from the prior ensemble on the right side of equation (2.3). In equation (2.3), $\boldsymbol{D}^i$ is a matrix of perturbed observed data, whose columns are samples of the distribution $d_j^i \sim \mathcal{N}\left(\boldsymbol{d}^{obs}, \alpha^i \boldsymbol{C}_{dd}\right)$.

We obtain the covariance matrices $\boldsymbol{C}_{xy}^i$ and $\boldsymbol{C}_{yy}^i$ from the ensemble estimates. First, we define the matrix of ensemble anomalies

$$\boldsymbol{A}^i = \boldsymbol{X}^i \frac{1}{\sqrt{N-1}}\left(\boldsymbol{I}_N - \frac{1}{N}\boldsymbol{1}_N \boldsymbol{1}_N^T\right) = \boldsymbol{X}^i \boldsymbol{\Pi}_N, \tag{2.5}$$

where $\boldsymbol{1}_N$ is a column vector with size N whose elements are one and $\boldsymbol{I}_N$ is the identity matrix with size $N \times N$. We also define the matrix of predicted ensemble anomalies

$$\boldsymbol{Y}^i = \left(\boldsymbol{D}^{sim}\right)^i \boldsymbol{\Pi}_N. \tag{2.6}$$

Then, it is possible to estimate the covariance matrices

$$\boldsymbol{C}_{xy}^i = \boldsymbol{A}^i\left(\boldsymbol{Y}^i\right)^T, \tag{2.7}$$

and

$$\boldsymbol{C}_{yy}^{i} = \boldsymbol{Y}^{i}\left(\boldsymbol{Y}^{i}\right)^{T}, \tag{2.8}$$

The inflation factor α^i is predefined based on the number of iterations, N_{mda}, and following the condition

$$\sum_{i=1}^{N_{mda}} \frac{1}{\alpha^i} = 1. \tag{2.9}$$

Here, we apply distance-based Kalman gain localization to alleviate the impact of spurious correlations due to the limited ensemble size. Moreover, we performed all data assimilations using constant inflation factors and $N_{mda} = 10$, which is a conservatively high value to assure stable results.

2.4 Methodology to mitigate the effect of model error

The general methodology to mitigate the effect of spatially correlated model-related errors in seismic data assimilation comprises five steps, as represented in Figure 2.1. We start by performing conventional data assimilation, where we neglect the influence of model-related errors. We then analyze the residuals from this data assimilation to perform a qualitative diagnosis and define the model-error parameter, $\boldsymbol{q}$. In the next step, we repeat the data assimilation, including $\boldsymbol{q}$ as an uncertain parameter to calibrate, adopting a weak-constraint formulation. Next, we consider the possibility of improving the data assimilation through an iterative update of the covariance matrix of total observation errors, following the workflow from Oliver and Alfonzo (2018a). In step five, we verify the final ensemble, validate the results, and perform the production forecast under uncertainty.

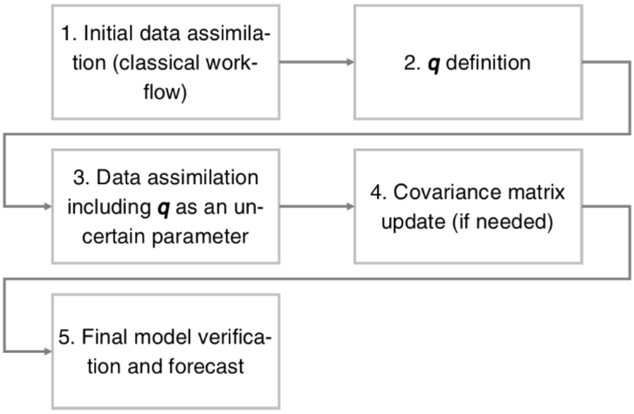


Figure 2.1: Methodology for mitigating model error impacts.

We acknowledge that different sources of model-related errors can influence well data assimilation (EVENSEN, G., 2020; LU; CHEN, 2020). Nevertheless, our methodology focuses on treating the correlated model-related errors in TLS data assimilation.

In the following subsections, we provide more details of each step shown in Figure 2.1.

2.4.1 Initial data assimilation

In the first step of Figure 2.1, we perform a standard data assimilation process using an iterative ensemble smoother. We opted to apply ESMDA in our tests, as described in the section 2.3, but one could choose different ensemble-based methods (CHEN, Yan; OLIVER, 2013; EVENSEN, Geir *et al.*, 2019; LUO *et al.*, 2015).

Our forward model, $\boldsymbol{g}$, comprises the reservoir flow simulator and a petroelastic model. The former provides the production rates, along with pressure, saturation, and fluid properties that are the PEM inputs for computing the acoustic impedance variations. The parameters, which form the vector $\boldsymbol{x}$, are reservoir properties, such as porosity, permeability, fault transmissibilities, relative permeability curves, pore compressibility, and well indexes. The prior distribution of these parameters follows the geological knowledge and an ensemble of flow models represents it. One generates each sample of the grid parameters using geostatistical simulation and assigns values to scalar parameters randomly. Note that the grid parameters are updated directly in the reservoir simulation scale, and only one three-dimensional grid exists during the history-matching process. Nevertheless, the null cell positions are different in each realization, following the porosity and permeability values. The data to assimilate, $\boldsymbol{d}^{obs}$, originate from the wells and time-lapse seismic. The differences between the observed and simulated data cause changes in the parameter vector, as defined in equation (2.3).

2.4.2 Definition of the model-error parameter

In the initial data assimilation, Figure 2.1 step 1, we solved the strong-constraint problem, in which the forward model is assumed perfect and all deviations between simulated and observed data cause parameter changes. Nevertheless, if this premise is not valid, inconsistent parameter updates will occur. Moreover, the final residuals will exhibit higher magnitudes and correlations, revealing the model-error effects (OLIVER, D.S.; ALFONZO, 2018a). Therefore, in this subsection, we use the residual from the initial data assimilation to define the error-related parameter $\boldsymbol{q}$. It is worth mentioning that we consider that the model errors only impair the TLS simulation. Therefore, $\boldsymbol{q}$ will not affect the production data assimilation, or equivalently, it will be zero for all the well data points.

To assume an imperfect forward model, one can define an error model with uncertain parameters, which will change during the data assimilation workflow (EVENSEN, Geir, 2019). As with any parameter, it is necessary to define the error parameters' prior distributions and how these parameters will affect the forward model outputs. For simplicity, we define $\boldsymbol{q}$ as an additive term in the simulated acoustic impedance variations

$$\left(\boldsymbol{d}_j^{sim}\right)_{TLS} = g\left(\boldsymbol{x}_j\right)_{TLS} + \boldsymbol{q}_j, \tag{2.10}$$

where the imperfect model computes the acoustic impedance variation $g\left(\boldsymbol{x}_j\right)_{TLS}$. Therefore, $\boldsymbol{q}$ will relate directly to the neglected model error $\boldsymbol{\varepsilon_g}$ in equation (2.1) and to the residuals. Consequently, we will use the behavior of the residuals to define this parameter's prior distribution.

After the initial data assimilation, we define the vector of updated parameters $\left(\boldsymbol{x}_j^a\right)^{(0)}$, which are the columns of the matrix $(\boldsymbol{X}^a)^{(0)}$. The TLS residuals from the initial data assimilation are

$$\boldsymbol{p}_j^{(0)} = (\boldsymbol{d}^{obs})_{TLS} - g\left(\left(\boldsymbol{x}_j^a\right)^{(0)}\right)_{TLS}. \tag{2.11}$$

We focus on the magnitude of the mean of the residuals with significant spatial correlations, which is an indicator of disparity between the forward model and the data (OLIVER, D.S.; ALFONZO, 2018a). First, we compute the ensemble mean of the residuals

$$\overline{\boldsymbol{p}}^{(0)} = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{p}_j^{(0)}, \tag{2.12}$$

which represents part of the observed data that the ensemble of models could not match, on average. In other words, a high magnitude $\overline{\boldsymbol{p}}^{(0)}$ indicates a significant bias in the parameters, model imperfection, or observation bias.

Rammay et al. (2019) mentioned that it is essential to limit the degrees of freedom to avoid overfitting when defining an error model. We limit the degrees of freedom by focusing on the long-range correlated residuals in this work. Here we refer as long-range to the impedance variations occurring in a portion of the reservoir more extensive than the fluid substitution signal, such as the waterflood acoustic impedance increase around injectors. The exact criterium to define the long-range will vary in each case due to the fluid substitution signal sizes, the presence of other variations in the reservoir, and grid size.

We filter the vector $\overline{\boldsymbol{p}}^{(0)}$ to remove the spatially uncorrelated fluctuations, variations related to the fluid fronts in the reservoir, or other signals with small areal influence. We define the filtered version of the residuals mean as $\overline{\overline{\boldsymbol{p}}}^{(0)}$. We apply a 2D moving average filter in the residuals images because our data set comprises a sequence of time-lapse seismic maps, one for each seismic horizon. Nevertheless, other low-pass image filters could be applied. We chose the moving average window that is long enough to attenuate all the signals with small areal influence to the point that they are imperceptible in the maps. The moving average window size is a tuning parameter in the methodology, and we discuss its influence in the results section.

After filtering the ensemble mean of the residuals, we define it as our prior estimate of the mean value of the additive error parameter, $\overline{\boldsymbol{q}}^{f} = \overline{\overline{\boldsymbol{p}}}^{(0)}$. This definition implies that the higher the model imperfection, the higher the mean residuals will be, which relates to the criterium that Rammay et al. (2021) used to select part of the initial residuals due to model imperfection. Moreover, one should note that $\overline{\boldsymbol{q}}^{f}$ is only the prior estimate of the error-related parameter's mean, and this variable will be calibrated in the next data assimilation (step 3 of Figure 2.1).

We assume that the additive error parameter pertains to a Gaussian distribution with the prior mean $\overline{\boldsymbol{q}}^{f}$. Consequently, it is still necessary to define the prior standard deviation and a variogram model. Here, we assume that the first guess of the standard deviation is proportional to the prior mean. This choice causes uncertainty in $\boldsymbol{q}^{f}$ to be higher where the average residuals are also higher in magnitude. Finally, we estimate the initial variogram model of $\boldsymbol{q}^{f}$ using a match of the experimental variogram of $\overline{\boldsymbol{q}}^{f}$.

We acknowledge that the definition of the prior distribution of $\boldsymbol{q}$ is arbitrary, and other definitions can work better in different cases. Nevertheless, these choices have some favorable characteristics that improved the data assimilation quality in our tests and should benefit future practical applications. Firstly, $\boldsymbol{q}$ will assume high absolute values only in regions where the absolute value of the filtered mean residuals is also high. Therefore, our methodology will not significantly influence cases where the model imperfections and observation bias are minor. Secondly, by setting the variogram of $\boldsymbol{q}^{f}$ with roughly the same format of the experimental variogram from $\overline{\overline{\boldsymbol{p}}}^{(0)}$, we limit the degrees of freedom of the additive error parameter to prevent it from compensating for fluid-front-related signals. This process will avoid overfitting. In this case, overfitting through $\boldsymbol{q}$ would cause weak updates of the reservoir parameters and poor model calibration.

After defining a prior distribution for the error-related parameters from the initial data assimilation residuals, we repeat the data assimilation, including $\boldsymbol{q}$. Note that the new data assimilation will be necessary only if the prior $\boldsymbol{q}$ distribution assumes high absolute values. We describe this process in the following subsection.

2.4.3 Data assimilation including $\boldsymbol{q}$ as an uncertain parameter

We acknowledge that it is crucial to improve the models as much as possible before adopting other techniques to overcome model imperfections in data assimilation. During the model improvement, one should consider reviewing the parameter's prior probability distributions, including other uncertain parameters, and adding unmodelled processes affecting the observed data. In this subsection, we assume that one performed all the feasible model improvements before performing new data assimilation, including $\boldsymbol{q}$.

Figure 2.2 shows the workflow for repeating the data assimilation, including the error-related parameter $\boldsymbol{q}$ as an uncertain variable to calibrate. The forward model $\boldsymbol{g}$ is the same as the initial data assimilation, comprising the reservoir flow simulator and the PEM, but we include the additive term $\boldsymbol{q}$ that accounts for model imperfections in the seismic data simulation (Figure 2.2 step 2). Since $\boldsymbol{q}$ is also an uncertain parameter, ESMDA will update the augmented parameter matrix $\boldsymbol{Z}^T = (\boldsymbol{X}^T, \boldsymbol{Q}^T)$, where $\boldsymbol{Q}$ is a matrix whose columns are the realizations of the additive error parameter $\boldsymbol{q}_j$. Therefore, the ESMDA update equation becomes

$$\boldsymbol{Z}^{i+1} = \boldsymbol{Z}^i + \boldsymbol{\rho} \circ \boldsymbol{C}_{zy}^i \left(\boldsymbol{C}_{yy}^i + \alpha^i \boldsymbol{C}_{dd}\right)^{-1} \left(\boldsymbol{D}^i - \left(\boldsymbol{D}^{sim}\right)^i\right), \tag{2.13}$$

where the covariance that relates the parameters to the data are ensemble estimates,

$$\boldsymbol{C}_{zy}^i = \boldsymbol{Z}^i \boldsymbol{\Pi}_N \left(\boldsymbol{Y}^i\right)^T. \tag{2.14}$$

Apart from equations (2.13) and (2.14), this data assimilation follows the same procedure described in sections 2.3 and 2.4.1. This method follows the theoretical foundation provided by Evensen (2019).

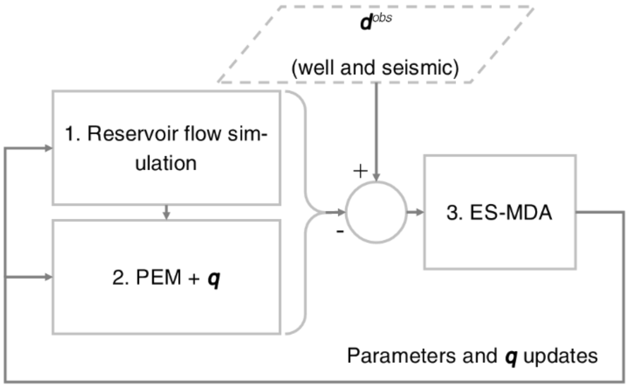


Figure 2.2: Data assimilation workflow with a weak-constraint formulation.

2.4.4 Update of the covariance matrix of observation errors

After performing the data assimilation, including $\boldsymbol{q}$ to compensate for bias and model errors with long-range correlations, it is also possible to update the covariance matrix of observation errors (Figure 2.1, step 4). The objective of this step is to account for remaining correlated errors that $\boldsymbol{q}$ was unable to compensate for due to the limited degrees of freedom, if necessary. We consider this step case-dependent because the error parameter's inclusion may be enough to compensate for the significant correlated error effects. Moreover, this step follows the general methodology to calibrate imperfect models from Oliver and Alfonzo (2018a).

The residuals from the last data assimilation are

$$\boldsymbol{p}_j^{(k)} = \boldsymbol{d}^{obs} - g\left(\left(\boldsymbol{x}_j^a\right)^{(k)}\right) - \left(\boldsymbol{q}_j^a\right)^{(k)}, \tag{2.15}$$

where k is the data assimilation number with the error parameter. According to Oliver and Alfonzo (2018a), it is possible to estimate the covariance matrix of total observation errors from

$$\boldsymbol{C}_D^{(k+1)} \approx \frac{1}{N} \boldsymbol{p}_j^{(k)} \left(\boldsymbol{p}_j^{(k)}\right)^T. \tag{2.16}$$

Nevertheless, in most practical applications, this estimate will not be full rank because of $N \ll m$. Therefore, we adopted an analytical model for the covariance matrix that we estimate by fitting an experimental variogram of the residuals. An exponential model provided a good match in our experiments, and other studies described a similar procedure (LU; CHEN, 2020; SILVA NETO, Gilson M.; DAVOLIO; SCHIOZER, 2020). Furthermore, we used the average of the diagonal elements from the matrix of the right side equation (2.16) as a representative value for the diagonal elements of the updated covariance matrix. Alfonzo and Oliver (2020) also assumed a stationarity condition when assimilating TLS data. This condition may simplify the algorithm for generating the perturbed observation data, and it avoids issues related to close to zero variance values in points of the grid where there are minor changes in the saturation and pressure in the ensemble of models.

After estimating $\boldsymbol{C}_D^{k+1}$, we may repeat the data assimilation, replacing $\boldsymbol{C}_{dd}$ for the updated covariance matrix of the total observation error. Note that $\boldsymbol{C}_{dd}$ is commonly a diagonal matrix whose diagonal elements are the variance of the measurement noise or an inflated value (ALFONZO; OLIVER, 2020; SILVA NETO, Gilson M.; DAVOLIO; SCHIOZER, 2020; SUN; VINK; GAO, 2017). As this is an iterative procedure, one may repeat this process until the changes in the data assimilation results are negligible.

2.4.5 Final model verification and production forecast

In the last step of Figure 2.1, we perform a final verification of the models, checking their consistency based on data match, parameter updates, prior available information, and technical knowledge about the case. In this step, we recommend checking the magnitude of $\boldsymbol{q}_j^a$ and comparing it to the observed data to understand which data aspects the models were unable to match. These results facilitate an interdisciplinary analysis, where the reservoir engineer may provide feedback to the geologists and geophysicists regarding the reservoir model capability to match the time-lapse seismic data. Furthermore, the ensemble average of $\boldsymbol{q}_j^a$ provides information about how the error parameter changed the model's output to provide an appropriate match to the observed data.

After validation, one will be able to use the calibrated ensemble of reservoir models to perform production forecasts.

2.5 NQDS metric

In this work, we use the Normalized Quadratic Deviation with Sign (NQDS) to assess the well data misfit (AVANSI; MASCHIO; SCHIOZER, 2016). The NQDS is a quadratic norm, which includes a sign that identifies if a model is mostly overestimating or underestimating the data. We define NQDS as

$$\begin{aligned}(NQDS_j)_l = &\frac{\mathbf{1}_{m_l}^T\left[\left(\boldsymbol{d}_j^{sim}\right)_l - (\boldsymbol{d}^{obs})_l\right]}{\left|\mathbf{1}_{m_l}^T\left[\left(\boldsymbol{d}_j^{sim}\right)_l - (\boldsymbol{d}^{obs})_l\right]\right|} \\ &\times \frac{\left[\left(\boldsymbol{d}_j^{sim}\right)_l - (\boldsymbol{d}^{obs})_l\right]^T\left[\left(\boldsymbol{d}_j^{sim}\right)_l - (\boldsymbol{d}^{obs})_l\right]}{[(\boldsymbol{d}^{obs})_l \times \tau_l + \varphi_l]^T[(\boldsymbol{d}^{obs})_l \times \tau_l + \varphi_l]},\end{aligned} \tag{2.17}$$

where $\boldsymbol{d}_j^{sim}$ is the simulated data by model j in the ensemble, $\boldsymbol{d}^{obs}$ is the observed data, the subscript l indexes a data type, such as the water rate of a particular well, m_l is the number of data points of this type, τ is a tolerance value, and φ is a constant. The tolerance relates to the accuracy of the measurements and pertains to the interval $\tau \in [0,1]$, while the constant is a small positive number that avoids division by zero. These scalars are user-defined, and we used tolerances of 10% for water rate and 5% for bottom hole pressure. The constants were 10 for water rate and 0.01 for bottom hole pressure. These are typical values for the benchmark case that we worked with (AVANSI; MASCHIO; SCHIOZER, 2016; FORMENTIN *et al.*, 2019).

It is worth mentioning that this work only applied the NQDS metric to evaluate well data history matching. However, it is also possible to apply it to quantify TLS data match in maps, as demonstrated in (ALMEIDA; DAVOLIO; SCHIOZER, 2020; DANAEI *et al.*, 2020).

2.6 Applications

We tested our methodology using two examples: a simple 2D case and a realistic field-scale application. In the first test, the low computational requirements enable us to test different configurations of the methodology to assess the sensitivity to some of its parameters. Also, the field-scale case mimics practical application challenges.

2.6.1 2D case

The 2D test case is a toy problem mimicking some reservoir application challenges. We built a simplistic 2D reservoir model with 40×40 cells, where we included four producers and four injectors in a staggered line drive pattern depicted in Figure 2.3. The wells operate with a fixed bottom-hole pressure. We generated a random Gaussian porosity field that represented our synthetic truth, which is unknown during the data assimilation. The logarithm of permeability for the synthetic truth was a linear function of porosity, with an added Gaussian perturbation. We provide additional information about this case in Table 2.1. We simulated this simplistic reservoir model using MATLAB Reservoir Simulation Toolbox (LIE, 2019).

We ran the synthetic truth for two years of production with waterflooding and computed the normalized impedance variation using a linear proxy defined by

$$\Delta I_{PNORM} = a_s \frac{(S_w - S_{wi})}{\max(S_w - S_{wi})} + a_p \frac{(p - p_i)}{\max(|p - p_i|)}, \tag{2.18}$$

where a_s is the saturation sensitivity and a_p is the pressure sensitivity (DANAEI *et al.*, 2020). We used $a_s = 0.85$ and $a_p = 0.15$ in the synthetic truth. The observed data comprised this normalized impedance variation with added Gaussian noise with zero mean, standard deviation of 0.06, and an exponential five-cell horizontal correlation (Figure 2.3 a). We did not include production data in this example.

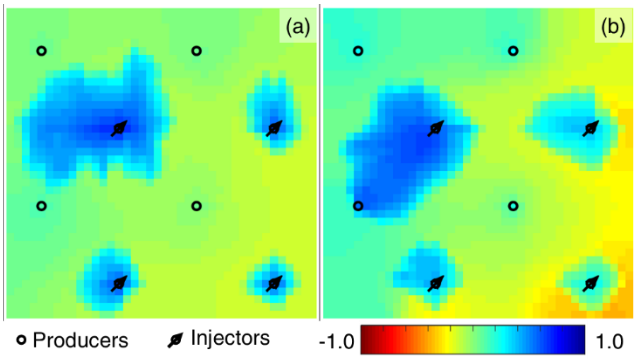


Figure 2.3: 2D case – synthetic acoustic impedance difference observed data (a) and one example of prior simulated data with model error (b).

Table 2.1: Additional information about the 2D example.

Characteristic	Value
Cell size	$50 \times 50 \times 50\ m$
Average porosity	0.15
Average permeability	$132\ mD$
Fluid mobility ratio (water/oil)	5
Fluid densities	Water: $1000\ kg/m^3$ Oil: $700\ kg/m^3$
Initial pressure	$200\ bar$

The forward model comprised a reservoir flow simulation model and a PEM, as shown in Figure 2.2. In this example, we used 100 reservoir models built from limited information about the synthetic truth because only the properties at the well locations were known. We used a simple Gaussian simulation to generate samples of the properties at the rest of the grid. We used an exponential variogram with a six-cell correlation length and a maximum standard deviation of 0.05 for the porosity. The log-permeability and porosity relation followed the same pattern of the synthetic truth. The uncertain parameters to calibrate were the porosity and the horizontal permeability at each cell. Figure 2.4 (a) displays the initial mean porosity distribution.

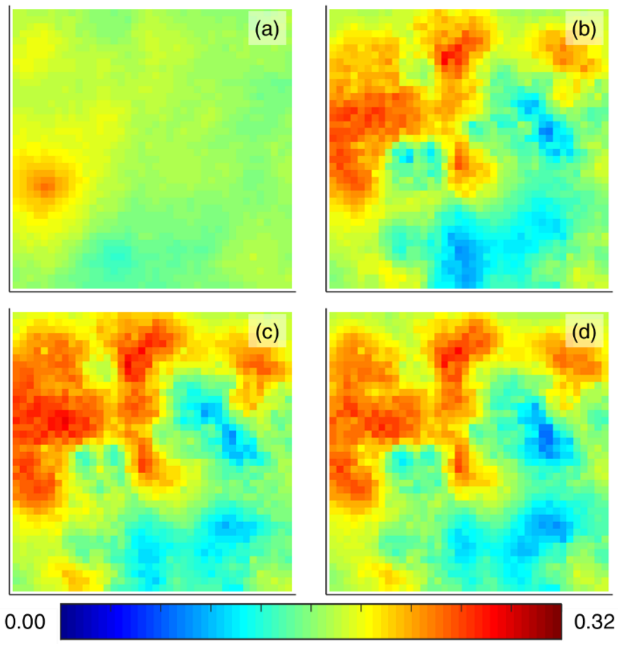


Figure 2.4: 2D case – mean porosity maps: (a) prior distribution, (b) posterior distribution with a "correct" forward model, (c) posterior distribution neglecting the model error, and (d) posterior distribution including q parameter to compensate for the model errors.

The PEM of the forward model was identical to the synthetic truth, equation (2.18), in the quasi-ideal case, where there are no significant model errors. Aiming at mimicking a model error effect, we changed the pressure sensitivity term of equation (2.18) to $a_p = 0.3$. Figure 2.3 b shows the normalized acoustic impedance difference from one simulation model with the model error before the data assimilation. Although this pressure sensitivity mismatch was artificial in this example, this is a common challenge in real field applications, as discussed in (MALEKI; DAVOLIO; SCHIOZER, 2019).

2.6.2 Field-scale case

In our field-scale example, we used the benchmark case UNISIM-I-H (MASCHIO, C. *et al.*, 2013). The authors built this case using information from a real offshore reservoir in the Namorado Field, Campos Basin, Brazil. Avansi and Schiozer (2015) described the synthetic truth for this benchmark. The observed data that we refer to here as "measured" comes from this fine-scale reference model. Although we do not use information about the "true" parameters during the data assimilation, this realistic benchmark allows us to compare the final results with the reference for validation purposes. However, we do not compare the estimated grid parameters directly with the synthetic truth due to scale differences.

The simulation models of the UNISIM-I-H benchmark were built using partial information from the synthetic truth at the location of the 14 producers and 11 water injectors.

The wells' boundary conditions are the total liquid rate for the producers and water rate for the injectors during the history. During the forecast, these boundary conditions change to fixed bottom-hole pressure.

The uncertain grid attributes to calibrate are the net-to-gross ratio, the directional permeabilities, and the porosity. Moreover, there are uncertain scalar parameters, such as the water-oil contact at the east of the reservoir, the rock compressibility, the water relative permeability exponent and terminal value, and well productivity and injectivity. In this benchmark case, one describes these uncertain properties using an ensemble of 500 reservoir models whose grid parameters were generated using geostatistical simulation (MASCHIO, C. *et al.*, 2013). We simulated the reservoir models using IMEX version 2017 (CMG, COMPUTER MODELLING GROUP LTD., 2017).

We provide additional information about this case in Table 2.2. One can find more details about the case in the following references (AVANSI; SCHIOZER, 2015; MASCHIO, C. *et al.*, 2013). Furthermore, other studies in the literature applied this benchmark in data assimilation tests using ensemble-based methods (DANAEI *et al.*, 2020; EMERICK, Alexandre A., 2019, 2018; SILVA NETO, Gilson M.; DAVOLIO; SCHIOZER, 2020; SOARES; MASCHIO; SCHIOZER, 2019).

Table 2.2: Additional information about the field-scale example.

Characteristic	Value
Cell size	$100 \times 100 \times 8\ m$
Average porosity	0.14
Average horizontal permeability	$26\ mD$
Average vertical permeability	$5\ mD$
Fluid mobility ratio (water/oil)	From 0.9 to 3.0 (uncertain)
Fluid densities	Water: $1010\ kg/m^3$ Oil: $866\ kg/m^3$
Initial pressure	$321\ bar$

In this example, our observed data includes oil rate, water cut, gas-oil ratio, and bottom-hole pressures for the producers; water rate and bottom-hole pressure for the injectors. All observed data contains random noise. The end of the historical period in this study is 2618 days after the start of the production. At this time, we computed two synthetic time-lapse seismic data sets. The first, called PEM data, originated from the petroelastic model applied to the reference model and upscaled to the simulation grid. The second, called INV data, includes seismic modeling and inversion after the PEM to account for the resolution loss and interference between different reservoir layers. The seismic modeling comprises a vertical 1D convolution with a wavelet. Besides, we applied the coloured inversion algorithm (LANCASTER; WHITCOMBE, 2000) to obtain relative acoustic impedance data. One may find more information about this process in the studies (DAVOLIO; SCHIOZER, 2019; DE SOUZA, 2018). Both seismic data sets were upscaled to the reservoir simulation grid, and we work with four acoustic impedance variation maps, seismic horizons, as our observed data.

We compare the two seismic data sets in Figure 2.5. There are significant differences between the PEM and the INV data sets. Nevertheless, our forward simulation model only considers the petroelastic model after the reservoir flow simulation to estimate the seismic response, a common practice in data assimilation. Therefore, when we apply the INV data set, our model becomes significantly imperfect due to missing physics. The difference between the PEM and the INV data sets represents the model error considered in this example. However, we do not use the knowledge about the PEM data when we assimilate the INV data.

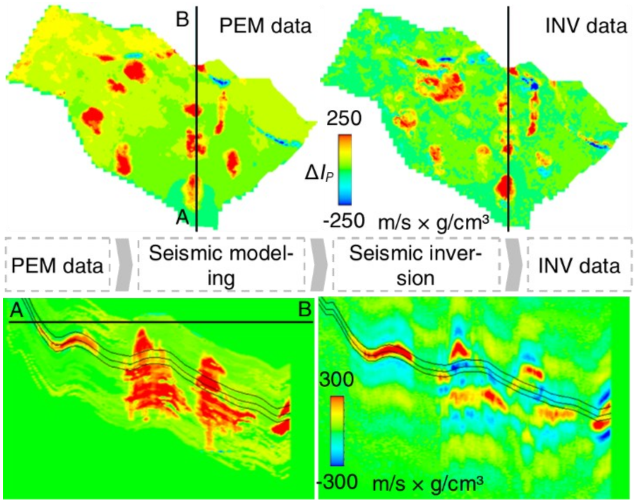


Figure 2.5: Field-scale application – comparison between the two data sets, the quasi-ideal data, PEM data, is on the left, and the realistic data, INV data, is on the right.

In this field-scale example, the synthetic data generation and the forward simulation models use the same petroelastic model. The PEM applies the known Gassmann equation (GASSMANN, 1951) to compute the saturated rock properties and relations close to the Hertz-Mindlin model (MAVKO; MUKERJI; DVORKIN, 2009) to describe the dry rock. The Batzle-Wang correlations (BATZLE; WANG, 1992) represent the fluid properties. Furthermore, we consider two minerals, shale and quartz, and we estimate the equivalent mineral properties using the mean of the Hashin-Shtrikman bounds (HASHIN; SHTRIKMAN, 1963; apud AVSETH; MUKERJI; MAVKO, 2005).

One may perform seismic data assimilation at different levels, such as seismic amplitude, impedance, or pressure and saturation (GOSSELIN *et al.*, 2003). Here, we compare time-lapse acoustic impedance changes, which is a popular choice because it avoids the seismic modeling in the forward simulation and dispenses the inversion to pressure and saturation.

2.7 Results and discussion

2.7.1 2D case

In the 2D case, we admitted a significant model error in the pressure sensitivity term of the PEM, a_p in equation (2.18). In other words, the forward model has a more pronounced pressure sensitivity than the reference model used to generate the synthetic data. We assimilated the synthetic seismic data of Figure 2.3 (a) twice using the imperfect forward model and a "correct" model.

2.7.1.1 Influence of the spatially correlated model errors in the TLS data assimilation

We start our analysis by verifying the possible effects of the model's disparity in data assimilation using the standard data assimilation method. The red histogram Figure 2.6 shows the effect of the model error in the porosity distribution if one ignores it by applying a strong-constraint classical formulation. The figure depicts a histogram of the relation between the calibrated mean porosity, influenced by model errors, divided by the mean porosity in a quasi-ideal case. There is a noticeable tendency to overestimate the porosity due to the model's discrepancy in the former test, causing a porosity relation greater than one.

In this case, it is easy to provide a physical interpretation of the shift in Figure 2.6 since we know the cause of the model error. The model predicts a higher pressure-related impedance variation that is not present in the data. One way to compensate for this difference is to increase the porosity, reducing the flow simulation's pressure variations. However, since the discrepancy arises from a petroelastic model error, the data assimilation process causes unphysical parameter changes. One could argue that it would be straightforward to consider uncertainty in the pressure-related parameter of the PEM, which could also be heterogeneous in the reservoir. However, we did not adopt this procedure to admit the cases where it is complex to identify the cause of the error or where there are multiple causes.

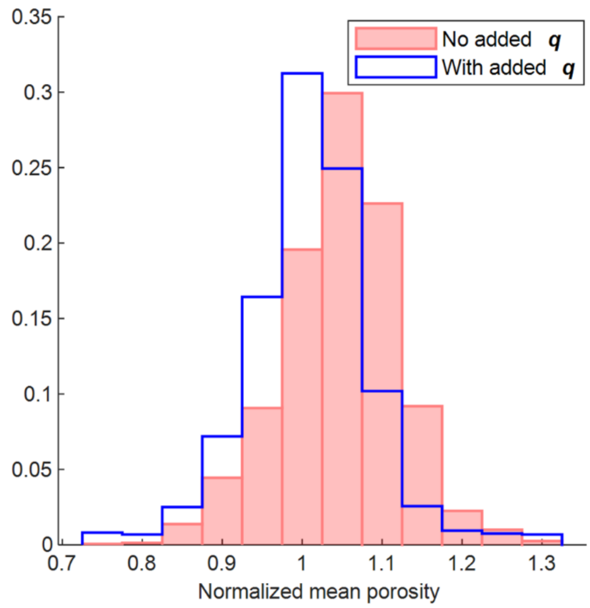


Figure 2.6: 2D case – normalized mean porosity histograms.

We can also analyze the model error effect in the maps (b) and (c) of Figure 2.4, which have the same tendency to increase the porosity in a particular path in the west, north, and northeast regions of the model, when compared to the prior (a). Nevertheless, it is noticeable that the high porosity region is more pronounced in the reddish areas in the (c) map due to the correlated error effect. Furthermore, the low porosity sections, in blue, are attenuated in the (c) map when compared to (b).

The differences depicted in Figure 2.4 and Figure 2.6 increased the overall porosity deviation by 7.3% compared to the data assimilation with the "correct" model. We estimate this incremental deviation through

$$\Delta\delta = \frac{\delta}{\delta_{ideal}} - 1, \tag{2.19}$$

where $\Delta\delta$ is the deviation increase, δ is the absolute average deviation, and δ_{ideal} is the absolute average deviation of the quasi-ideal case. We computed these porosity deviations from the "true" porosity field on a model-by-model basis.

2.7.1.2 Mitigation of the model error effects

After analyzing the model error impacts in the initial data assimilation, we applied our methodology to compensate for it. We started by computing the residuals' ensemble average from the last data assimilation with the imperfect PEM, as shown in Figure 2.7 (a). Since we want to limit the degrees of freedom in the $\boldsymbol{q}$ parameter, we apply a moving average to filter these residuals, as shown in Figure 2.7 (b). We used a five-cell length in this moving average. After that, we define the map in Figure 2.7 (b) as the $\boldsymbol{q}$ mean and standard deviation, following the intuition that the uncertainty in $\boldsymbol{q}$ should be higher where the mean residuals are also higher. The variogram model of $\boldsymbol{q}$ is also defined using a experimental variogram match from the data in Figure 2.7 (b). In this case, we used a Gaussian model,

$$\frac{\xi}{\sigma^2} = 1 - \exp\left(-\frac{h^2}{L^2 a}\right), \tag{2.20}$$

with $L = 18$ cells and $a = 0.35$, where we measure the distance h in number of model cells, ξ is the semivariogram, and σ^2 is the variance. We generated a prior ensemble of $\boldsymbol{q}$ with 100 realizations and applied it in new data assimilation, including the additive term as an uncertain parameter.

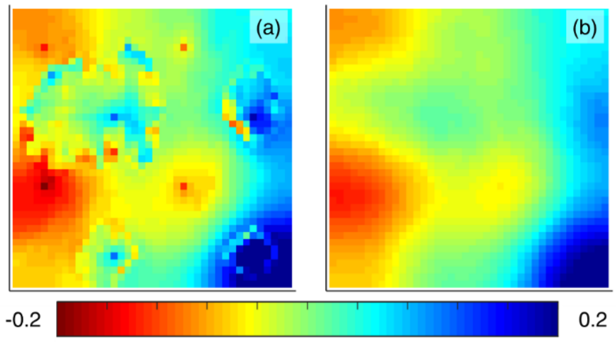


Figure 2.7: 2D case – definition of the prior q distribution from the residuals of the initial data assimilation: (a) ensemble mean of the residuals and (b) filtered mean residuals.

The blue histogram in Figure 2.6 shows the comparison between the new data assimilation, including $\boldsymbol{q}$ as a parameter, and the quasi-ideal case. The error-related term addition corrected the porosity overestimation tendency by centralizing the porosity relation around one. Comparing maps (b) and (d) in Figure 2.4, we identify only minor differences, confirming the improvement from the map (c), where we neglected model error. Furthermore, the deviation in the porosity distribution from the synthetic truth in the case where we added the $\boldsymbol{q}$ parameter increased by 0.1%, which is significantly lower than the 7.3% increment when we neglected the model error.

These results indicate that it was beneficial to add the $\boldsymbol{q}$ term in the forward model and include it in the data assimilation. However, it is essential to verify if our methodology would jeopardize the data assimilation if we apply it to a case where the spatially correlated model errors are minor.

2.7.1.3 Methodology test in a case without significant spatially correlated model error

We repeated the procedure described in the last subsection when the forward PEM is identical to the synthetic truth. We call this situation a quasi-ideal case because we minimized the known pressure-related model error, but other error sources in the data may exist. Our objective was to verify if the addition of $\boldsymbol{q}$ would impair the data assimilation.

The first difference when we applied our methodology to the quasi-ideal case occurred in the definition of the prior $\boldsymbol{q}$ distribution, as depicted in Figure 2.8. Comparing Figure 2.8 (b) and Figure 2.7 (b), we notice that the magnitude of $\boldsymbol{q}$ is greatly reduced when the model errors are negligible, which makes for the reduced impact of the methodology. The insignificant change due to $\boldsymbol{q}$ is desirable, as there are no significant model error causes.

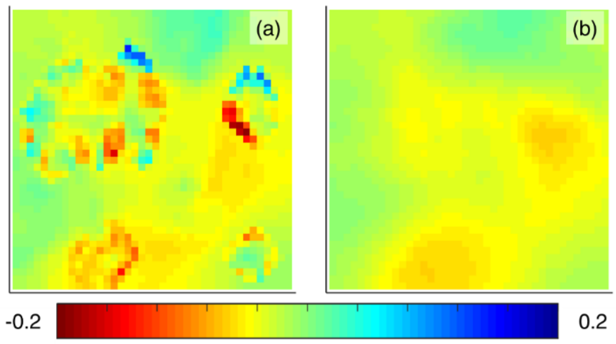


Figure 2.8: 2D case – definition of the prior q distribution from the residuals of the data assimilation without model error: (a) ensemble mean of the residuals and (b) filtered mean residuals.

We checked the impact of the methodology in this quasi-ideal case in Figure 2.9. The mean porosity maps (a) and (b) are similar, which indicates that our methodology did not play a significant role in this case. It is possible to confirm this observation by checking the porosity deviation from the synthetic truth on a model-by-model basis. The inclusion of the model-error parameter increased the porosity deviation by only 3.7%, which we consider minor compared to the risk of neglecting the model error (7.3% deviation increase). One of the causes of the former increment is the reduction of the impact of the data, which occurs in the weak-constraint formulation.

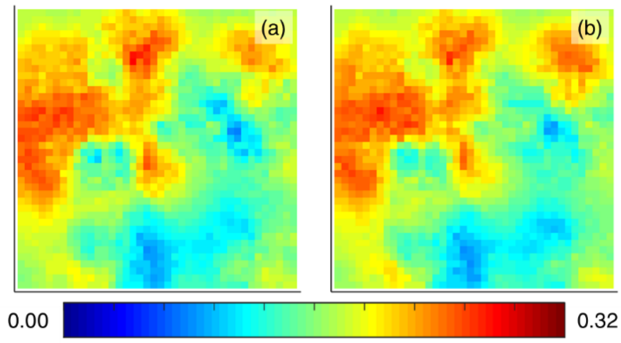


Figure 2.9: 2D case – mean porosity maps: (a) posterior distribution with a "correct" forward model and (b) posterior distribution with a "correct" forward model and added $\boldsymbol{q}$.

2.7.1.4 Varying the parameters of the $\boldsymbol{q}$ prior distribution

When we applied our methodology in the 2D case, some prior $\boldsymbol{q}$ distribution parameters were arbitrary: the moving average window's length and the relation between the standard deviation and the mean prior values. Therefore, we repeated the data assimilation with different combinations of these properties, verifying the impact on the results. Here we focus on the incremental deviation in the porosity mean, comparing it with the quasi-ideal case.

First, we varied the moving average window length from zero, no moving average, to nine cells, as shown in Figure 2.10 (a). If the moving average window is too small, it is possible that the $\boldsymbol{q}$ parameter has enough degrees of freedom to compensate for part of the data deviations caused by the reservoir parameters, weakening their updates. On the other hand, if the moving average window is too large, $\boldsymbol{q}$ will not be able to compensate for the spatial variations of the model error effect on the impedance variations. Therefore, we opted to use a moving average window ranging from three to seven cells (five cells in the previous tests). It is worth mentioning that the vertical axis in Figure 2.10 quantifies the deviation change compared to the quasi-ideal case. Therefore, a small negative value indicates that applying the proposed methodology was slightly better than the quasi-ideal case. However, we consider that this minor improvement is due to the problem's statistical nature, and the proposed methodology does not aim to provide a better parameter estimation than a model-error-free case.

We also varied the ratio between the standard deviation and the mean of the prior $\boldsymbol{q}$ distribution in Figure 2.10 (b). A low standard deviation causes $\boldsymbol{q}$ to exhibit low variation from the estimated mean distribution and not to change significantly during the data assimilation. A higher prior $\boldsymbol{q}$ standard deviation increases the overall variability, leading to more significant $\boldsymbol{q}$ updates and lowering the reservoir parameter updates. In this test, all the data assimilations with added $\boldsymbol{q}$ improved the data assimilation results, but applying the standard deviation equal to the mean led to slightly better results.

This sensitivity test does not aim to find a general rule for the influence of the moving average window and the standard deviation magnitude because one cannot generalize Figure 2.10 results to other more complex cases. However, it is crucial to note that all the cases with added $\boldsymbol{q}$ led to improved data assimilation results regarding the estimated porosity quality. This result indicates that it is vital to consider the model error in the data assimilation process, even if it is not feasible to optimize the error parameter's prior distribution.

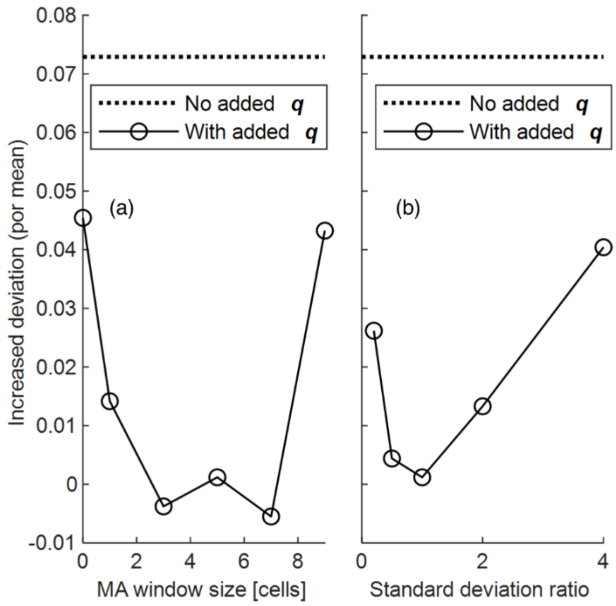


Figure 2.10: 2D case – the influence of moving average window (a) and standard deviation and mean ratio (b).

2.7.2 Field-scale case

2.7.2.1 Influence of the spatially correlated model errors in the initial TLS data assimilation

We start the field-scale test analysis verifying the effect of the model-related errors in the production data assimilation quality, shown in Figure 2.11. The results are presented in terms of NQDS, as described in section 2.5. Only water cut and bottom-hole pressure are present in the figure because these are the challenging data to match in the current benchmark case. We compare the prior distribution, in grey, with the well data assimilation, in blue, and the well and TLS data assimilation, neglecting the model-related errors in red. The ± 5 NQDS horizontal lines in the figure represent a standard quality threshold. Looking at the water production data, we see a good match for most models, with close to zero deviations after the well data assimilation with or without seismic. Furthermore, the TLS data seems to improve the history matching of PROD010. One plausible explanation for this improvement is the fact that the TLS data includes information regarding the waterfronts around the producer, which may shift the solution towards a better well data match when there is a biased prior ensemble. We did not identify evidence of the model error effect in the water rate deviations during the history data period.

Figure 2.11 (b) presents the bottom hole pressure (BHP) deviations. Comparing the prior, the well data assimilation, and the well and TLS data assimilation ignoring the model error, it is possible to notice a systematic tendency to obtain a positive deviation in the latter. Although relatively small in magnitude, this tendency indicates that the pressure is systematically higher than the data, possibly indicating higher average pressure in the porous medium.

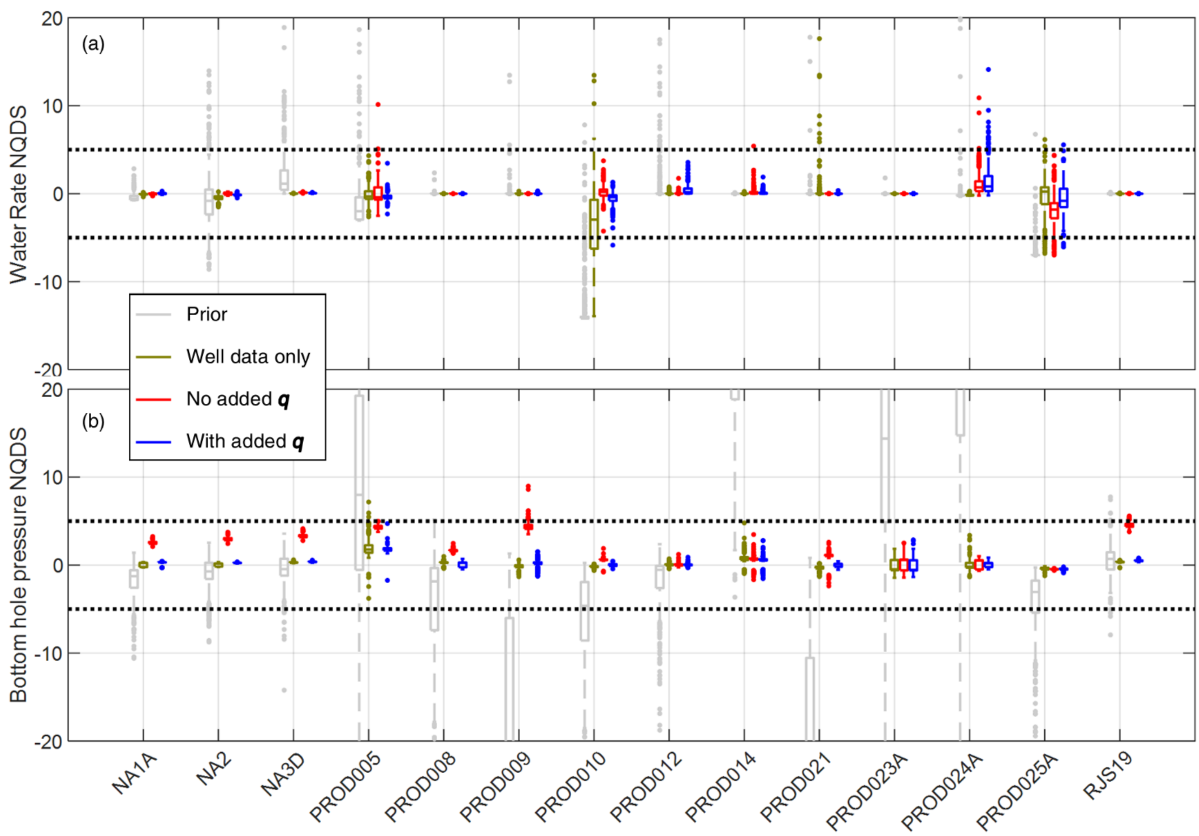


Figure 2.11: Field-scale case – history matching NQDS: water rate (a) and well bottom hole pressure (b).

The pressure behavior relates to the overestimated pore volume that we evidence in the Volume of Oil in Place (VOIP) cumulative distribution curves in Figure 2.12. It is clear that the seismic data, assimilated ignoring model errors, led to a significantly increased volume that impacted the wells' bottom hole pressure, in Figure 2.11 (b). This VOIP increase due to TLS data has a physical interpretation. On average, the forward model simulated acoustic impedance variations with higher absolute values than the observed data because the former does not include seismic resolution loss and inter-layer interference. Therefore, the data assimilation process tends to increase the volume to decrease simulated data values.

In Figure 2.12, the assimilation of well data alone in this particular case tends to reduce the reservoir volume. Since this is a synthetic case, it was possible to include the synthetic truth volume as a reference in the figure.

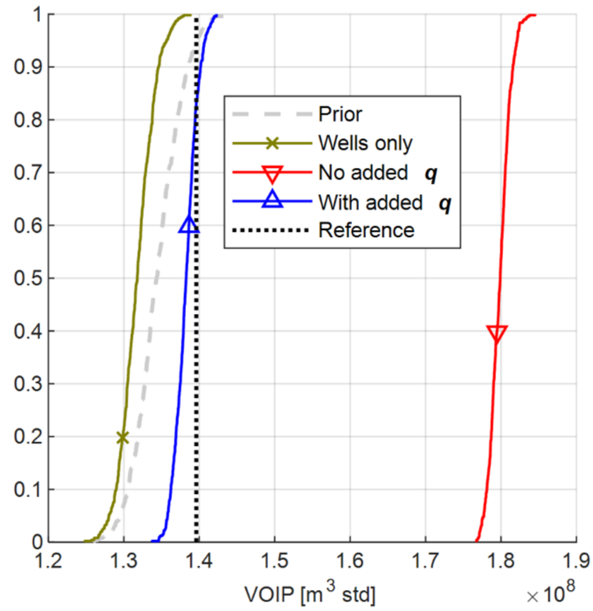


Figure 2.12: Field-scale case – Volume of Oil in Place (VOIP) cumulative distribution curves.

Figure 2.13 displays PROD010 water cut and bottom-hole pressure curves as an example of history matching quality. We chose this well because it is the worst one in terms of the water cut match, and the prior ensemble shows relatively high deviations in the bottom-hole-pressure match for this producer. Corroborating the data in Figure 2.11, the well data assimilation alone greatly improved the PROD010 data match in terms of water production and pressure for most of the 500 models. The TLS data incorporation improved the water-cut match from the case with only production data. However, higher pressure values at the end of the history period in (e) are noticeable, resulting from the exaggerated volume increase without model-error treatment.

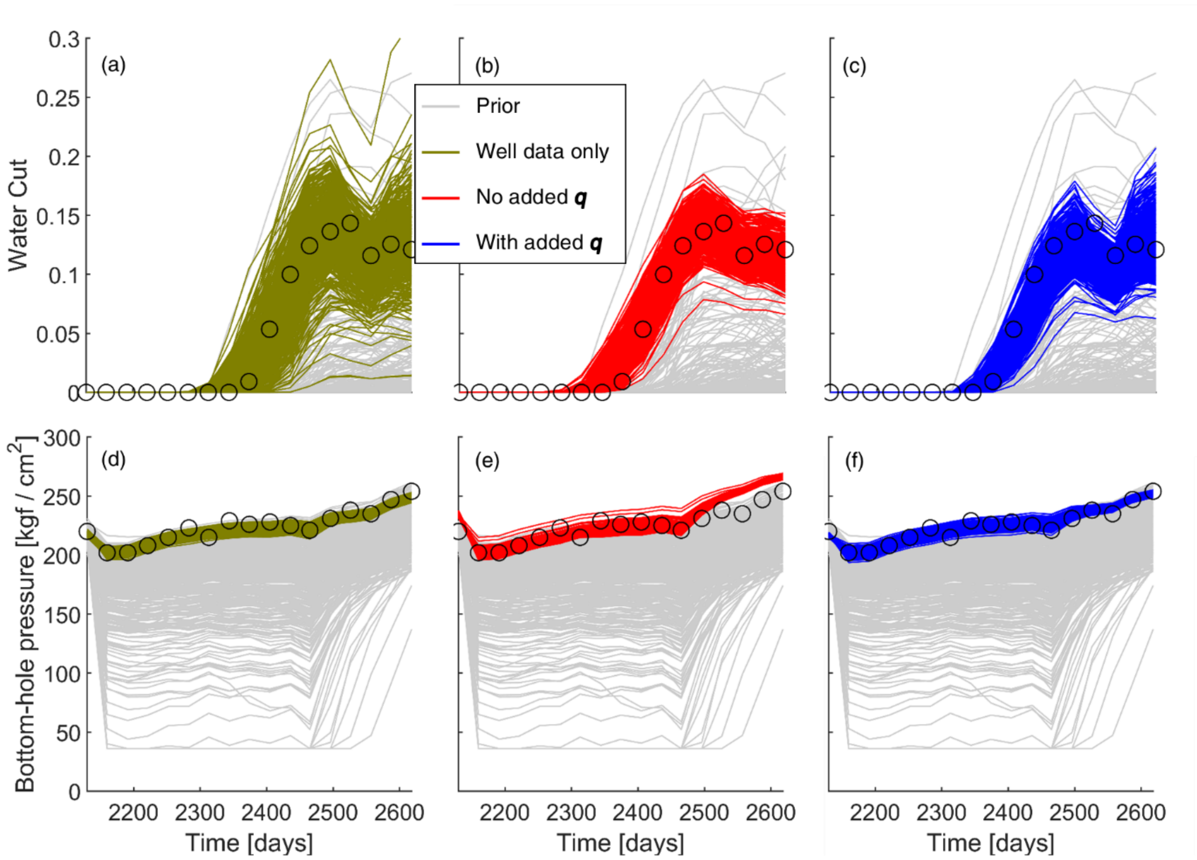


Figure 2.13: Field-scale case – PROD010 data match. The figure presents water-cut (a, b, and c) and bottom-hole pressure (d, e, and f). It compares the prior ensemble with the data assimilation results using only well data (a and d), using well and TLS data without *q* (b and e), and with *q* (c and f).

Figure 2.14 enables the analysis of how the data assimilation changed the porosity distribution in the reservoir simulation models. It only shows one reservoir layer, but it is enough to analyze the tests' main features. The well data assimilation mean porosity map (e) is similar to the prior (a), which is a consequence of the lack of spatial information from this data source. The minor standard deviation decrease in (f) compared to (b) corroborates the previous interpretation. There are noticeable differences in the two random samples, (g) compared to (c) and (h) compared to (d), but the updated models maintain the porosity trends from the initial models.

When we compare the models calibrated with well and seismic data, neglecting model errors, with the Prior and Well ensembles, we notice significant differences in Figure 2.17. Firstly, the TLS data caused a major standard deviation reduction (j) compared to (b), resulting from the large amount of data distributed in space. Secondly, the neglected model error led to a porosity increase throughout the reservoir (i, k, and l).

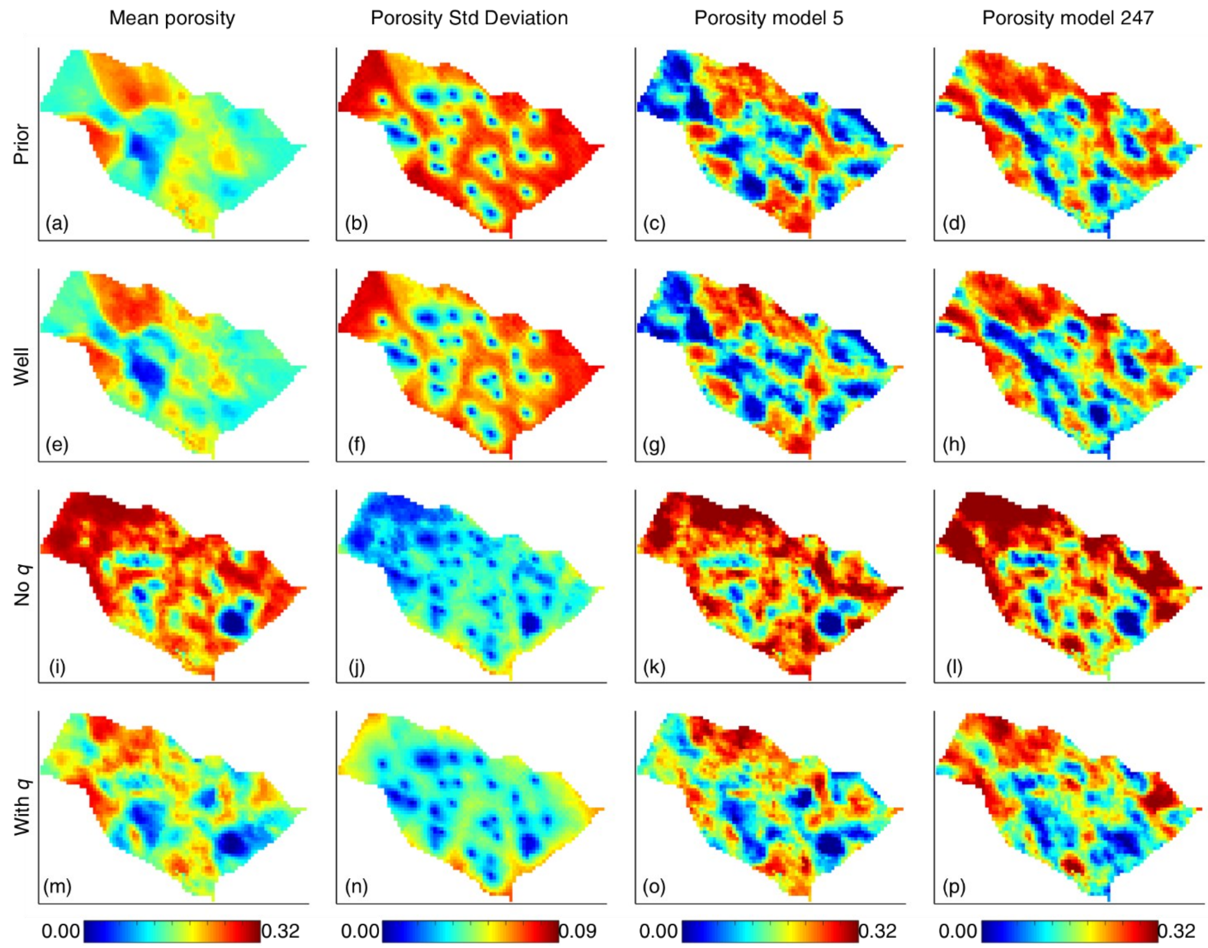


Figure 2.14: Field-scale case – layer 8 porosity maps. The figure displays the mean porosity (a, e, i, and m), standard deviation (b, f, j, and n), sample 5 (c, g, k, and o), and 247 (d, h, l, and p) out of 500 models. It compares the prior ensemble (a, b, c, and d), the ensemble calibrated with well data (e, f, g, and h), the ensemble incorporating well and TLS data without *q* (i, j, k, and l), and with *q* (m, n, o, and p).

Figure 2.15 provides the same information as Figure 2.14, but for the permeability. We applied a logarithmic transformation to this variable due to its log-normal distribution. In this field example, the horizontal permeability is partially correlated to the porosity. Similarly to the porosity analysis, the well data assimilation did not cause significant permeability changes (e, g, and h) compared to the prior ensemble (a, c, and d). It is possible to notice a standard deviation reduction in (f) compared to (b), as expected in any data assimilation, but it was minor. Again, the inclusion of TLS data neglecting model errors caused a major permeability increase (j, k, and l) and standard deviation reduction (j).

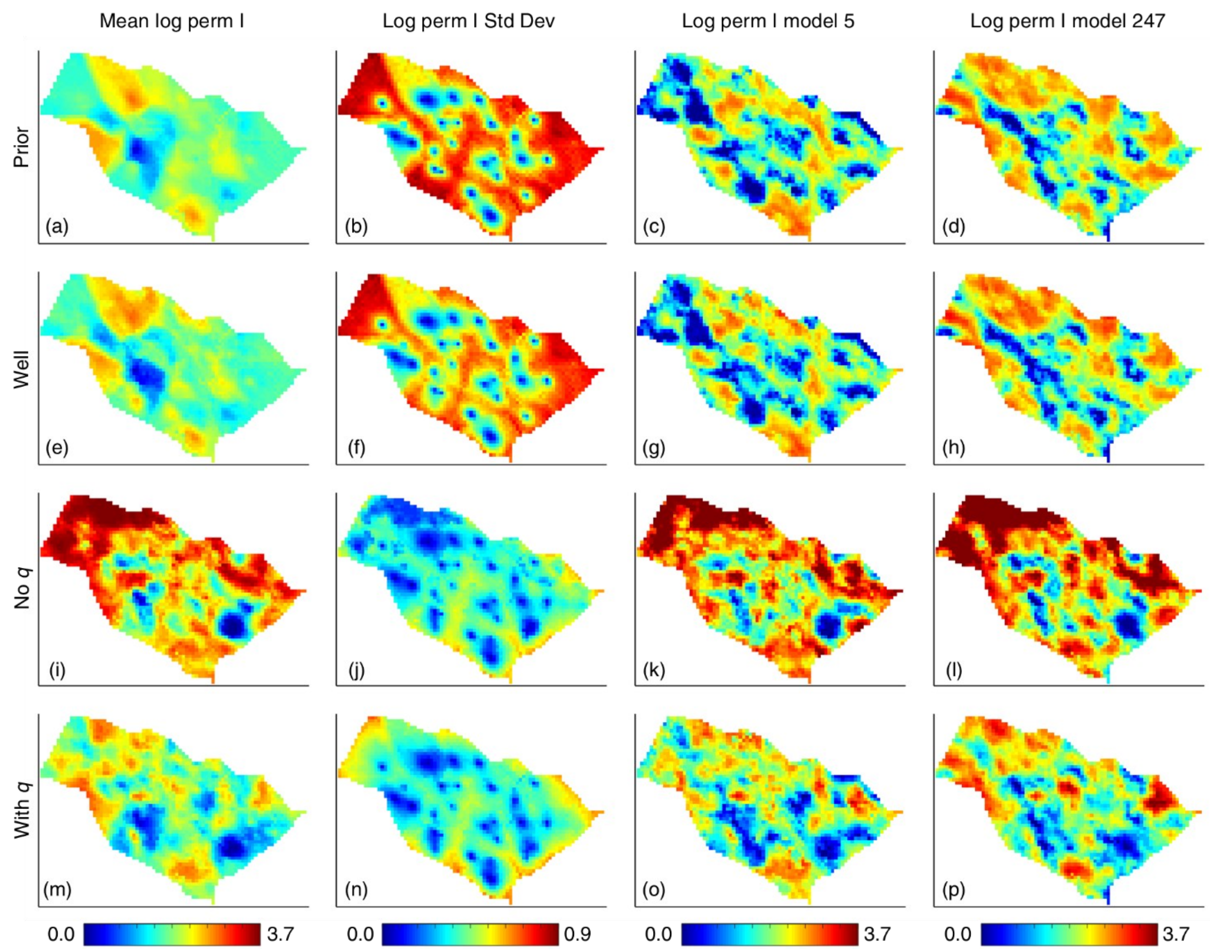


Figure 2.15: Field-scale case – layer 8 logarithm base 10 of the horizontal permeability maps. The figure displays the mean logarithm of the permeability (a, e, i, and m), standard deviation (b, f, j, and n), sample 5 (c, g, k, and o), and 247 (d, h, l, and p) out of 500 models. It compares the prior ensemble (a, b, c, and d), the ensemble calibrated with well data (e, f, g, and h), the ensemble incorporating well and TLS data without q (i, j, k, and l), and with q (m, n, o, and p).

The assimilation of TLS data ignoring the model errors also impairs the well rate forecasts, as shown in figures 2.16, 2.17, and 2.18. Figure 2.16 (a) compares the cumulative oil production forecasts after the assimilation of well data and this data together with TLS neglecting the model-related errors. One may notice that the added TLS data led to models that overestimate the oil production due to the exaggerated reservoir volume (Figure 2.12). This behavior is not compatible with the synthetic truth, whose response is also shown in Figure 2.16 (a).

We continue the analysis of the oil production forecast on a well-by-well basis in Figure 2.17. We compute the normalized average deviations using

$$\delta_l = \frac{\sum_i^{m_l} \left| d_l^{sim} - d_l^{obs} \right|}{\sum_i^{m_l} d_l^{obs}}, \tag{2.21}$$

where m_l is the number of points of this type. Note that the assimilation of TLS and well data neglecting model errors led to normalized average deviations worse than when we calibrated the models only with well data. We observed the same behavior in the water rate forecast in Figure 2.18. These results indicate that, in our example, the inclusion of time-lapse seismic data with an imperfect forward model led to a worse ensemble of models than one would obtain using only the well data.

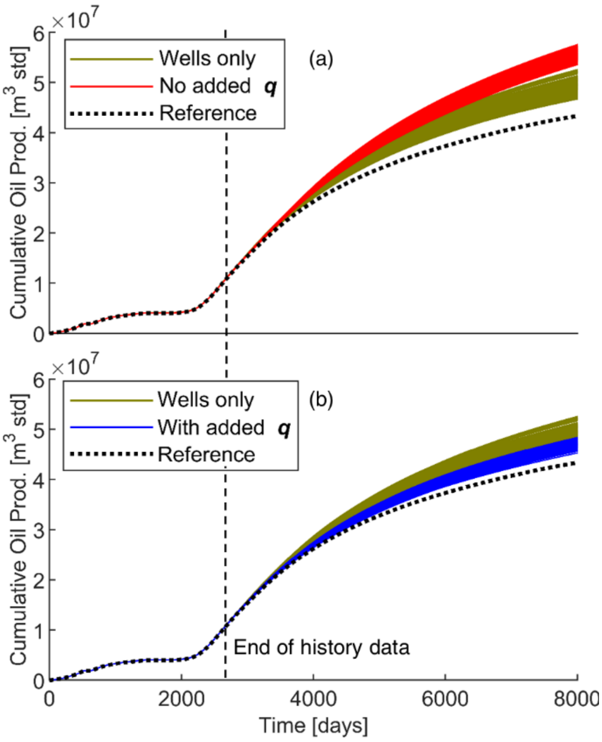


Figure 2.16: Field-scale case – cumulative oil production, comparing the models after assimilating well data with the models after assimilating TLS and well data, neglecting model errors (a), and considering it through the added q parameter (b).

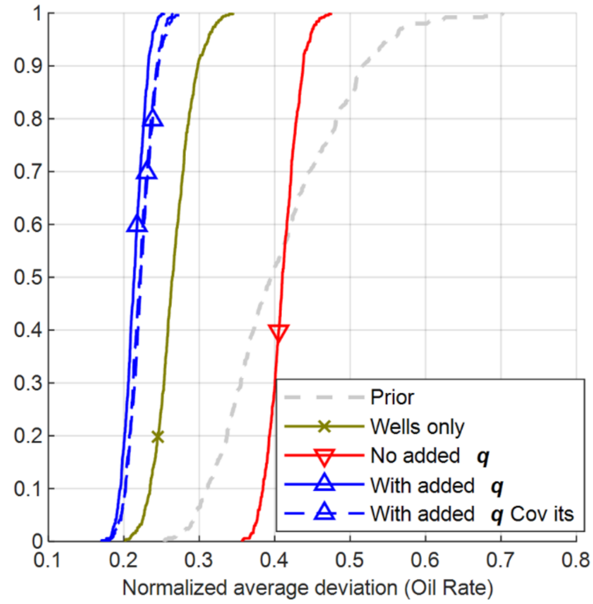


Figure 2.17: Field-scale case: normalized average deviation in the oil rate forecast (well-by-well comparison).

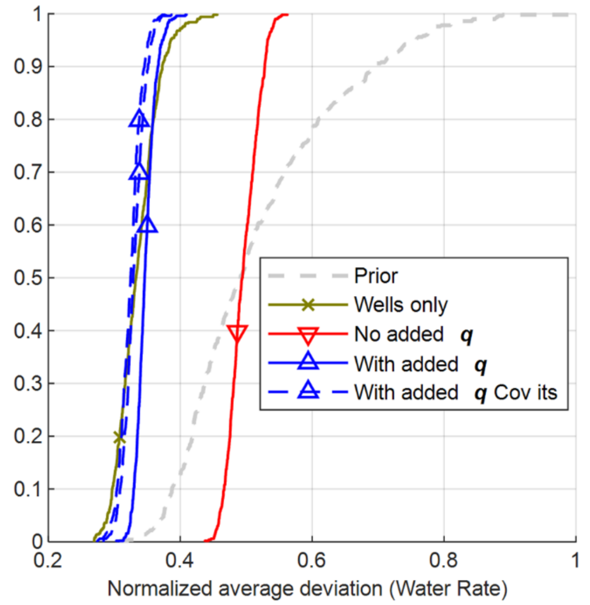


Figure 2.18: Field-scale case – normalized average deviation in the water rate forecast (well-by-well comparison).

It is worth mentioning that in order to validate the updated models in a field application, one should check the data mismatch, the parameter changes, and the production forecast. In the current analysis, the two calibrated ensembles, well only and including TLS data neglecting model errors, provided acceptable data mismatch since most of the models led to lower mismatch than the ± 5 NQDS threshold for all the wells and data types (Figure 2.11). However, the exaggerated VOIP increase in Figure 2.12, porosity increase in Figure 2.14 (j, k, and l), and permeability increase in Figure 2.15 (j, k, and l) evidence the unplausible parameter changes due to model imperfection. Since this example is a benchmark case, investigated in previous data assimilation studies (BERTOLINI; MASCHIO; SCHIOZER, 2015; CAVALCANTE *et al.*, 2017; EMERICK, Alexandre A., 2019, 2018; MASCHIO, Célio; SCHIOZER, 2018, 2019, 2016; OLIVEIRA; SCHIOZER; MASCHIO, 2017; SOARES; MASCHIO; SCHIOZER, 2019), it is possible to validate the models by comparing the production forecast with the reference results. In the current application, this analysis demonstrated that the TLS data assimilation neglecting model errors jeopardized the forecast compared to the case with only well data. In a practical application, where the forecast data is not available, it is possible to perform a similar test by reserving the last data points from the history to check the models.

2.7.2.2 Mitigation of the model error effects

To mitigate the model error effects that jeopardized the TLS data assimilation in the previous section, we started by computing the residuals using equation (2.11). We then computed the ensemble mean of these residuals, as shown in Figure 2.19 (a), (c), (e), and (g), respectively, for maps 1 to 4. These mean residuals include acoustic impedance differences related to the reservoir waterfronts present in the observed and simulated data, affecting relatively small areas. Moreover, significant long-range correlated variations indicate a discrepancy between the forward model and the observed data or strong bias in the parameters' prior distribution.

We define the prior distribution for $\boldsymbol{q}$ focusing on these long-range correlated differences. We start by filtering the mean residuals, defining the moving average window length to remove the fluid front-related variations. In this case, an eight-cell long window was enough. We show the filtered residuals in Figure 2.19 (b), (d), (f), and (h) for the four TLS maps. We then define these maps as the $\boldsymbol{q}$ prior mean and standard deviation. Moreover, we use the experimental variogram computed using the filtered maps, which was a Gaussian model with a range of roughly 14 cells and $a = 0.45$.

When compared to the analysis of the 2D case, it is possible to say that we started with conservative values for the prior $\boldsymbol{q}$ distribution. First, the moving average window is relatively large, limiting the degrees of freedom of $\boldsymbol{q}$. Moreover, by setting its prior standard deviation to the same value as the mean, we also impose a limit on how much $\boldsymbol{q}$ can compensate for the current residuals. In the 2D case, we tested standard deviations up to four times the estimated mean and obtained improved calibration in all tests.

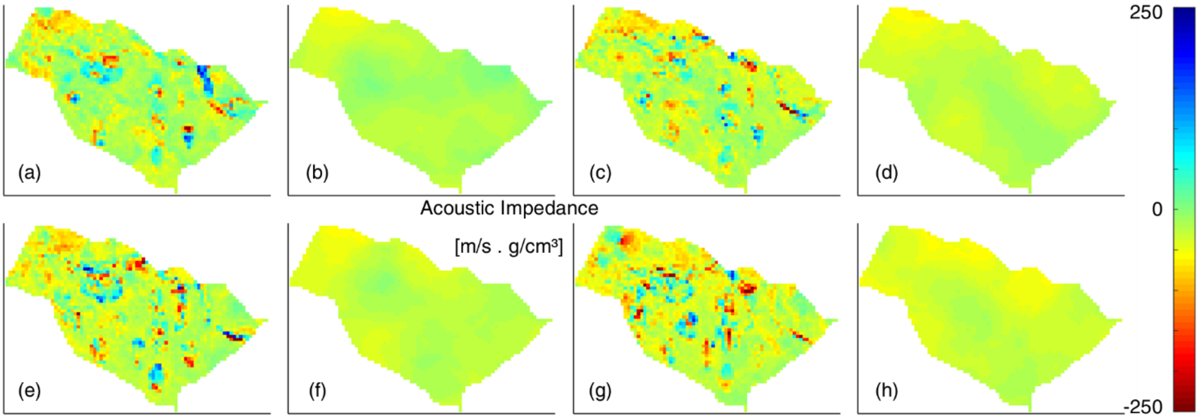


Figure 2.19: Field-scale case – mean acoustic impedance residuals of the initial data assimilation and filtered mean residuals using an eight-cell window moving average, we show horizon 1 (a and b), 2 (c and d), 3 (e and f), and 4 (g and h), respectively.

After estimating $\boldsymbol{q}$ prior distribution, we repeated the data assimilation process, including the additive error in the forward PEM and calibrating it during the workflow, as shown in Figure 2.2. This new data assimilation with added $\boldsymbol{q}$ led to well history matching results similar to the previous cases, as Figure 2.11 depicts. In terms of water rate match (Figure 2.11 (a)), the results were roughly the same as incorporating TLS data without accounting for model errors. Nevertheless, we noticed a slight improvement of the NQDS metric for the BHP, indicating that the inclusion of $\boldsymbol{q}$ corrected pressure tendency.

The pressure corrections that we identified in the BHP data are related to the VOIP adjustment, as shown in Figure 2.12. The addition of $\boldsymbol{q}$ in the data assimilation improved the reservoir characterization in terms of pore volume distribution. Our calibrated models exhibited volumes consistent with the prior distribution and the reference value after calibrating with TLS data, accounting for spatially correlated model errors.

Looking at a particular well behavior, PROD010 exhibited a good water-cut match in Figure 2.13 (c), comparable to the case without $\boldsymbol{q}$ (b) and better than the assimilation of production data alone (a). Furthermore, the model error treatment corrected the positive pressure deviation at the end of the history, as one can notice in Figure 2.13 (f) compared to (e). This improvement is a consequence of the volume correction.

We can check the model error treatment effect on the porosity and permeability distributions in Figure 2.14 and Figure 2.15, respectively. The model error treatment corrected the tendency to overestimate porosity and permeability (m, o, and p compared to i, k, and l). As expected, the weak-constraint formulation attenuated the standard deviation reduction (n compared to j). However, the uncertainty reduction is still significantly more pronounced in (n) than in the Well case (f), which one expects due to the spatially distributed information from TLS.

Figure 2.14 and Figure 2.15 also enable analyzing how the TLS data assimilation, mitigating the model error effects, updated the reservoir properties. The seismic spatially distributed information caused a significant change in the parameter distributions (m, o, and p) compared to the prior ensemble (a, c, and d). For instance, the prior ensemble assumes intermediate porosity and permeability mean values in the southwest. However, the seismic information led to low porosity and permeability values in this region. Nevertheless, comparing the two calibrated samples (o and p) to the prior models (c and d), the parameter distributions seem plausible. In other words, one could admit that the four models, the two prior and the two updated, are possible realizations of the same reservoir.

It is essential to check if the incorporation of TLS data with our methodology helped improve the production forecast compared to when only well data calibrated the models. We first analyze the forecast in terms of cumulative oil production in Figure 2.16 (b). The TLS data assimilation with added $\boldsymbol{q}$ reduced the oil production overestimation observed in the models calibrated only with production data. This improvement, although significant, was not enough to eliminate the oil production overestimation tendency in this example.

The time-lapse seismic data improved the oil rate forecast on a well-by-well basis when we added $\boldsymbol{q}$, as depicted in Figure 2.17. The deviations in oil rates were smaller than when only well data were available, indicating that the TLS data improved reservoir characterization. In this particular case, the TLS did not significantly improve the water rate forecast compared to the case where only well data is present, as shown in Figure 2.18. However, both calibrated ensembles are significantly better than the prior ensemble in forecasting water production.

It is crucial to notice that all results of figures 2.16, 2.17, and 2.18 indicate that the addition of $\boldsymbol{q}$ significantly improved the well forecasts compared to the case when we ignored the model errors, which is the classical approach. Furthermore, one can notice that the mean $\boldsymbol{q}$ is not zero, indicating that this application includes model bias. However, our methodology does not distinguish between model errors and bias, compensating for both effects. This behavior is typical of other weak-constraint approaches (EVENSEN, Geir, 2019).

2.7.2.3 Analyzing the effect of $\boldsymbol{q}$ in the observed data

Since $\boldsymbol{q}$ is an additive term in our forward model, we can interpret it as a modification to either the model or the data. Therefore, it is possible to compare the final distribution of the $\boldsymbol{q}$ parameter with the observed data to understand what kind of signal we are compensating for. Furthermore, we can use the final value of $\boldsymbol{q}$ to reinterpret the data with the geophysicists by using the geological models' information to identify uncertain aspects of the seismic attribute. We highlight that we did not change our observed data, as we assimilated the actual INV data in all the previous tests. Nevertheless, by formulating the problem as a weak constraint, we incorporate in $\boldsymbol{q}$ the data aspects that we cannot explain using our flow models and our forward petroelastic model.

In the first column of Figure 2.20 (a, d, g, and j), we show our observed data from the seismic inversion. If we reinterpret this data using the mean value of the calibrated $\boldsymbol{q}$ to remove the characteristics that the forward model could not explain, we obtain the maps in the second column of Figure 2.20 (b, e, h, and k). We do this mathematically by subtracting the mean of $\boldsymbol{q}$ from the data. Note that this "modified" data resembles the quasi-ideal case, which is the PEM data that we show in the third column of Figure 2.20 (c, f, i, and l). The most significant change is the bluish acoustic impedance difference in the areas that do not exhibit impedance variations related to water replacing oil. One can identify these fluid-related signals as the more intense bluish impedance increase, surpassing $250\frac{m}{s}\frac{g}{cm^3}$, or by their position around the injectors. The forward seismic modeling and the seismic inversion distorted the data and led to values close to zero, or negative, in regions without water saturation increase in the INV data. The addition of $\boldsymbol{q}$ compensated for this tendency. Therefore, one could say that $\boldsymbol{q}$ mimicks the unmodeled effects of seismic resolution loss and inversion in this example.

It is worth mentioning that, in a practical case, the quasi-ideal case would not be available, and one would be able to analyze the two first columns of Figure 2.20. Nevertheless, the observed data analysis with the $\boldsymbol{q}$ modification would contribute to multidisciplinary work with engineers, geologists, and geophysicists to improve the models and the observed data. In this process, it may be possible to improve the model by including missing physics and new uncertain parameters and enhance the data through reinterpretation and reprocessing.

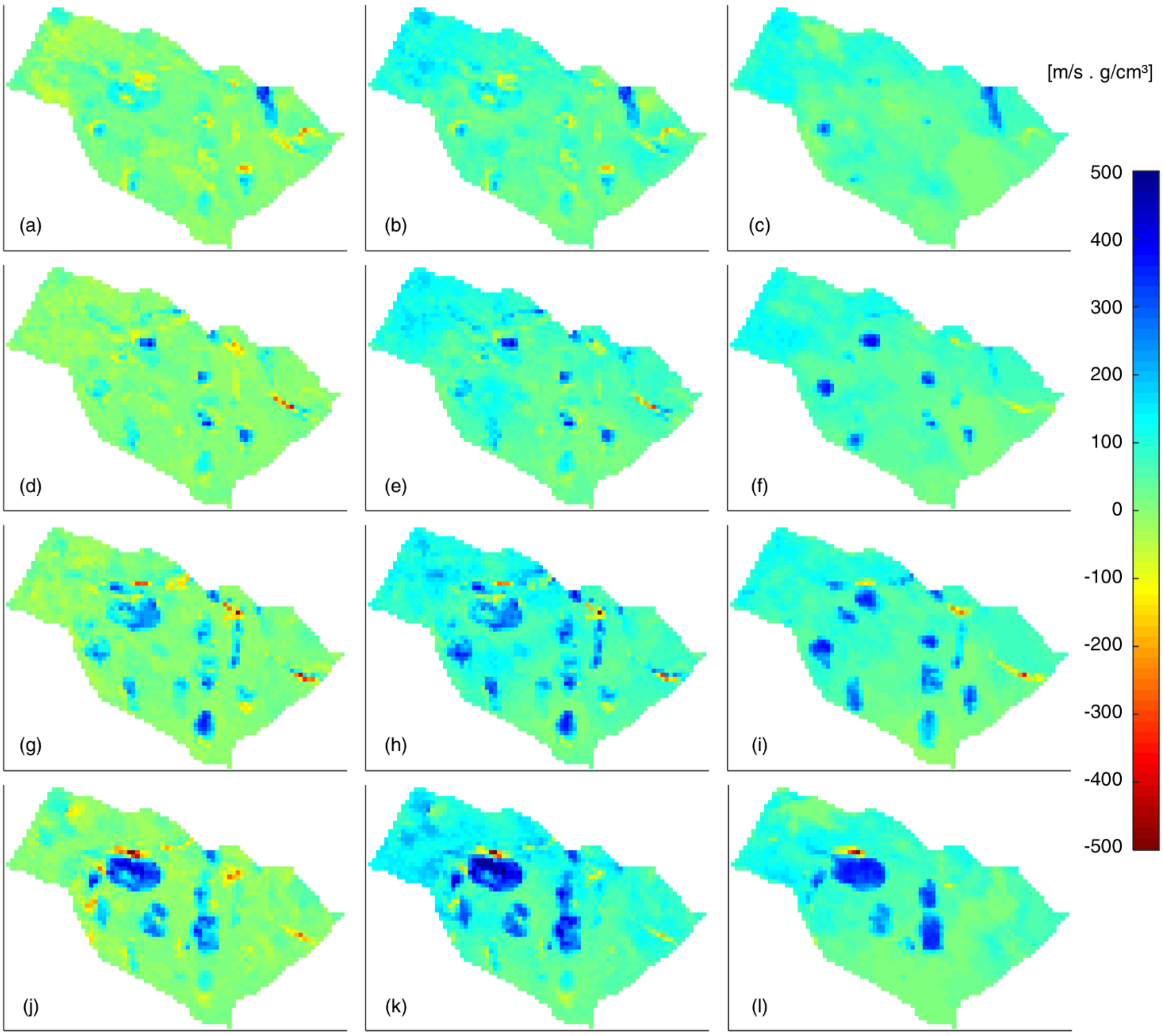


Figure 2.20: Field-scale case – analysis of the effect of *q* in the observed acoustic impedance variation. The four horizons of the INV data are in (a), (d), (g), and (j). The four horizons of the INV data "modified" by the mean of *q* after the calibration are in (b), (e), (h), and (k). The quasi-ideal PEM data are in (c), (f), (i), and (l).

2.7.2.4 Update of the covariance matrix of observation errors

After the addition of the error-related parameter to compensate for the data characteristics that the models cannot explain, it is possible to follow an iterative approach to update the covariance matrix of total observation errors, as proposed by Oliver and Alfonzo (2018a). We included this process in the workflow of Figure 2.1, and its objective is to compensate for other correlated errors that the parameter ***q*** could not absorb, due to the limited degrees of freedom.

In this work, we update the covariance matrix of observation errors, $\boldsymbol{C}_{dd}$, using the residuals of the data assimilation with the added $\boldsymbol{q}$ parameter. We estimate the variance of $\boldsymbol{C}_D^{k+1}$ from the residual magnitude, and we use an experimental variogram of these residuals to generate the updated covariance matrix. In our tests, we started the data assimilation with an uncorrelated covariance matrix with a constant standard deviation of $30\frac{m}{s} \times \frac{g}{cm^3}$. This value balances TLS and well data objective functions, which form the second term on the right side of equation (2.2). After two iterations of the covariance update, we obtained a standard deviation of roughly $49\ \frac{m}{s} \times \frac{g}{cm^3}$ and an exponential variogram model with a four-cell correlation length, roughly $400\ m$. Nevertheless, this update did not significantly impact the calibrated models' behavior, as one can verify in the dashed curves of figures 2.17 and 2.18. We noticed a minor increase in the oil rate metric in Figure 2.17 and a slight improvement in the water rate metric in Figure 2.18.

We can mention two reasons for the small influence of the covariance update in our example. The initial variance of observations that we chose to balance the objective functions was relatively close to the updated value. Furthermore, the addition of the $\boldsymbol{q}$ parameter seemed to compensate for the correlated errors due to model limitation.

2.7.2.5 Application in a quasi-ideal situation

Aside from the field application using the inverted data, it is essential to check if our methodology would jeopardize the data assimilation in a quasi-ideal case, where it would not be fundamental to compensate for model-related errors. Here, we perform this test using the data directly from the reference petroelastic model (PEM maps) without seismic modeling and inversion (see Figure 2.5 (a)). It is worth mentioning that this data assimilation is not ideal because we still have scale differences and the reference model has some features that are not included in the simulation models, which have a coarser grid. Therefore, we refer to this case as a quasi-ideal situation.

Following the workflow of Figure 2.1, we performed regular data assimilation without including the $\boldsymbol{q}$ parameter. Thereafter, we computed the residuals from the data assimilation and filtered it, using an eight-cell moving average, to estimate the prior distribution of $\boldsymbol{q}$. Figure 2.21 depicts the original and filtered residual maps. Comparing them with the maps of Figure 2.19, we note that the mean residuals are closer to zero in this quasi-ideal application. This difference corroborates the idea that the missing physics originated most of the previous correlated errors using the INV data.

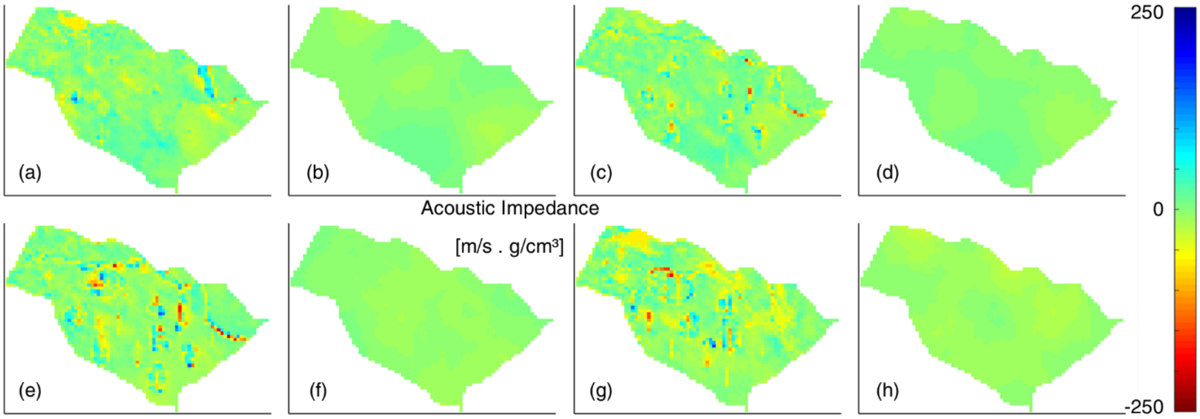


Figure 2.21: Field-scale quasi-ideal case – mean acoustic impedance residuals of the initial data assimilation and filtered mean residuals using an eight-cell window moving average, we show horizon 1 (a and b), 2 (c and d), 3 (e and f), and 4 (g and h), respectively.

After estimating the prior distribution of $\boldsymbol{q}$, we performed new data assimilation, adding this parameter to the petroelastic model's output, as shown in Figure 2.2. One would expect that the changes from this error-related parameter's inclusion would not significantly impact the results. First, we checked the VOIP distribution in Figure 2.22. Note that the inclusion of $\boldsymbol{q}$ caused a minor shift in the volumes, which are still close to the reference value. Although not significant, this change is consistent with the idea that the weak-constraint formulation tends to weaken the seismic data's influence. Furthermore, $\boldsymbol{q}$ may also be compensating for scale-related differences because the PEM data runs through an upscaling process before comparing it with the simulated responses.

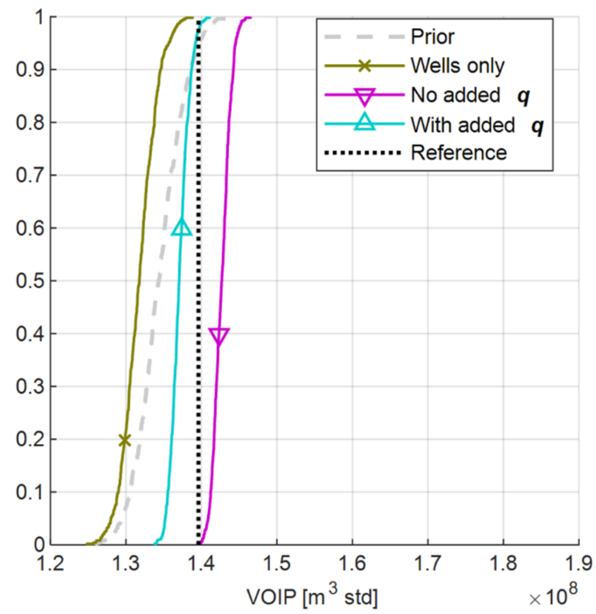


Figure 2.22: Field-scale quasi-ideal case – Volume of Oil in Place (VOIP) cumulative distribution curves.

We compared the calibrated model's production forecasts with and without the $\boldsymbol{q}$ parameter in the data assimilation. Looking at the oil production forecasts comparisons in Figure 2.23, we notice that the error parameter's inclusion caused a slight improvement of the oil rate deviations compared to the reference on a well-by-well basis. We observed the same tendency in the water rate forecast deviations in Figure 2.24. These results indicate that the addition of $\boldsymbol{q}$ compensated the minor effects related to the scale differences. Nevertheless, its influence is marginal when there is no significant correlated error source.

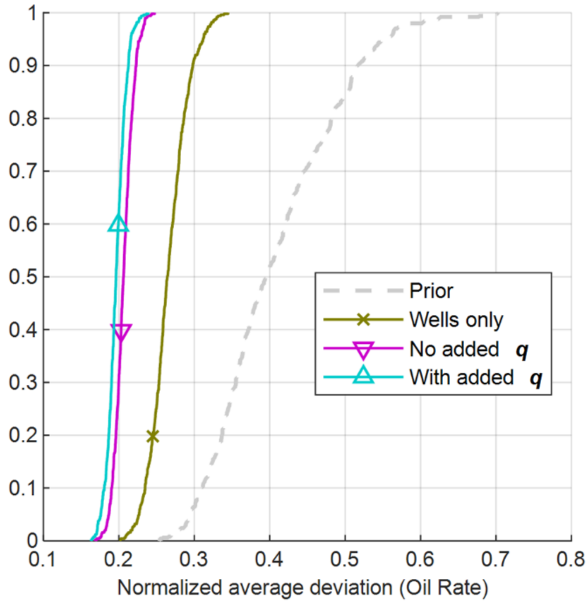


Figure 2.23: Field-scale quasi-ideal case: normalized average deviation in the oil rate forecast (well-by-well comparison).

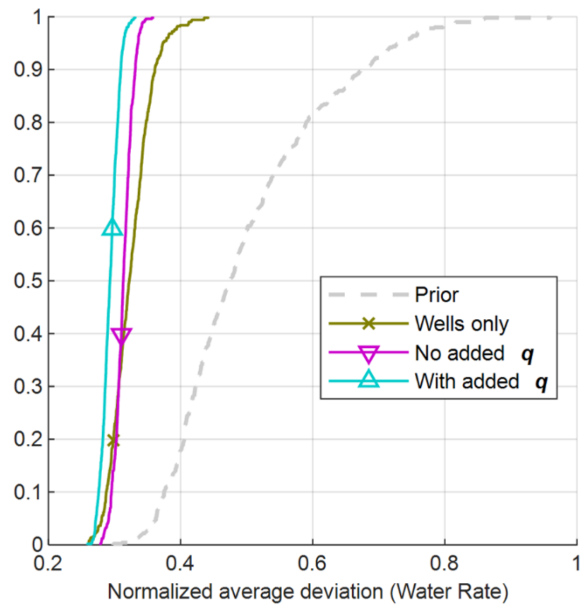


Figure 2.24: Field-scale quasi-ideal case: normalized average deviation in the water rate forecast (well-by-well comparison).

It is interesting to note that, after applying our methodology, the realistic INV data led to similar model calibration as the quasi-ideal PEM data. We can demonstrate that by comparing the production forecast metrics in figures 2.17 and 2.23, for the oil rates, and figures 2.18 and 2.24 for the water rates. These similar forecasts indicate that, apart from the distortions compensated by $\boldsymbol{q}$, the seismic modeling and inversion processes did not cause major information loss that would impair the data assimilation in this example.

2.8 Summary and conclusions

In this work, we proposed a practical methodology to account for model-related errors in time-lapse seismic data assimilation. We augmented the parameter vector to include an additive error parameter in the calibration process. Therefore, we adopted a weak-constraint formulation, which mitigates the effects of model error and bias, avoiding unphysical updates of the reservoir parameters. We tested the methodology in two synthetic cases, a simple 2D case and a realistic field-scale 3D case, which presents the typical challenges of a real reservoir. In short, the main contribution is a new methodology to compensate for model error and bias in time-lapse seismic data assimilation that does not require a high-fidelity low-error model.

The specific conclusions of this work are:

- Model errors with long-range correlations can jeopardize the data assimilation results, causing unphysical parameter changes.

- We managed to improve the quality of TLS data assimilation in the presence of this type of error using a weak-constraint formulation.
- After applying our methodology, the TLS data improved the production forecasts, even under the influence of model-related errors. This forecast improvement had nearly the same quality as the quasi-ideal case, where the model errors are minor.
- Our methodology can also handle model bias as a specific type of model error.
- It is not straightforward to optimize the definition of the prior distribution of the error-related parameter in the weak-constraint formulation. Nevertheless, it is essential to consider it in data assimilation when there are significant correlated model-related errors. Our results show that the impact of neglecting this type of error is greater than the effect of applying our methodology to a case where it is insignificant.
- We provided a practical way to estimate the prior distribution of the error parameter using the residuals from a calibrated ensemble that is not optimal but improved the model updates in our tests.
- As with any weak-constraint formulation, the proposed methodology reduces the magnitude of the reservoir parameter's updates due to the data. In our approach, this reduction intensifies as the model limitations increases.
- Comparing the observed data with the additive error parameter enables a multidisciplinary analysis of the data and the models. In this analysis, the additive error parameter relates to part of the data that the models could not represent. Therefore, it may be possible to improve the geological models or to reinterpret the time-lapse seismic data.

2.9 Nomenclature

Variables

$\mathbf{1}$	Column vector whose elements are 1
a	Semivariogram Gaussian model parameter
a_p	Impedance variation pressure sensitivity
a_s	Impedance variation saturation sensitivity
$\boldsymbol{A}$	Matrix of ensemble anomalies
$\boldsymbol{C}_{dd}$	Covariance matrix of measurement noise
$\boldsymbol{C}_{D}$	Covariance matrix of total observation errors
$\boldsymbol{C}_{xx}$	Covariance matrix of the parameters
$\boldsymbol{C}_{xy}$	Covariance matrix between the parameters and the simulated response
$\boldsymbol{C}_{yy}$	Covariance matrix of the simulated response

$\boldsymbol{C}_{zy}$	Covariance matrix between the augmented parameters and the simulated response
$\boldsymbol{d}$	Data vector
$\boldsymbol{D}$	Data matrix
$\boldsymbol{g}$	Forward model
h	Distance
$\boldsymbol{I}$	Identity matrix
$\mathcal{J}$	Cost function
$\boldsymbol{K}$	Kalman gain matrix
L	Semivariogram length
m	Number of data points (measurements)
n	Number of parameters
N	Ensemble size
N_{mda}	Number of data assimilations in an ESMDA run
$NQDS$	Normalized quadratic norm with sign
p	Pressure
$\boldsymbol{p}$	Residuals between observed and simulated response
$\overline{\boldsymbol{p}}$	Ensemble mean of the residuals
$\overline{\overline{\boldsymbol{p}}}$	Filtered mean residuals
$\boldsymbol{q}$	Additive model-error parameter vector
$\overline{\boldsymbol{q}}$	Ensemble mean of the additive model-error parameter vector
$\boldsymbol{Q}$	Additive model-error parameter matrix
S	Fluid saturation
$\boldsymbol{x}$	Parameter vector
$\boldsymbol{x}^{true}$	"True" parameter vector
$\boldsymbol{X}$	Parameter matrix
$\boldsymbol{Y}$	Matrix of predicted ensemble anomalies
$\boldsymbol{Z}$	Augmented parameter matrix
α	Inflation factor
δ	Average absolute deviation
δ_{ideal}	Average absolute deviation of the quasi-ideal test
$\boldsymbol{\varepsilon_d}$	Measurement noise
$\boldsymbol{\varepsilon_g}$	Model error
ΔI_{PNORM}	Normalized acoustic impedance variation
$\Delta\delta$	Deviation increase
ξ	Semivariogram
$\boldsymbol{\Pi}$	Operation that removes the mean and scales the matrix
$\boldsymbol{\rho}$	Localization matrix
σ	Standard deviation
τ	Observed data error tolerance
φ	Observed data error constant

Subscripts

i	Initial value
j	Element in the ensemble
l	Data type
N	Size equal to the ensemble size

TLS	Time-lapse seismic data
w	water
wi	Initial (connate) water

Superscripts

a	Updated estimate
f	Prior or background estimate
i	ESMDA iteration
(k)	Iteration of the covariance matrix update methodology
obs	Observed, related to the measurements
sim	Simulated
T	Matrix transpose

2.10 Acknowledgments

We conducted this work with the support of Libra Consortium (Petrobras, Shell Brasil, Total, CNOOC, CNPC, and PPSA) within the ANP R&D levy as "commitment to research and development investments" and Energi Simulation. The authors are grateful for the support of the Center of Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil), and Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil). Also, a special thanks to CMG and Schlumberger for software licenses.

3 SUBSPACE ENSEMBLE RANDOMIZED MAXIMUM LIKELIHOOD WITH LOCAL ANALYSIS FOR TIME-LAPSE SEISMIC DATA ASSIMILATION

Authors:

Gilson Moura Silva Neto

Ricardo Vasconcellos Soares

Geir Evensen

Alessandra Davolio

Denis José Schiozer

3.1 Abstract

Time-lapse seismic data assimilation has been drawing the reservoir engineering community's attention over the past few years. One of the advantages of including this kind of data to improve the reservoir flow models is that it provides complementary information compared to the wells' production data. Ensemble-based methods are some of the standard tools used to calibrate reservoir models using time-lapse seismic data. One of the drawbacks of assimilating time-lapse seismic data involves the large data sets, mainly for big reservoir models. This situation leads to high-dimensional problems that demand significant computational resources to process and store the matrices when using conventional and straightforward methods. Another known issue associated with the ensemble-based methods is the limited ensemble sizes, which cause spurious correlations between the data and the parameters and limit the degrees of freedom. In this work, we propose a data assimilation scheme using an efficient implementation of the Subspace Ensemble Randomized Maximum Likelihood method with local analysis. This method reduces the computational requirements for assimilating big data sets because the number of operations scales linearly with the number of observed data points. Furthermore, by implementing it with local analysis, we reduce the memory requirements at each update step and mitigate the effects of the limited ensemble sizes. We test two local analysis approaches, one distance-based, and another correlation-based. We apply these implementations to two synthetic time-lapse seismic data assimilation cases, one

2D example, and one field-scale application that mimics some of the real field challenges. We compare the results to reference solutions and to the known Ensemble Smoother with Multiple Data Assimilations using Kalman gain distance-based localization. The results show that our method can efficiently assimilate time-lapse seismic data, leading to updated models that are comparable to other straightforward methods. The correlation-based local analysis approach provided results similar to the distance-based approach, with the advantage that the former can be applied to data and parameters that do not have specific spatial positions.

Abbreviations:

EnRML – Ensemble Randomized Maximum Likelihood

ESMDA – Ensemble Smoother with Multiple data Assimilations

IES – Iterative Ensemble Smoother

PEM – Petro-elastic Model

SEnRML – Subspace Ensemble Randomized Maximum Likelihood

Keywords:

EnRML, Iterative Ensemble Smoother, Seismic History Matching, Local analysis, Adaptive localization

3.2 Introduction

Time-lapse seismic data provide dynamic information about the spatial distribution of the reservoir properties. One may consider it complementary to the well production data because the latter is sparsely spaced but more frequent in time than the former (STEPHEN, Karl D.; MACBETH, 2008). Therefore, there are efforts in place to use this type of data quantitatively to calibrate reservoir models (EMERICK, Alexandre A., 2016; LORENTZEN *et al.*, 2019; STEPHEN, Karl D. *et al.*, 2006; TOLSTUKHIN; LYNGNES; SUDAN, 2012; ULLMANN DE BRITO; CALETTI; MORAES, 2011). Nevertheless, the seismic data contributes to augmenting the data assimilation process's complexity, particularly by significantly increasing the size of the data set. This aspect becomes especially challenging for giant oil field applications due to the number of data points and parameters. Therefore, it is crucial to develop efficient methods to assimilate the available data in this context.

Gu and Oliver (2007) proposed an iterative method for sequential data assimilation called Ensemble Randomized Maximum Likelihood (EnRML) Filter. Later, Chen and Oliver (2012) presented EnRML as an iterative ensemble smoother (IES), which circumvented the need to restart reservoir simulation runs. Chen and Oliver (2013) improved the method by using the idea of the Levenberg-Marquardt method and including a step length parameter. Later, Raanes et al. (2019) revised EnRML as an iterative ensemble smoother, simplifying the method both conceptually and computationally, and Evensen et al. (2019) proposed an efficient implementation of the EnRML IES. After the modifications, the authors referred to the method as the Subspace EnRML (SEnRML). The computational cost of the SEnRML algorithm scales only linearly with the number of parameters and the data, making it suitable for the assimilation of big data sets into big reservoir models. Nevertheless, the authors only tested the global update implementation of the method, where they used the entire data set to update all model variables into a single step.

The application of the subspace inversion with a low-rank representation of the covariance matrix of observation error, as in SEnRML, reduces the observed data size's impact on the data assimilation. Evensen (2004) explained this idea and Skjervheim et al. (2007) presented this approach for time-lapse data assimilation. The projection of the covariance matrix of the observation errors into the subspace generated by the ensemble anomalies enables limiting the matrix's dimension to less than the ensemble size. Furthermore, the projection of the data into this subspace reduces its size to the same dimension.

Many researchers addressed the challenge of assimilating time-lapse seismic big data sets by employing different strategies. Liu and Grana (2020) reduced the data dimension by applying a deep representation learning method called deep convolutional autoencoder. Yin et al. (2019) proposed a method to reduce the number of data points using the correlation between time-lapse seismic data and the wells' cumulative volume data when multiple monitors are available. Both studies employed Ensemble Smoother with Multiple Data Assimilations (EMERICK, Alexandre A.; REYNOLDS, 2013a) as the data assimilation method. Soares et al. (2019) significantly reduced the number of assimilated data points using a dictionary learning technique. Luo et al. (2017) proposed a history-matching framework in which they perform data reduction using a wavelet-based method. Later, Luo et al. (2018) and Lorentzen et al. (2019) applied this idea to the Brugge field benchmark and to the Norne field, respectively. Other authors achieved data reduction by representing the time-lapse seismic data in terms of the position and distances of fluid fronts (ABADPOUR; BERGEY; PIASECKI, 2013;

LEEUWENBURGH; ARTS, 2014; TRANI; ARTS; LEEUWENBURGH, 2013; ZHANG, Yanhui; LEEUWENBURGH, 2017).

In a practical application of ensemble-based methods in reservoir history matching, the reservoir flow simulation's computational costs limit the ensemble size. Usually, the number of models in the ensemble ranges from 100 to 500 models, while the number of parameters and observed data points may reach the magnitude of 10^5 to 10^7. Consequently, we face two significant issues related to small ensemble sizes (EMERICK, Alexandre; REYNOLDS, 2011). The covariance matrices' ensemble estimates tend to produce a non-zero correlation between the data and the parameters, even for physically uncorrelated regions. Moreover, the degrees of freedom available to update the parameters is also limited. Hence, two conventional approaches to circumvent these issues involve applying localization or local analysis.

Kalman Gain localization (ZHANG, Yanfen; OLIVER, 2011) has become a popular procedure in reservoir history matching with ensemble-based methods (EMERICK, Alexandre A.; REYNOLDS, 2013b; LACERDA; EMERICK; PIRES, 2019; RANAZZI; SAMPAIO, 2019; SOARES; MASCHIO; SCHIOZER, 2018). Nevertheless, its application in a global update approach, in which one uses all parameters and data points at a single update step, requires constructing the large dimension Kalman Gain matrix. This implementation may lead to slower convergence when compared to local analysis (CHEN, Yan; OLIVER, 2017). In the efficient implementation of SEnRML (EVENSEN, Geir *et al.*, 2019), one does not form the Kalman Gain matrix, which would directly relate the parameter changes and the innovations. An alternative approach to circumvent the issues related to the limited ensemble size is to perform local analysis. Although less addressed than the Kalman Gain localization in reservoir history-matching problems, some studies reported the use of local analysis (FAHIMUDDIN; AANONSEN; SKJERVHEIM, 2010a; SKJERVHEIM, Jan-arild; EVENSEN, 2011; ZHAO; REYNOLDS; LI, 2008).

Sakov and Bertino (2011) compared the covariance localization and the local analysis for the Ensemble Kalman Filter (EnKF, EVENSEN, Geir, 1994). They showed that both methods could produce comparable results. Chen and Oliver (2017) reported similar findings when comparing local analysis schemes to Kalman gain localization for reservoir history-matching problems. All the methods can provide equally good history-matching results if the localization lengths are adequately tuned. However, finding these lengths in a practical problem is not a trivial task. For instance, Emerick and Reynolds (2011) suggested defining them as the prior ensemble correlation lengths, summed to the sensitivity range of the data.

One alternative to the data selection and tapering of the influences based on the distance from the parameters is to compute the localization or local analysis from the correlation between them. This approach does not require tuning the localization lengths and it applies to nonlocal data and parameters. For instance, Bishop and Hodyss (2007) proposed attenuating the spurious correlations in ensemble-based data assimilation by raising them to a power. Evensen (2009), in chapter 15, provides a brief review of pioneer studies addressing adaptive localization, including Bishop and Hodyss' (2007) work. Later, Luo et al. (2018) proposed an automatic method to compute a threshold and select the data for the analysis based on the absolute value of the estimated correlation. Luo et al. (2019) applied this method to assimilate the real production data from the Norne Field. Also, Luo and Bhakta (2020) improved the previous method's computational efficiency and proposed a technique to compute smooth tapering coefficients.

In this work, we propose two local analysis schemes for the SEnRML method. The first is a distance-based approach that selects and weights the data that influences each parameter from the physical distance between them. The second is an automatic method that uses the correlation between the data and the parameters to compute the influences. We test the method with two synthetic examples, one 2D case, and one field-scale example. In both tests, we compare SEnRML with local analysis to the widely known Ensemble Smoother With Multiple Data Assimilations (ESMDA) with distance-based Kalman gain localization (EMERICK, Alexandre A.; REYNOLDS, 2013a).

The specific objectives of the present work are:

- Propose the local analysis schemes for SEnRML, both distance-based and automatic.
- Test the subspace efficient implementation of the method in time-lapse data assimilation using a realistic case.
- Compare the distance-based and the automatic schemes for data selection and tapering in a local analysis scheme for time-lapse seismic data assimilation.

3.3 Subspace EnRML

The Subspace EnRML (SEnRML) formulation applied in this work follows the revision made by Raanes et al. (2019) and the efficient implementation for big data sets and big models described by Evensen et al. (2019). Here, we assume there is a forward model ($\boldsymbol{g}$) that provides the simulated data $\boldsymbol{d}^{sim} \in \Re^{m\times 1}$, given the parameters $\boldsymbol{x} \in \Re^{n\times 1}$

$$\boldsymbol{d}^{sim} = \boldsymbol{g}(\boldsymbol{x}). \tag{3.1}$$

Also, this model can determine these data precisely if we use the “true” value parameters, $\boldsymbol{x}^{true}$. We also have the measurements $\boldsymbol{d}^{obs} \in \Re^{m\times 1}$ regarding the “true” output of the forward model and the measurement errors $\boldsymbol{\varepsilon} \in \Re^{m\times 1}$, which are Gaussian and have zero mean, defined as

$$\boldsymbol{d}^{obs} = \boldsymbol{g}(\boldsymbol{x}^{true}) + \boldsymbol{\varepsilon}. \tag{3.2}$$

In this context, the data assimilation problem consists of calibrating the model’s parameters vector using prior knowledge and the observed data. We may represent our prior knowledge regarding the models’ parameters with an ensemble of N realizations, $\boldsymbol{X}^f = (\boldsymbol{x}_1^f, \boldsymbol{x}_2^f, \dots, \boldsymbol{x}_N^f)$.

Evensen (2003) expressed the updates of the models’ parameters as a linear combination of the ensemble perturbations, which also occurs in the case of a change in variables proposed by Hunt et al. (2007). Raanes et al. (2019) and Evensen et al. (2019) applied an equivalent change of variables that simplified the EnRML method, both conceptually and computationally. Using this definition, the updates of the parameters are a linear combination of the ensemble anomalies ($\boldsymbol{A}$),

$$\boldsymbol{X}^a = \boldsymbol{X}^f + \boldsymbol{A}\boldsymbol{W} = \boldsymbol{X}^f\left(\boldsymbol{I}_N + \frac{1}{\sqrt{N-1}}\boldsymbol{W}\right), \tag{3.3}$$

where the matrix $\boldsymbol{W} \in \Re^{N\times N}$ defines the linear combination of $\boldsymbol{A}$ that represents the models’ updates. One can demonstrate the right side of equation 3.3, by using the $\boldsymbol{W}$ matrix property, that the sum of the elements in each column is equal to zero (EVENSEN, Geir *et al.*, 2019). Therefore, by applying this change of variables, the data assimilation problem changes to searching the matrix $\boldsymbol{W}$, which has a lower dimension than searching $\boldsymbol{X}^a \in \Re^{n\times N}$ directly. We obtain $\boldsymbol{A}$ by normalizing and subtracting the mean from the matrix with the prior parameters,

$$\boldsymbol{A} = \boldsymbol{X}^f \frac{1}{\sqrt{N-1}}\left(\boldsymbol{I}_N - \frac{1}{N}\boldsymbol{1}_N\boldsymbol{1}_N^T\right) = \boldsymbol{X}^f\boldsymbol{\Pi}_N, \tag{3.4}$$

where $\boldsymbol{1}_N \in \Re^{N\times 1}$ is a vector whose elements are equal to 1, $\boldsymbol{I}_N \in \Re^{N\times N}$ is the identity matrix, and $\boldsymbol{\Pi}_N$ is a projector that transforms $\boldsymbol{X}^f$ into the ensemble anomalies.

Similarly to the parameters, we represent the observed data by an ensemble of perturbed measurements $\boldsymbol{D} \in \Re^{m\times N}$, formed by sampling the distribution $\boldsymbol{d}_j \sim \mathcal{N}(\boldsymbol{d}^{obs}, \boldsymbol{C}_{dd})$, where $\boldsymbol{C}_{dd} \in \Re^{m\times m}$ is a predefined covariance matrix of the measurement errors.

Finally, the problem consists of finding the matrix $\boldsymbol{W}$ that minimizes the cost function,

$$\mathcal{J}(\boldsymbol{w}_j) = \frac{1}{2}\boldsymbol{w}_j^T\boldsymbol{w}_j + \frac{1}{2}\left[\boldsymbol{g}(\boldsymbol{x}^f + \boldsymbol{A}\boldsymbol{w}_j) - \boldsymbol{d}_j\right]^T \boldsymbol{C}_{dd}^{-1}\left[\boldsymbol{g}(\boldsymbol{x}^f + \boldsymbol{A}\boldsymbol{w}_j) - \boldsymbol{d}_j\right], \tag{3.5}$$

where the first term on the right relates to the parameters' updates compared to the prior estimates, whereas the second term pertains to the data mismatch. It is possible to obtain the set of vectors $\boldsymbol{w}_j$ that minimizes equation 3.5 by employing the Gauss-Newton scheme. The iterative procedure for updating $\boldsymbol{W}$ (EVENSEN, Geir *et al.*, 2019) is

$$\boldsymbol{W}^{i+1} = \boldsymbol{W}^i - \gamma\left[\boldsymbol{W}^i - \left(\boldsymbol{S}^i\right)^T\left(\boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^T + \boldsymbol{C}_{dd}\right)^{-1}\boldsymbol{H}^i\right], \tag{3.6}$$

where γ is the step-length parameter, $\boldsymbol{S}^i$ is called the matrix of predicted and deconditioned ensemble anomalies

$$\boldsymbol{S}^i = \boldsymbol{Y}^i\left(\boldsymbol{I}_N + \boldsymbol{W}^i\boldsymbol{\Pi}_N\right)^{-1}, \tag{3.7}$$

and $\boldsymbol{Y}^i$ is the matrix of predicted ensemble anomalies,

$$\boldsymbol{Y}^i = \left(\boldsymbol{D}^{sim}\right)^i\boldsymbol{\Pi}_N. \tag{3.8}$$

Note that $\left(\boldsymbol{D}^{sim}\right)^i = \boldsymbol{g}\left(\boldsymbol{X}^i\right)$ consists of the simulated data obtained after running the reservoir simulator with the vectors of parameters in the updated $\boldsymbol{X}$ matrix at iteration i, $\boldsymbol{X}^i$. $\boldsymbol{H}^i$ represents the innovations

$$\boldsymbol{H}^i = \boldsymbol{S}^i\boldsymbol{W}^i + \boldsymbol{D} - \left(\boldsymbol{D}^{sim}\right)^i. \tag{3.9}$$

Before implementing the method, it is necessary to define a procedure to compute the inverse of $\left(\boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^T + \boldsymbol{C}_{dd}\right)$. Evensen et al. (2019) obtained this inverse by first computing the truncated and economic singular value decomposition (TSVD) of $\boldsymbol{S}^i$,

$$\boldsymbol{S}^i = \boldsymbol{U}^i\boldsymbol{\Sigma}^i\left(\boldsymbol{V}^i\right)^T, \tag{3.10}$$

keeping a maximum of $N-1$ significant singular values. To avoid scaling issues, it is essential to rescale the $\boldsymbol{S}^i$ using a diagonal matrix whose elements are the inverse of the data's standard deviation before computing the TSVD. To keep the equation consistent, it is also necessary to apply the rescaling procedure to the innovations and the $\boldsymbol{C}_{dd}$ matrix. After, we use the approximation

$$\boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^T + \boldsymbol{C}_{dd} \approx \boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^T + \boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^+ \boldsymbol{E}\boldsymbol{E}^T \left(\boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^+\right)^T, \tag{3.11}$$

where the superscript $+$ denotes the Moore-Penrose pseudoinverse. In equation 3.11, we first approximated the matrix $\boldsymbol{C}_{dd} \approx \boldsymbol{E}\boldsymbol{E}^T$, and then projected this matrix onto the subspace defined by $\boldsymbol{S}$ (EVENSEN, Geir *et al.*, 2019). We obtain the vectors that form $\boldsymbol{E} \in \Re^{m \times N_E}$ by sampling the distribution $\mathcal{N}(0, \boldsymbol{C}_{dd})$ and dividing the results by $\sqrt{N_E - 1}$. We can use a number of samples to form $\boldsymbol{E}$ that is equal to the ensemble size, $N_E = N$, or we can increase this number to reduce sampling errors. Here, we used $N_E = 10N$, but we also tested a smaller size in the computational cost section, obtaining similar results. After substituting (3.10) in (3.11), we find

$$\boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^T + \boldsymbol{C}_{dd} \approx \boldsymbol{U}^i\boldsymbol{\Sigma}^i \left(\boldsymbol{I}_N + \left(\boldsymbol{\Sigma}^i\right)^+ \left(\boldsymbol{U}^i\right)^T \boldsymbol{E}\boldsymbol{E}^T \boldsymbol{U}^i \left(\left(\boldsymbol{\Sigma}^i\right)^+\right)^T\right) \left(\boldsymbol{\Sigma}^i\right)^T \left(\boldsymbol{U}^i\right)^T. \tag{3.12}$$

Hereafter we apply the eigenvalue decomposition

$$\boldsymbol{Z}^i\boldsymbol{\Lambda}^i\left(\boldsymbol{Z}^i\right)^T = \left(\boldsymbol{\Sigma}^i\right)^+ \left(\boldsymbol{U}^i\right)^T \boldsymbol{E}\boldsymbol{E}^T \boldsymbol{U}^i \left(\left(\boldsymbol{\Sigma}^i\right)^+\right)^T, \tag{3.13}$$

that we can compute efficiently from the SVD of $\left(\boldsymbol{\Sigma}^i\right)^+ \left(\boldsymbol{U}^i\right)^T \boldsymbol{E}$. Finally, we can obtain the inverse of $\left(\boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^T + \boldsymbol{C}_{dd}\right)$ from

$$\left(\boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^T + \boldsymbol{C}_{dd}\right)^{-1} \approx \left(\boldsymbol{U}^i\left(\boldsymbol{\Sigma}^i\right)^+ \boldsymbol{Z}^i\right) \left(\boldsymbol{I}_{N_\Lambda} + \boldsymbol{\Lambda}^i\right)^{-1} \left(\boldsymbol{U}^i\left(\boldsymbol{\Sigma}^i\right)^+ \boldsymbol{Z}^i\right)^T. \tag{3.14}$$

Both $\boldsymbol{I}_{N_\Lambda}$ and $\boldsymbol{\Lambda}$ are diagonal matrices whose dimension is $N_\Lambda \times N_\Lambda$, where $N_\Lambda \leq N - 1$. We show the sizes of the matrices in equations 3.6 and 3.14 in **Table 3.1**. From the dimensions, one can quickly note that these equations scale linearly with m.

Table 3.1: Dimensions of the variables in SEnRML.

Matrix	$\boldsymbol{W}^i$	$\boldsymbol{S}^i$	$\boldsymbol{U}^i$	$\boldsymbol{\Sigma}^i$	$\boldsymbol{Z}^i$	$\boldsymbol{\Lambda}^i$	$\boldsymbol{H}^i$
Dimension	$N \times N$	$m \times N$	$m \times N_\Lambda$	$N_\Lambda \times N_\Lambda$	$N_\Lambda \times N_\Lambda$	$N_\Lambda \times N_\Lambda$	$m \times N$

One should note that the application of equation 3.14 into equation 3.6 implicates the product of the innovations by $\left(\boldsymbol{U}^i\right)^T$, which reduces its dimension to $N_\Lambda \times N$. This result has a beneficial effect on decreasing the number of operations. Nevertheless, it may also cause information loss. Here, we try to mitigate the latter's impacts by reducing the data used for each analysis step through the local analysis scheme, addressed in the following section.

3.4 Local analysis schemes

We separate the local analysis procedure into three parts, the parameter segregation into local groups, the data selection and tapering of their influence, and the calibration of the parameters, see Figure 3.1. We propose two ways of selecting the data and computing the tapering vectors, one distance-based and one correlation-based. We address the latter as automatic because after defining the formulation, there is no need to tune any parameter, as one needs to tune the localization lengths for regions of the reservoir in the distance-based scheme.

We address each step depicted in Figure 3.1 in the next subsections.

Figure 3.1: Local analysis scheme general workflow.

3.4.1 Segregation of parameters

At this step, we need to segregate the parameters into local groups to perform the calibration in independent stages during the third step of Figure 3.1. Chen and Oliver (2017) state that the idea behind this segregation is to update all the related parameters with the same data using a single analysis equation. When one starts grouping variables contained by different cells, the definition of the groups of parameters becomes a tradeoff between the computational cost and precision of the data selection.

In 3D models, a common approach to segregation is to group all the parameters located in the grid's vertical columns of cells (CHEN, Yan; OLIVER, 2017). We acknowledge that, in some cases, there may be differences between the correlated data related to parameters located in different reservoir layers. Nevertheless, in this work, we decided to group sets of columns of cells for two reasons. First, the distance-based localization methods used for comparison purposes commonly consider the horizontal distance to define the data selection (CHEN, Yan; OLIVER, 2017; EMERICK, Alexandre; REYNOLDS, 2011; EMERICK, Alexandre A., 2016). Secondly, by grouping the parameters into cell columns, we reduce the number of analysis steps and the computational costs compared to segregating the cells vertically.

3.4.2 Distance-based data selection and influence tapering

In our implementation of the SEnRML method with local analysis, it is necessary to select the data and compute one tapering vector, $\boldsymbol{\rho}_k \in \Re^{m\times 1}$, for each group of parameters. Therefore, we defined the following procedure to compute them from the physical distance between the local groups and the data:

1. Select a local group of parameters to update.
2. Include all the data whose position is inside the group without tapering.
3. Select all data within a distance from the group of parameters.
4. Determine a tapering vector to reduce the impact of the data located outside the group.

We apply the widely known Gaspari and Cohn (1999) function to compute the tapering values to reduce the selected data's influence outside the group of parameters. One can determine the maximum distance of the selected data from a minimum tapering value. It is also necessary to define the critical lengths for the Gaspari and Cohn’s function, which is not straightforward for time-lapse seismic data assimilation (EMERICK, Alexandre A., 2016).

3.4.3 Automatic data selection and influence tapering

We also proposed an automatic correlation-based local analysis scheme in this work, in addition to the distance-based data selection. In this case, we select the data and compute the correlation-based tapering using an adaptation of the method by Luo and Bhakta (2020). We start by estimating the standard deviation of the spurious correlations, σ_ϵ, due to the limited ensemble size using the asymptotic approximation

$$\sigma_\epsilon \approx 1/\sqrt{N}. \tag{3.15}$$

After defining the standard deviation, the correlation threshold is

$$\theta = \sigma_\epsilon\sqrt{2\ln\left(n/n_p\right)}, \tag{3.16}$$

where n/n_p is the number of parameters of each type, e.g., porosity, directional permeabilities, and it is the same for all parameter types. It is important to note that the threshold varies with the inverse of the ensemble size's square root because a larger N reduces the sampling errors in the covariance estimation, its drawback being the higher computational costs. For a detailed explanation on the threshold estimation, refer to (LUO; BHAKTA, 2020; LUO; BHAKTA; NÆVDAL, 2018).

After obtaining the correlation thresholds, we can compute the tapering vectors for each local analysis group. We start by obtaining the correlation matrix between the prior realizations of parameters and the simulated data

$$\boldsymbol{R}_k = \mathrm{corr}\left(\boldsymbol{X}_k^f, \left(\boldsymbol{D}^{sim}\right)^f\right), \tag{3.17}$$

where $\boldsymbol{R}_k \in \Re^{\#\boldsymbol{X}_k \times m}$, and $\#\boldsymbol{X}_k$ are the number of parameters in the local analysis group. Luo and Bhakta (2020) proposed a tapering rule following the intuition that higher correlation coefficients should lead to higher tapering values. Based on this principle, they proposed an arbitrary pseudo-distance dummy variable

$$(z_k)_{l,c} = \left(1 - \left|(r_k)_{l,c}\right|\right)/(1-\theta), \tag{3.18}$$

where the elements $(z_k)_{l,c}$ are related to $(r_k)_{l,c}$, which are the elements of the matrix $\boldsymbol{R}_k$. Note that the index $l = 1,2,\dots,\#\boldsymbol{X}_k$ indexes the parameters in the group, and the index $c = 1,2,\dots,m$ refers to the data. The results based on equation 3.18 were satisfactory on Luo and Bhakta's (2020) applications using ensemble sizes of 100 and 103 models, and our 2D application with 100 models. Nevertheless, equation 3.18 has some undesired characteristics. Firstly, the pseudo-distance does not go to zero if θ tends to zero. One would expect this behavior because a large ensemble would not require data influence tapering. Secondly, it does not guarantee that the tapering goes to zero as the correlation goes to zero, or it is much lower than the threshold. Thirdly, even if the correlation is double the threshold, it does not necessarily result in tapering values close to 1. This tendency would be desirable because it means that the correlation is above the sampling error level, and the data is correlated to the parameter. Furthermore, in our tests with increasing ensemble sizes, which reduce the threshold, equation 3.18 led to somewhat homogeneous tapering coefficients. This influence tapering format caused the data assimilation to underestimate the final ensemble variance.

Therefore, we changed this equation to

$$(z_k)_{l,c} = \max\left(1.67 - 0.67\left|(r_k)_{l,c}\right|/\theta\,, 0\right), \tag{3.19}$$

whose results are roughly equal to equation 3.18 when $\theta \approx 0.4$, which is the condition that the original tapering rule led to favorable results. Note that equation 3.19 is also arbitrary, and we do not claim that this formulation is optimal in any sense. However, it has some desired features. First, it causes tapering close to zero if the correlation is much lower than the threshold and tapering equals zero if θ tends to zero. Secondly, it causes the tapering to go to one if the correlation is at least 2.5 times higher than the threshold. This value came from the match of equation 3.18 with $\theta \approx 0.4$. Finally, these characteristics do not change if we change the number of models in the ensemble. Consequently, we can apply it to different ensemble sizes for the same case, resulting in similar responses, differently from the original pseudo-distance dummy variable defined in equation 3.18. We provide details later in the Results section.

In our implementation, we need one vector $\bar{\mathbf{z}}_k \in \Re^{1\times m}$ to represent the tapering for each local analysis group. We obtain it using a statistic of the distribution of $(z_k)_{l,c}$, for $l = 1, 2, \dots, \#\mathbf{X}_k$. The proper statistic depends on the definition of local analysis groups, and we do not provide an optimal choice for a general case. Here, we used a percentile of the $(z_k)_{l,c}$ distribution among the parameters within the local groups. We tuned this percentile to cause a clear separation between the data closely related to the group, tapering close to 1, and the data that should not be related to it, tapering close to 0. Looking at the extremes, the option of taking the maximum values of $(z_k)_{l,c}$ means that only the data correlated to all parameters in the group will significantly influence the calibration. On the other hand, taking the minimum values of $(z_k)_{l,c}$ means that any data that relates to any parameters in each group will be relevant. Although this definition relies on knowledge about the case, it is possible to define this rule before starting the data assimilation process. Furthermore, if it is possible to work with local groups comprising a few parameters, or only parameters related to the same subset of the data, the definition of the vectors $\bar{\mathbf{z}}_k$ should be trivial. We provide more details about the selection of $\bar{\mathbf{z}}_k$ in the results section of the field-scale case.

After obtaining the pseudo-distances, we apply the Gaspari and Cohn (1999) function to compute the tapering vectors for each group, $\boldsymbol{\rho}_k \in \Re^{m\times 1}$, using the elements in $\bar{\mathbf{z}}_k^T$ as the argument of the function. Finally, we select all data whose tapering values are above a predefined minimum value, which is usually a small positive number, e.g., 10^{-2}. In this approach, the correlations estimated from the prior ensemble define the data that will influence each parameter group.

3.4.4 Calibration of parameters

Apart from using a different subset of the data points to calibrate each local group, we apply the tapering vector to change the data weights, based on the distance or the correlation. We opted to apply the tapering of ensemble anomalies and innovations, similarly to Chen and Oliver's (2017) tapering of observation. To do that, we change the $\mathbf{S}_k^i$ matrix to

$$\widehat{\mathbf{S}}_k^i = \left(\boldsymbol{\rho}_k^{1/2}\mathbf{1}_N^T\right) \circ \mathbf{S}_k^i, \tag{3.20}$$

where $\boldsymbol{S}_k^i$ comes from equation 3.7 and the index k relates to the local group. Furthermore, we change the innovations to

$$\widehat{\boldsymbol{H}}_k^i = \widehat{\boldsymbol{S}}_k^i \boldsymbol{W}_k^i + \left(\boldsymbol{\rho}_k^{1/2} \mathbf{1}_N^T\right) \circ \left[\boldsymbol{D}_k - \left(\boldsymbol{D}^{sim}\right)_k^i\right]. \tag{3.21}$$

Sakov and Bertino (2011) showed that tapering of ensemble anomalies and innovations with the square root of the tapering elements would have an equivalent effect as changing the observations' variance with the tapering elements' inverse if the observation errors were not correlated. Nevertheless, this condition does not hold in our tests. We opted to apply the tapering to the anomalies and innovations for two reasons. Firstly, it may be considered a more general approach than changing the observations' variance if the errors are correlated, as Sakov and Bertino (2011) commented. Secondly, we avoided forming the $\boldsymbol{C}_{dd}$ matrix to apply the tapering.

The SEnRML implementation with the local analysis scheme described in this section simplifies the computation of each step, in the sense that it uses only a subset of the data and parameters. Nevertheless, it is necessary to repeat the operations for each local group. Therefore, the procedure reduces the memory consumption during the computations, but it tends to increase the number of operations. The increase is more dramatic if one segregates the parameters into more local groups. Nevertheless, the local analysis steps are independent, and one can readily parallelize them.

3.5 Applications

We tested SEnRML with local analysis in two time-lapse seismic data assimilation examples. Firstly, we use a simple 2D case that allows us to have a large ensemble solution. Hence, it does not require localization and may serve as a reference solution to other small ensemble tests. Secondly, we apply SEnRML in a field-scale example that mimics some of the real field challenges. We describe each example in the following two subsections.

3.5.1 2D case

The 2D example consists of a 40×40 grid with four producer wells and four injectors. We built a synthetic truth from which we generate the observed data. **Figure 3.2a** depicts the reference model's porosity field and the well location, while Figure 3.2b illustrates the relationship between the porosity and permeability. In addition, we present the general characteristic of the example in **Table 3.2**. This low-fidelity model does not include a real reservoir's general complexities, but it mimics a time-lapse seismic data assimilation problem with fast forward simulation. We list the following simplifications of our 2D model: the relationship between the porosity and permeability is arbitrary and close to log-linear; there are no residual fluid saturations; we maintain the material balance through water injection; our data assimilation time is arbitrary; our data is normalized.

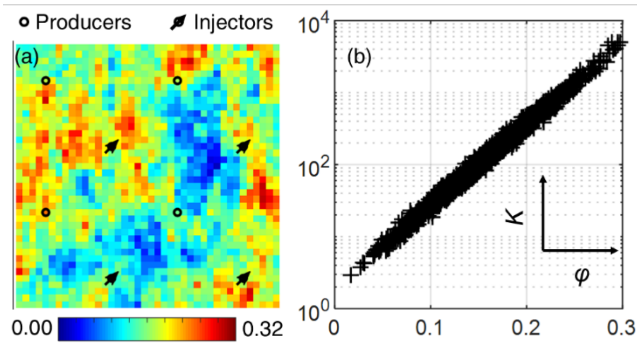


Figure 3.2: 2D example – synthetic truth. The porosity field is in (a), and the permeability-porosity relationship is in (b).

Table 3.2: Characteristics of the 2D case.

Characteristic	**Value**
Grid size	$40 \times 40 \times 1$ cells
Cell size	$50 \times 50 \times 50\ m$
Average porosity	0.15
Average permeability	$132\ mD$
Fluid mobility ratio (water/oil)	5
Fluid densities	Water: $1000\ kg/m^3$ Oil: $700\ kg/m^3$
Initial pressure	$200\ bar$
Number of active TLS data points	1600
Reservoir simulator	MATLAB Reservoir Simulation Toolbox (LIE, 2019)

In this 2D case, the waterflooding process causes changes in the pressure and saturation in 2 years. We generated a normalized impedance variation map by applying a simple petroelastic proxy model

$$\Delta I_{PNORM} = a_s \times (S_w - S_{wi})/\max(S_w - S_{wi}) + (a_s - 1) \times (p - p_i)/\max(|p - p_i|), \tag{3.22}$$

which is related to the work of (DANAEI *et al.*, 2020). We note that $\Delta I_{PNORM} \in [-1,1]$ and $a_s = 0.85$ means that 85% of the impedance variations are due to saturation changes, and 15% are related to pressure changes. Figure 3.3a depicts the noise-free observed data. We added a gaussian noise with zero mean, standard deviation of 0.06, and an exponential correlation with a length of 5 cells, as shown in Figure 3.3b. We represented this uncertainty in the observed data by 5000 realizations of the map.

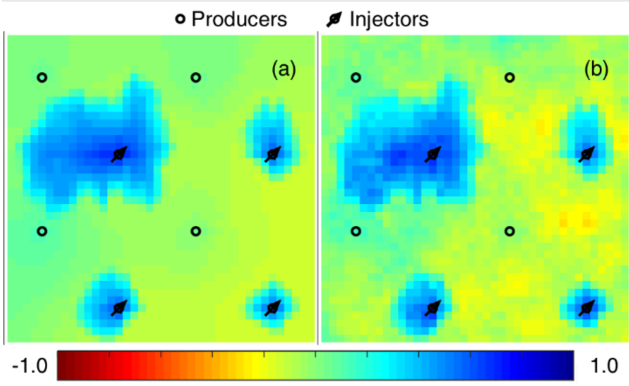


Figure 3.3: 2D case – observed data. The noise-free data is in (a), and the perturbed observed data is in (b).

Based on partial data about the synthetic truth model, we created a large initial ensemble of 5000 simulation models. At the location of each well, we sampled the porosity from a low variance distribution ($\sigma_{\phi wells} \approx 10^{-4}$), whose mean was the actual value shown in Figure 3.2a. Using this available information, we generated the models by applying a simple Gaussian Simulation. We used an exponential correlation model with a range of six cells, in which the maximum standard deviation was roughly 0.05. Figure 3.4a and b show the prior porosity mean and standard deviation maps, respectively.

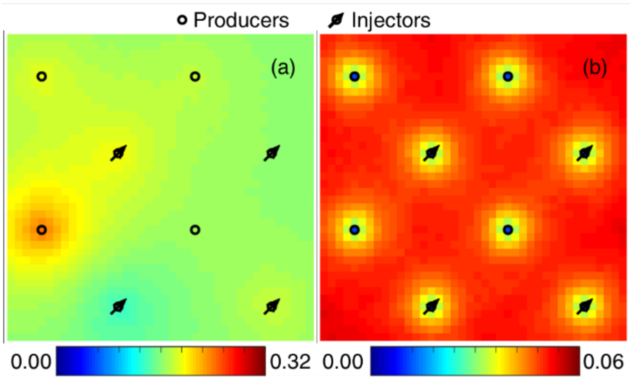


Figure 3.4: 2D case – maps of the prior mean (a) and standard deviation (b) of the porosity field.

3.5.2 Field-scale case

We used the simulation models from the UNISIM-I-H benchmark case (MASCHIO, C. *et al.*, 2013) as our field-scale example. Our motivation for testing the SEnRML method with local analysis in this problem is to face some of the real field application challenges. The models have 14 producer wells, 4 vertical and 11 horizontal, and 11 horizontal injectors. Figure 3.5 shows 2D maps of the mean porosity field of layers 1, 3, 9, and 12, including the well placement at each layer. In this example, we calibrate the reservoir porosity, permeabilities at each direction, and net-to-gross ratio, each with 38,466 active points. Therefore, the total number of parameters is almost 200,000, whose prior uncertainty we represent by an ensemble consisting of 500 models. We present other characteristics of this example in Table 3.3. For more details about the simulation models, we refer to (MASCHIO, C. *et al.*, 2013).

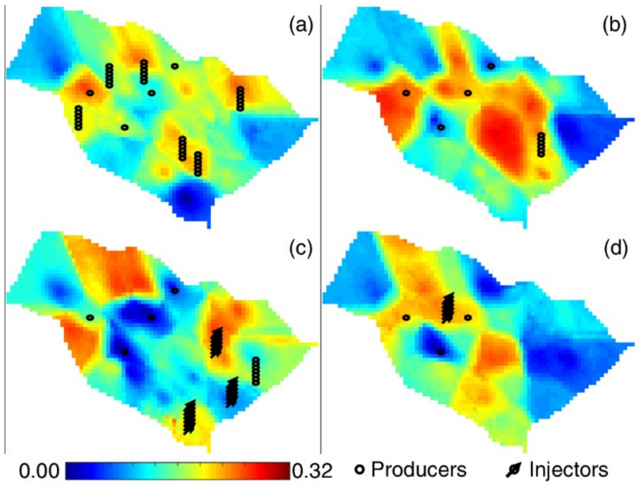


Figure 3.5: Field-scale case – prior mean porosity distribution of the reservoir layers 1 (a), 3 (b), 9 (c), and 12 (d), with the wells' placement. The figure only shows the completions in each layer. Not all wells are displayed.

Table 3.3: Characteristics of the field-scale case.

Characteristic	**Value**
Grid size	$81 \times 58 \times 20$ cells
Cell size	$100 \times 100 \times 8\ m$
Average porosity	0.14
Average horizontal permeability	$26\ mD$
Average vertical permeability	$5\ mD$
Fluid mobility ratio (water/oil)	From 0.9 to 3.0 (uncertain)
Fluid densities	Water: $1010\ kg/m^3$ Oil: $866\ kg/m^3$
Initial pressure	$321\ bar$
Number of active TLS data points	9324
Reservoir simulator	IMEX version 2017 (CMG, COMPUTER MODELLING GROUP LTD., 2017)

We generated acoustic impedance variation maps after 2618 days of production using a reference model that is not part of our simulation ensemble. To compare the final parameter distribution of the different methods to the synthetic truth, we used a reference model on the same scale as the simulation models. Furthermore, we used a petroelastic model (PEM), based on Gassmann's equation (GASSMANN, 1951), with dry-rock properties resembling the Hertz-Mindlin model (MAVKO; MUKERJI; DVORKIN, 2009). We estimated the fluids' properties using the Batzle-Wang correlations (BATZLE; WANG, 1992) and Wood's equations (MAVKO; MUKERJI; DVORKIN, 2009). Our rock model includes shale and quartz minerals, whose proportions were estimated using the net-to-gross ratio. The average of the Hashin-Shtrikman bounds (HASHIN; SHTRIKMAN, 1963; apud AVSETH; MUKERJI; MAVKO, 2005) provided the properties of the mineral mixture. For more details regarding this PEM, we recommend (DANAEI *et al.*, 2020).

To consider the vertical seismic resolution, we upscaled the impedance variation data in four horizons, each comprising 2 to 4 layers of the model. **Figure 3.6** depicts these horizons from the top to the bottom of the reservoir in (a) to (d), respectively. We perturbed the data with an arbitrary Gaussian noise with zero mean, standard deviation of $15\,\frac{g}{cm^3} \times \frac{m}{s}$, horizontal correlation length of $700m$, and no vertical correlation.

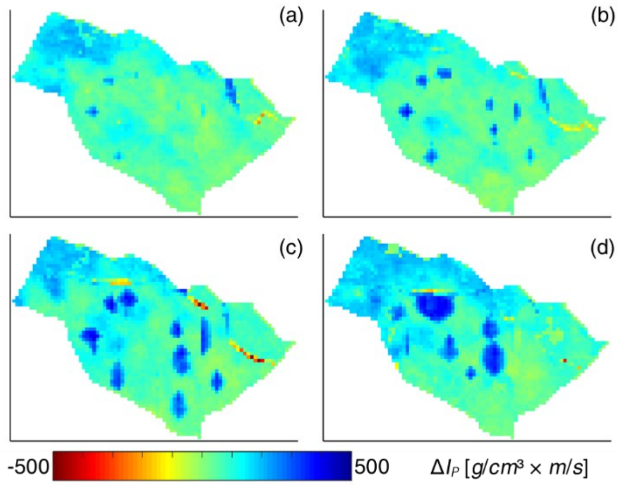


Figure 3.6: Field-scale case – maps of the observed data with added Gaussian noise. The data comprises four impedance variation horizons, which figures (a) to (d) depict from the top to the bottom of the reservoir.

3.5.3 Level of time-lapse seismic matching

There are different levels at which one may perform seismic history matching, e.g., seismic amplitude domain, elastic domain, and pressure and saturation changes (GOSSELIN *et al.*, 2003). We opted to work in the elastic domain by comparing acoustic impedance variations in our applications, combining pressure and saturation effects through a PEM. This domain is a popular option because it avoids seismic modeling in the forward simulations and circumvents the complicated seismic inversion to pressure and saturation fields (GOSSELIN *et al.*, 2003). Fahimuddin et al. (2010b) reported that the assimilation of time-lapse seismic data, using an ensemble-based method in the acoustic impedance domain, provided better results than the amplitude domain process. However, it is also possible to incorporate TLS data by comparing saturation changes, with the advantage of not requiring a PEM in the forward simulation (KETINENI *et al.*, 2020).

3.6 Results and discussion

We divide the results into two parts; the 2D case and the field-scale case, as shown by the two following subsections.

3.6.1 2D case

For the 2D case, we first discuss the large ensemble solution, followed by the small ensemble results, with local analysis. Then, we address the influences of the localization length, the local group size, and the automatic tapering formulation in the data assimilation results.

3.6.1.1 Large ensemble solution

Using a large ensemble, we can avoid the local analysis or the localization scheme, since the ensemble is large enough to deal with the known issues of low degrees of freedom and spurious correlations in the calibration process. We refer to this solution in this text as the quasi-ideal solution.

Figure 3.7 presents the evolution of the relative data deviation (a) and the relative deviation of the parameters (b) throughout the iterations of SEnRML with global analysis. We calculate these deviations for each ensemble member using equation 3.23, where $\boldsymbol{\zeta}$ refers to the simulated variable, $\boldsymbol{\xi}$ refers to the reference value for the same variable, $\boldsymbol{\eta}$ is a normalization factor, and N_{el} is the number of points. For instance, for the data deviation, $N_{el} = m$, $\boldsymbol{\zeta} = \boldsymbol{d}^{sim}$, $\boldsymbol{\xi} = \boldsymbol{d}^{\text{obs}}$, and $\boldsymbol{\eta} = \boldsymbol{d}^{std}$, which is a vector with the standard deviation of the measurement errors. For the deviation of the parameters, $N_{el} = n$, $\boldsymbol{\zeta} = \boldsymbol{x}^{a}$ or $\boldsymbol{\zeta} = \boldsymbol{x}^{f}$, depending on the iteration or prior ensemble, $\boldsymbol{\xi} = \boldsymbol{x}^{true}$ related to the synthetic truth, and $\boldsymbol{\eta}_n = \overline{x^{true}}$, which is the mean value of this type of parameter in the grid for the synthetic truth. We chose this latter normalization factor to avoid division by numbers close to zero in some regions.

$$\delta = \frac{1}{N_{el}} \sum_{n=1}^{N_{el}} |\zeta_n - \xi_n| / \eta_n \tag{3.23}$$

From the results of Figure 3.7, we can see that SEnRML, with global analysis, provided similar results to ESMDA for this problem. One would expect this similarity because the two methods solve just about the same problem using a large ensemble. Furthermore, it is possible to estimate that the calibration process converged after 4 to 5 iterations because the updates in both the data deviation and the parameter changes are negligible after this point.

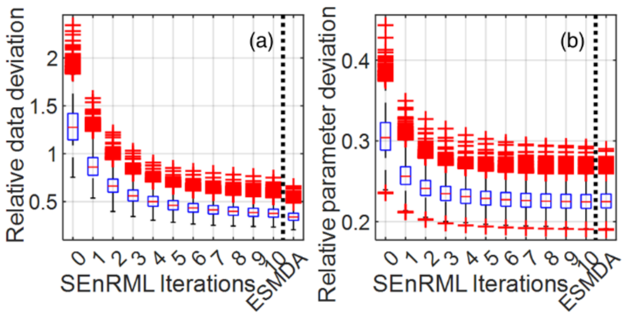


Figure 3.7: 2D large ensemble evolution of the data (a) and parameter (b) deviations throughout the 10 iterations using SEnRML with global analysis, compared to ESMDA with distance-based Kalman gain localization. The data deviation measures the distance between the simulated response and the measurements, while the parameter deviation indicates the difference between the calibrated models and the synthetic truth.

One can compare the porosity mean maps of the prior ensemble and the large ensemble's final solution (Figure 3.8a and b, respectively) as well as the standard deviation maps (Figure 3.8f and g, respectively). We show only the SEnRML solution because the differences to the ESMDA solution are nearly unnoticeable. As expected, the large ensemble mean porosity map is more heterogeneous than the prior distribution, following the main tendencies of the synthetic truth (Figure 3.2). The calibration process reduces the ensemble's variability, notedly between the injectors and the producers, as these regions contain the parameters that are more correlated to the impedance variations caused by the water injection in the reservoir.

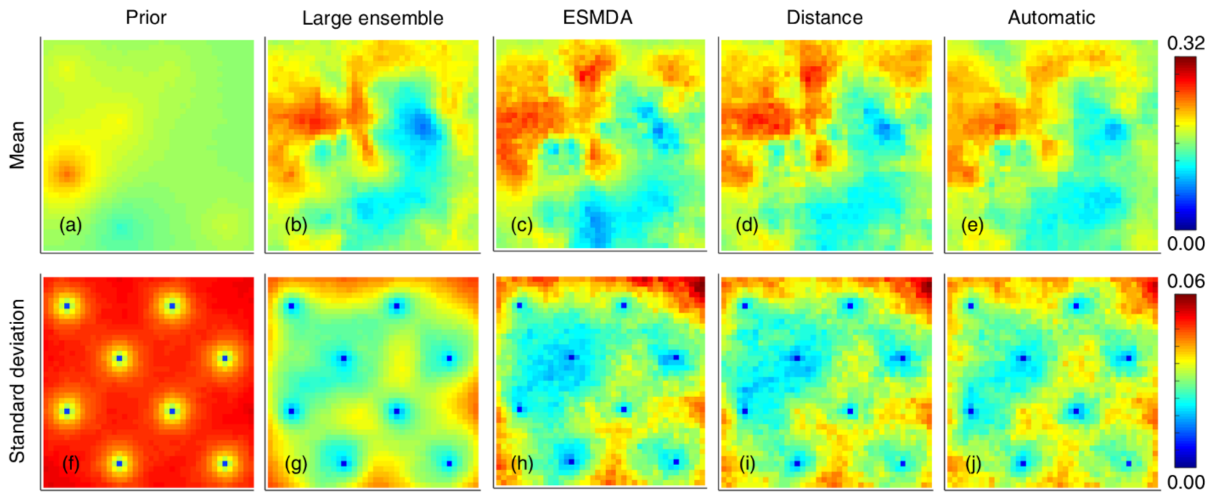


Figure 3.8: 2D case maps. The mean maps are at the top, and the standard deviation maps are at the bottom. The figure presents the prior (a and f), large ensemble (b and g), ESMDA with Kalman gain localization (c and h), SEnRML with distance-based local analysis (d and i), and SEnRML with automatic local analysis (e and j) results.

In the next sections, we analyze the results obtained using smaller ensembles. We use the large ensemble results as a reference for the final distribution of the parameters, and assume that SEnRML and ESMDA adequately sampled the posterior parameter distribution using the large ensemble. It is worth mentioning that, even though the sampling-related errors still influence the large ensemble solutions, we neglect this influence when compared to the errors in the small ensemble tests.

3.6.1.2 Small ensemble solution (100 models)

After the run in a quasi-ideal condition in terms of sampling issues, we tested both algorithms using a smaller ensemble of 100 models. We were expecting worse results for both methods due to the sampling problems and the information loss caused by the subspace projection in SEnRML. Therefore, we apply local analysis with smooth data influence tapering in SEnRML, and Kalman gain localization in ESMDA to try to mitigate the limited-ensemble-related problems. Also, we apply the widely known Gaspari-Cohn function (GASPARI; COHN, 1999) in both methods.

Table 3.4 shows the distribution of the parameters obtained by each method compared to the quasi-ideal solution. We computed the relative deviations using equation 3.23, using the large ensemble test values as both reference and the normalization factor. We see that the three methods, SEnRML, with distance-based and automatic local analysis, and ESMDA, led to relatively close results. When comparing them with the quasi-ideal test, the deviations were lower than 10% in both the mean and the standard deviation. SEnRML, with automatic local analysis, provided slightly better results, with a 7.6% deviation in the mean and 6.9% in the standard deviation.

Table 3.4: Relative deviations of the parameter distributions in the 2D case with an ensemble of 100 models.

Method:	SEnRML distance-based	SEnRML automatic	ESMDA
Mean	0.082	0.076	0.099
Standard deviation	0.088	0.069	0.094

One can confirm the results in Table 3.4 visually from the maps in Figure 3.8. Figure 3.8 b, c, d, and e illustrate that the porosity mean maps are alike. One exception is the low porosity region in the northeast section of the reservoir. This section varies in all the cases with the small ensemble, and it is smaller than the quasi-ideal solution. Another noticeable difference in the mean maps is the high porosity region close to the northeast corner in (c) and (d), which does not appear in the other maps. The ESMDA results in Figure 3.8 c also exhibit a low porosity region in the south that is more pronounced than in the other maps. However, these differences are minor.

Comparing the standard deviation maps in Figure 3.8 g, h, i, and j, we also note that the solutions are similar, except for some secondary aspects. First, the standard deviation maps obtained using 100 models in the ensemble are noisier. Also, the calibration using all three methods underestimated the ensemble variability in some regions between injectors and producers. We believe that the sampling noise in the correlation between parameters and data caused both issues. The local analysis scheme and the Kalman gain localization do not entirely negate that. The latter effect seems to be more severe in the distance-based methods, but the automatic scheme alleviated it.

Figure 3.9 presents the evolution of the relative deviation of the data (a) and the parameters (b) throughout the iterations of the SEnRML method with distance-based local analysis. We compare the results to ESMDA with the distance-based Kalman gain localization. It is possible to notice that the approximate convergence of the method occurs around iteration five and that the final deviations are close to ESMDA. The SEnRML method, with distance-based local analysis, led to slightly lower data deviation, with a lower spread.

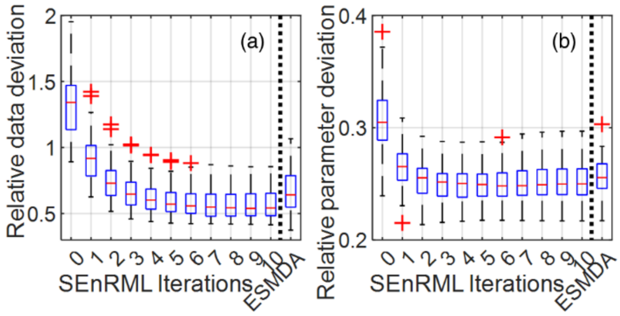


Figure 3.9: 2D small ensemble evolution of the data (a) and parameter (b) deviations throughout the 10 iterations using SEnRML with distance-based local analysis, compared to ESMDA with distance-based Kalman gain localization.

The same results for the automatic local analysis scheme appear in Figure 3.10. The variations of the data deviation and the parameter are minor after iteration 5. Furthermore, the final relative deviations are slightly lower than ESMDA.

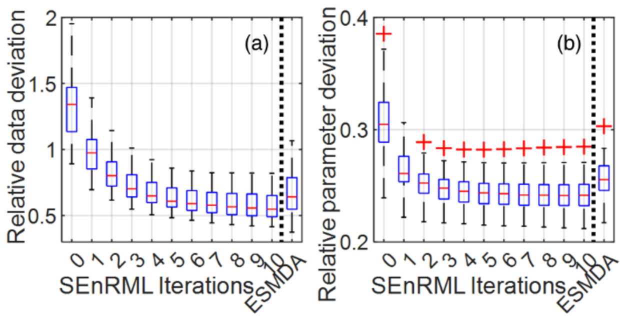


Figure 3.10: 2D small ensemble evolution of the data (a) and parameter (b) deviations throughout the 10 iterations using SEnRML with automatic local analysis, compared to ESMDA with distance-based Kalman gain localization.

3.6.1.3 Influence of the localization lengths

In all previous comparisons, one needed to define the localization length for the distance-based schemes. We chose the distances that nearly minimized the relative deviation of the mean and the standard deviation of the parameters when compared to the quasi-ideal case. To simplify this investigation, we only considered isotropic distances. We ran some data assimilations and varied the lengths to obtain these distances, as Figure 3.11 depicts. We notice that the best distance for SEnRML, with distance-based local analysis, and ESMDA, with distance-based localization, was of around 8 cells. Previous studies reported differences when comparing tapering of the ensemble anomalies and tapering of the Kalman gain matrix (CHEN, Yan; OLIVER, 2017). In this small 2D case, this difference was minor. However, focusing on the standard deviation curves in Figure 3.11, one can notice that SEnRML, with tapering of the ensemble anomalies, reached a better representation of the ensemble variability with about 7-cells range. At the same time, ESMDA, with Kalman gain localization, provided improved results with 8 cells.

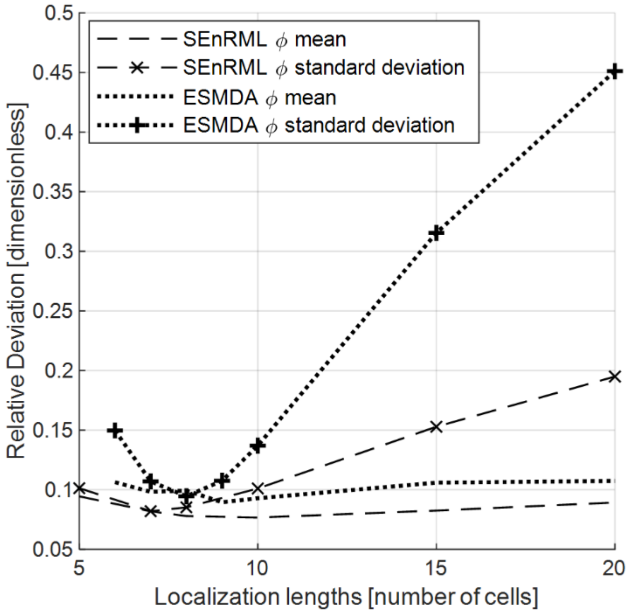


Figure 3.11: 2D small ensemble – relative deviation of the parameter distributions for different localization lengths.

3.6.1.4 Influence of the local group sizes

The results in Figure 3.11 are slightly different from the values in Table 3.4 because we used groups containing 3 × 3 cells in the tests of Table 3.4 and only one cell at a time in Figure 3.11. The choice of the groups' size in the local analysis scheme is a tradeoff between accuracy and computational costs. It is worth mentioning that, since the group analyses are independent, one can alleviate the impact of the computational cost by parallelizing the process. Nevertheless, if we increase the size of each analysis step from the parameters in a cell for a group containing 3 × 3 cells, we reduce the number of individual analysis computations by almost one order of magnitude. Therefore, we checked how larger local groups changed the relative deviation in the parameter distribution and presented these results in Figure 3.12. As expected, there is an overall increasing trend in the relative deviation as the group size grows. However, this increase in the deviations seems to be relatively small for groups that are significantly smaller than the localization length. Hence, we opted to use groups consisting of 3 × 3 cells in the other tests of this work.

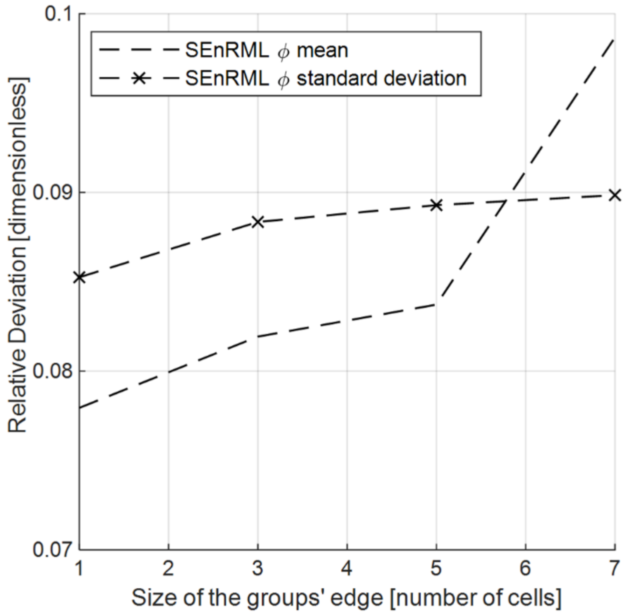


Figure 3.12: 2D small ensemble – relative deviation of the parameter distributions for different group sizes.

We generated the results of Figure 3.12 using distance-based tapering with 10 cells for the localization length with 1×1 cell groups. Moreover, since we compute distances from the group borders, we maintained the amount of data selected for each update nearly constant by adjusting the localization lengths for different group configurations.

3.6.1.5 Automatic data selection and tapering formulation

There are three aspects of the automatic data selection and influence tapering formulation that we needed to define for our applications. First, it was necessary to define how to obtain a representative tapering vector for the local groups. In the 2D case, we used the mean of the pseudo-distance variables for each group, which led to consistent results compared to the other distance-based methods. Moreover, it is necessary to define how one computes the correlation threshold and the pseudo-distance variable.

Luo and Bhakta (2020) proposed two methods for defining the correlation threshold. The first one consisted of estimating the sampling correlation noise by shuffling the ensemble and computing the correlation between parameters and data. The second method allows for reducing the computational costs by applying an asymptotic approximation to the standard deviation of the correlation noise, equation 3.15, resulting in one threshold for the data assimilation case. We compare these two approaches in Figure 3.13, where one can note that the simplified procedure provided a correlation threshold close to the mean of the thresholds from the shuffling method. Furthermore, comparing the data assimilation results for both approaches, we did not find any significant difference. Therefore, we opted to use the simplified method in our tests, avoiding the computation of the correlation between the parameters and the data using the shuffled ensemble.

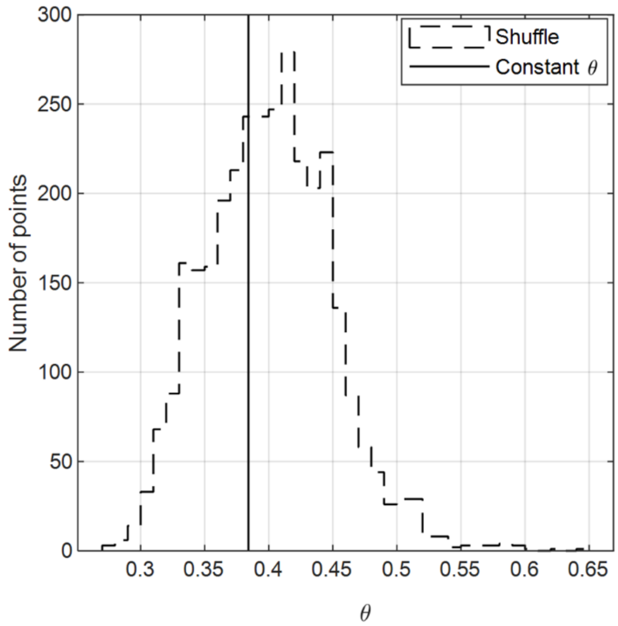


Figure 3.13: 2D small ensemble – a comparison of the methods for the correlation threshold.

Luo and Bhakta (2020) suggested computing the pseudo-distance variable using equation 3.18. Using this formulation with a threshold of around 0.4, we obtained an evident selection of the data that would influence each group. The method assigned low tapering values in the magnitude of 10^{-2} to the data that would not influence a particular group, whereas it assigned tapering close to 1 to the data that would.

Nevertheless, if we repeat the test with a larger ensemble, and consequently a lower θ, as in the case with, $\theta = 0.2$, this data segregation is not as evident as before. We show the behavior of the tapering function for two different threshold values in Figure 3.14. We limited the interval to $0.5 \leq \frac{|r|}{\theta} \leq 2.5$ because most of our estimated correlations lied inside this range in all tests. We see that the tapering using equation 3.18 is in the interval of $0.1 < \rho < 0.6$, which means that all the data would be selected with a similar weight. This result tends to limit the automatic tapering capability to mitigate the influence of the spurious correlations in the data assimilation, and it also attenuates the updates based on data that is closely related to the parameters.

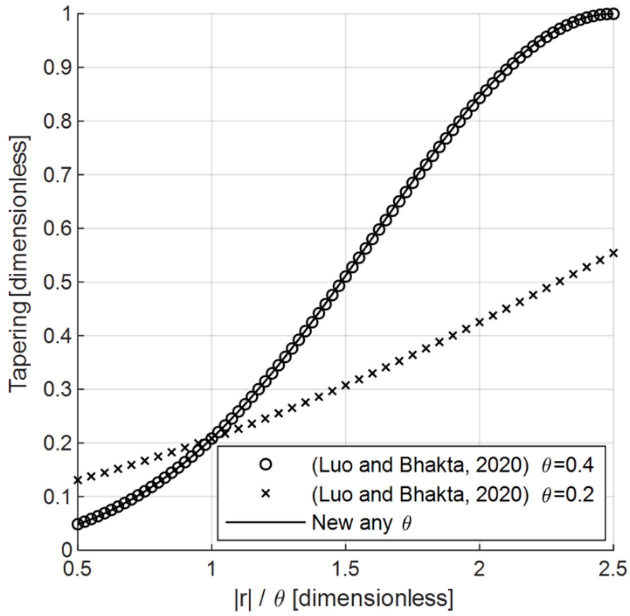


Figure 3.14: Automatic tapering as a function of the ensemble correlation and the threshold.

To investigate the influence of the tapering range at the data assimilation result in the 2D case, we used the definition of equation 3.18 and an ensemble of 400 models, which resulted in $\theta \approx 0.19$. Table 3.5 compares the final distribution of parameters using SEnRML with local analysis to the quasi-ideal solution. Differently from Table 3.4, the original formulation from Luo and Bhakta led to a more significant deviation than the distance-based scheme, notedly for the standard deviation. Although not shown in the table, the standard deviation values were biased towards lower values when compared to the quasi-ideal solution.

Table 3.5: Relative deviations of the parameter distributions in the 2D case with an ensemble of 400 models.

Method:	SEnRML distance-based	SEnRML (Luo and Bhakta 2020)	SEnRML New
Mean	0.044	0.051	0.044
Standard deviation	0.036	0.049	0.037

To improve the data assimilation results using different ensemble sizes, we propose a new function to compute the pseudo-distance (equation 3.19). We obtained this function by fitting it to equation 3.18 with a threshold of 0.4, close to the 2D case with 100 models. Although the modified formulation given in (3.19) avoided a tendency to underestimate the final ensemble variance in our tests, we do not claim that it is optimal in any sense. This optimization can be addressed in future studies.

One could argue that the tendency to underestimate the ensemble variability in this test is negligible because the difference in the values in Table 3.5 are low. Nevertheless, this impact will be more relevant in big reservoir models with more data because there will be more data points with a negligible correlation to a particular group's parameters. The field-scale case results, which we address in the following section, will evidence this fact.

3.6.2 Field-scale case

In the field-scale case tests, we considered four implementations of SEnRML: the global update; the automatic local analysis scheme with the original formulation from Luo and Bhakta (2020); the distance-based scheme; and the new formulation of the automatic local analysis scheme. We compare the results to ESMDA, with Kalman gain localization, using ten iterations and constant inflation factors. For the distance-based schemes, we chose localization lengths of $1000\ m$ for SEnRML and ESMDA. These choices are not optimal in any sense. In the automatic data selection and tapering method, the threshold is $\theta \approx 0.2$ and the local analysis groups consisted of vertical columns of 3×3 reservoir cells, including all the layers.

The field test's main limitation is that we do not have a quasi-ideal solution to compare the local analysis methods in detail. Therefore, to compare ESMDA and the different SEnRML method implementations, we used the relative deviations and the parameter updates. We computed all these metrics using equation 3.23. The data and parameter deviations are equivalent to the analysis of the 2D case. When computing the deviations for the well rates, we used the observed data from the synthetic truth as the reference, and the arithmetic mean of the simulated and observed rates, summed to a small number, as the normalization factor. This procedure avoided amplifying the deviations due to the division by small numbers.

In general, the methods provided similar results for the relative deviations with some minor differences, as we note in Figure 3.15. SEnRML with a global update, included for comparison, shows slightly worse results for the deviations of the data (a), the parameter (b), and the production rates (c and d). The two distance-based schemes provided the lowest results for the data deviation, whereas the automatic scheme led to slightly better results in terms of parameter deviation.

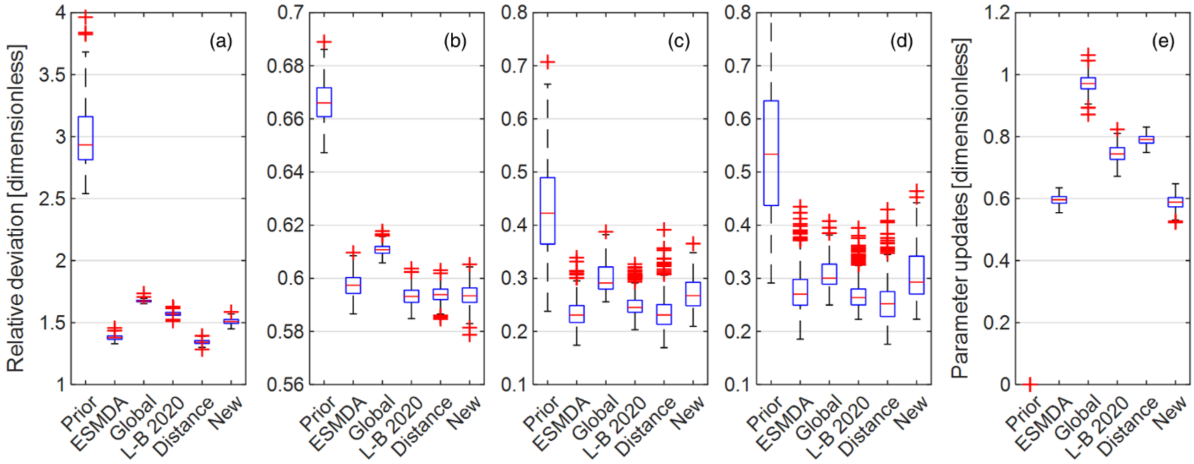


Figure 3.15: Field-scale case – comparison of the methods in terms of the relative deviation of the data (a), the parameters (b), the oil rate forecast (c), the water rate forecast (d), and the parameter updates (e).

Reducing the rates-related metrics indicates that the time-lapse seismic data improves the production forecast, as expected, even though we did not assimilate production data. Nevertheless, it is not straightforward to compare the minor differences between the methods using these criteria because there are nonlinear interactions between the parameters and possible error combination effects. Furthermore, it is crucial to include production data in the calibration of the parameters in a complete field study and consider the well history matching as a model quality assessment criterium. We did not include production data in this study because we wanted to focus on our method's capability in handling seismic data. A complete field application with our method will be the focus of a future study.

We also compare the methods in terms of the parameter updates in Figure 3.15e. As expected, SEnRML with global update causes more changes in the model parameters because it considers all the data, regardless of the distance or the correlation level. Comparing the two distance-based methods, one can note that SEnRML, with tapering of the ensemble anomalies, modifies the models more than ESMDA with Kalman gain localization.

By changing the pseudo-distance variable's formulation in the automatic tapering procedure, we intended to improve the data segregation based on their correlation with the group parameters. The lower parameter updates in the new formulation in Figure 3.15e, when compared to Luo and Bhakta (2020), indicates that this change was significant. The updates using the new formulation were also close to ESMDA. One could argue that the Luo and Bhakta (2020) case updates are close to the distance-based method. Nevertheless, this similar level occurs because it considers all the data to update each parameter using a tapering higher than 0.1, but it also attenuates the updates due to the highly correlated data with tapering values of around only 0.55 (see Figure 3.14). Also, this tapering combination tended to underestimate the ensemble variance, as we discuss later in this section.

Apart from comparing the final results, it is essential to verify the convergence speed of the methods. We use two metrics to analyze this: the relative data deviation and the parameter updates. These two variables are related to the objective function (equation 3.5). There was no noticeable difference between the two local analysis schemes in our tests regarding the number of iterations to achieve convergence. Both SEnRML implementations, with distance-based local analysis and automatic local analysis, seem to converge after about six iterations in this field example, in the sense that only minor incremental changes happen afterward (Figure 3.16). No significant differences were noted between the two formulations of the automatic schemes.

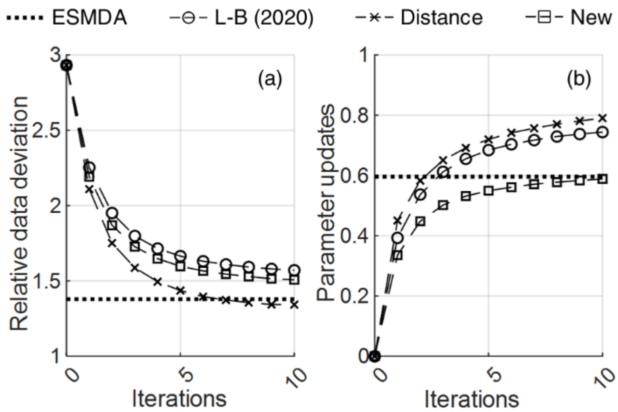


Figure 3.16: Field-scale case – the evolution of the data deviation (a) and the parameter updates (b) throughout the iterations using SEnRML.

Another important aspect regarding the time-lapse seismic data assimilation in this field-scale test is the spatial distribution of the reservoir properties and the ensemble variability after the calibration process. Here, we analyze these aspects using the mean porosity and the porosity standard deviation maps shown in Figure 3.17. As expected, all the data assimilation methods enhance the heterogeneity, since the prior ensemble mean map Figure 3.17a is smoother than all others. There are minor changes among the methods' results, but they seem to capture the same main aspects of the mean porosity distribution. For instance, all methods increased the porosity in the center of the reservoir and diminished it in the northwest area. Furthermore, we identified a tendency to obtain more extreme values for the reservoir properties at particular regions of the model using SEnRML, as seen in the high porosity region on the east of the model in Figure 3.17 i and k, when compared to c.

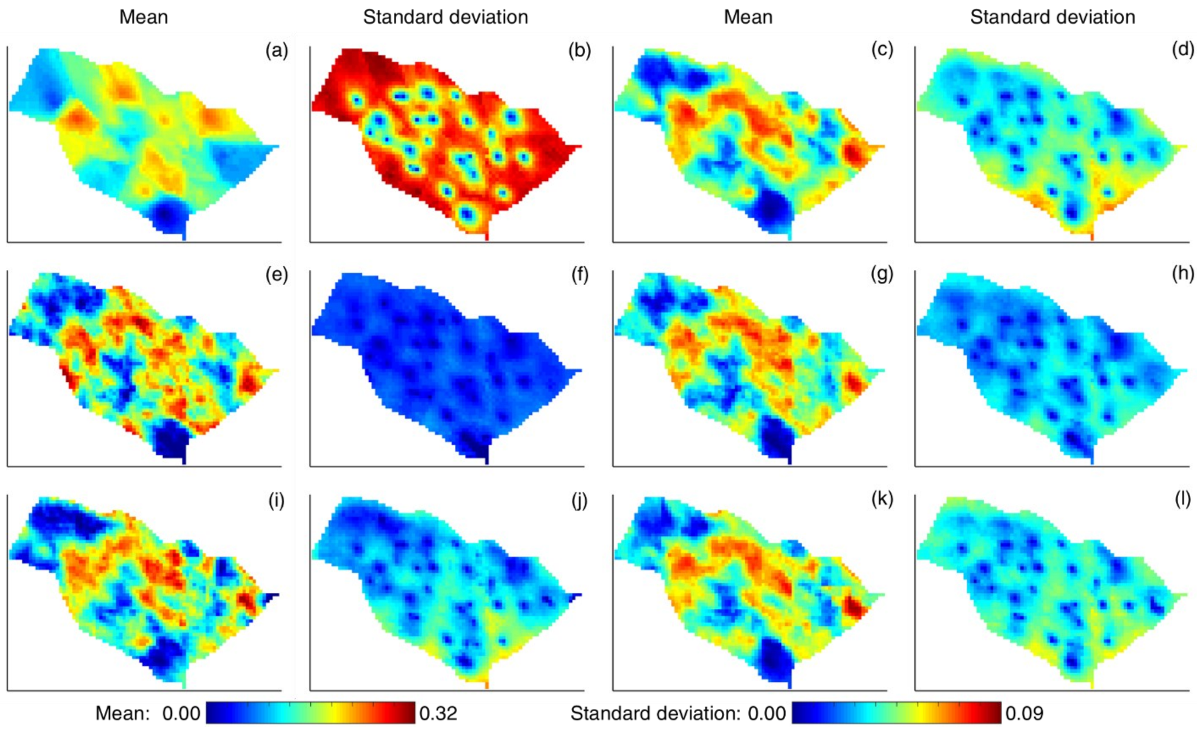


Figure 3.17: Field-scale case – maps of the porosity mean and standard deviation of the first layer. The figure presents the prior (a and b), ESMDA (c and d), EnRML with global analysis (e and f), automatic local analysis with the Luo et al. (2020) formulation (g and h), distance-based local analysis (i and j), and automatic local analysis with the new formulation (k and l).

The main differences among the tests occur in the standard deviation maps of Figure 3.17. SEnRML, with the original Luo and Bhakta (2020) formulation (Figure 3.17h), led to significantly lower standard deviations than the other methods with localization or local analysis (Figure 3.17d, j, and l). Compared to the global analysis (Figure 3.17f), the local analysis with the original formulation from Luo and Bhakta (2020) reduced the tendency to underestimate the ensemble variability. This, however, proved to be insufficient, particularly in the south of the reservoir, if we compare it to the other methods. The main reason for this behavior is that it uses all the data to update all the parameters with weights higher than 0.1 in our tests (Figure 3.14 for $\theta \approx 0.2$).

The new formulation for the automatic scheme improved the final ensemble variability when compared to the Luo and Bhakta (2020) formulation (Figure 3.17l), in the sense that it is more comparable to the distance-based schemes, particularly ESMDA (Figure 3.17d). The main difference between the new automatic formulation and ESMDA is the lower standard deviations in the southern border. It seems that the automatic scheme with our group configuration tended to include more data in the analysis for the border of the reservoir.

However, it is not clear if this inclusion is favorable to the data assimilation since a large ensemble solution is not available.

The standard deviation maps of the two distance-based schemes resembled in the south, but SEnRML led to lower variability in the northwest and east (Figure 3.17d and j). This result confirms the tendency previously depicted in Figure 3.15e. The tapering in SEnRML with local analysis seems to assign higher weights to the data for each group update when compared to ESMDA with Kalman gain localization.

The standard deviation maps of Figure 3.17 depict the general behavior of the uncertainty reduction and the final ensemble variability among the methods that we tested. Nevertheless, to provide a notion of the ensemble variability of all grid parameters throughout the reservoir layers, we also computed the average variance reduction compared to the prior ensemble. ESMDA with Kalman gain localization provided the highest final average variance, with 48% of the former. SEnRML, with the new automatic formulation for the local analysis, generated an ensemble whose variance, on average, is 39% of the former. Our distance-based local analysis with a 10-cell range resulted in 31% of the former variance. This result seems drastic in terms of variability reduction, indicating that our data tapering scheme may require shorter localization lengths. We also tested the distance-based local analysis scheme with an 8-cell range, and the variance was 36% of the former on average. Furthermore, the tendency to generate calibrated values close to each parameter's physical limits in some reservoir regions influences this result. Finally, the original formulation from Luo and Bhakta (2020) for the automatic local analysis and the global analysis resulted in 20% and 6% of the prior ensemble variance, respectively.

Despite secondary differences in the maps of Figure 3.17 and the final variance values, we consider that ESMDA with Kalman gain localization and SEnRML, with distance-based and automatic local analysis, produced comparable ensembles of reservoir models. The field-scale application results indicate that the methods can lead to equivalent calibration qualities if the localization or local analysis parameters are adequately set. Nevertheless, we did not optimize the localization lengths in any sense in our tests because this is out of the scope of this study.

3.6.2.1 Selecting data and combining the tapering coefficients for each group

In the field-scale tests, we opted to update all the cells in a column of the reservoir using the same data with the same tapering vectors. In the distance-based methods, this fact means that we only use the horizontal distance between the data and the cells or the group to select the data and compute the tapering. In the automatic local analysis scheme, it means that we use a single tapering vector for the data and ensemble anomalies. Therefore, we need to define a procedure to combine each group's pseudo-distances into a single vector.

In the 2D case, we used the mean value of the pseudo-distance among the group's parameters. Nevertheless, in the 3D case, this procedure would lead to low tapering values for all the data's influence because the correlation varies in the vertical direction. One could opt to segregate the parameters vertically to simplify the combination, but this procedure increases the analysis step's computational costs. An intermediate approach could be to group only similar reservoir formations vertically, but this was not tested in our examples. Furthermore, a group configuration with vertical segregation would differ from the familiar distance-based format used for comparison.

Therefore, we maintained the group distributed in vertical columns of cells and opted to combine the pseudo-distances by taking the 5^{th} percentile among the group's parameters. We used this number because each horizon of our seismic data corresponds to about two simulation model layers. Since our model has 20 layers, we may assume that each horizon is closely related to 10% of the reservoir, but it also influences the other layers. The 5^{th} percentile is related to the median of the correlations in the two layers. Therefore, the idea is to select all data points that influence the group, at least in the layers that correspond to the vertical data position.

We compare the two automatic formulations tapering values for one group and the top horizon data in Figure 3.18. The Luo and Bhakta (2020) formulation (a and c) led to tapering greater than 0.1 and lower than 0.6, which confirms the results in Figure 3.14. The new formulation (b and d) improved the selection of the data related to the group by assigning tapering close or equal to one to the more correlated data and in the magnitude of 10^{-2} to less-correlated points.

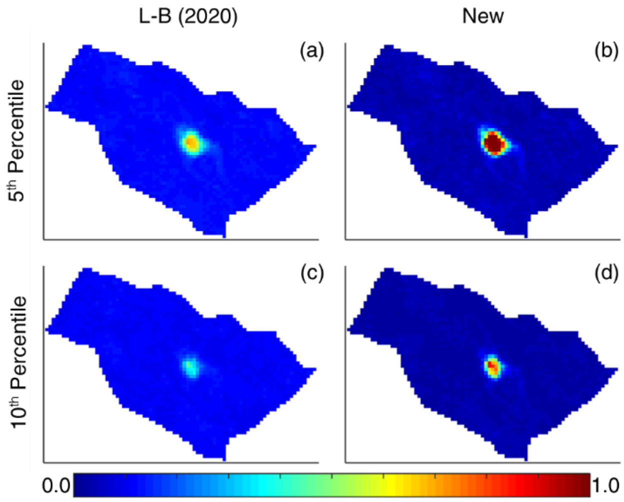


Figure 3.18: Field-scale case – map of the tapering for the 258th group, applied to the first horizon of the data. We compare the Luo and Bhakta (2020) formulation (a and c) to the new formulation (b and d). The result taking the 5th percentile is on the top (a and b), and the 10th percentile is on the bottom.

Changing the percentiles for combining the pseudo-distances has a similar effect to reducing the localization length (Figure 3.18). Unfortunately, this variation means that there is still a parameter related to the automatic scheme tuned to a particular case. Nevertheless, one could select it without running the iterations of the data assimilation by selecting the percentile that would associate tapering values as high as 1.0 to the correlated data and parameters. In future studies, we plan to avoid this tuning by proposing an efficient way to group parameters considering the correlation with the data.

3.7 Additional analysis: computational requirements with increasing data set size

We compared the SEnRML algorithm, implemented with two local analysis methods to ESMDA, with Kalman gain localization for seismic data assimilation in this work. We showed that the SEnRML algorithm with local analysis is promising for integrating big data sets because the analysis equations only scale linearly with the data set size. Nevertheless, our tests comprised relatively small models with data sets whose sizes were roughly 10^3 and 10^4 data points in the 2D and the field scale tests, respectively. These tests helped validate the SEnRML method with local analysis in seismic data assimilation by comparing it to ESMDA with our available computational resources.

We will investigate the assimilation of a bigger data set, comprising TLS and production data, using the SEnRML method with local analysis in a future study. Nevertheless, we included a simple test to show our method's benefits in applications with an increasing data set. In this test, we performed a model update in a single iteration using ESMDA with Kalman gain localization and SEnRML, with distance-based local analysis, using a 10-cell range in both methods. Aiming at testing these algorithms with different data set sizes, we repeated our data, perturbing it with random Gaussian noise, as if we acquired redundant seismic data at the same time. This test is not a usual TLS data assimilation problem, but it mimics the difficulties in handling big data sets.

We used the ESMDA algorithm for a large number of measurements reported by Emerick (2016). We updated the parameters using 5000 rows of the Kalman gain at a time, to reduce the memory requirements at this stage of the analysis scheme. By using this configuration, we guarantee that the peak memory demand occurs at the initial computations that do not depend on the definition of the number of rows. For more details about this ESMDA implementation, refer to Appendix B of (EMERICK, Alexandre A., 2016).

We start our analysis with the memory requirement, which was critical in our tests. Figure 3.19 depicts the memory demand of SEnRML and ESMDA with increasing data set size. As expected, the ESMDA requirement increased as the square of the data set size grew, enabling it to be tested with up to 10^5 data points, with our available resources. On the other hand, the growth of the SEnRML requirements was linear to the data size, enabling data assimilation of more than 3×10^5 with the same machine, demanding less than $35\ GB$. We also tested a faster configuration of our scheme by setting local groups of 5×5 columns of reservoir cells and reducing the E matrix number of columns to $4N$. In this case, we were able to perform the analysis step with 6×10^5 data points and required less than $35\ GB$. This is a tradeoff between accuracy and computational cost because it increases the sampling error in the $\boldsymbol{C}_{dd}$ representation and it divides the reservoir into coarser local groups. However, this tradeoff did not significantly impair the data or parameter relative deviation using our original field-scale test, without data repetition, as shown in Figure 3.20.

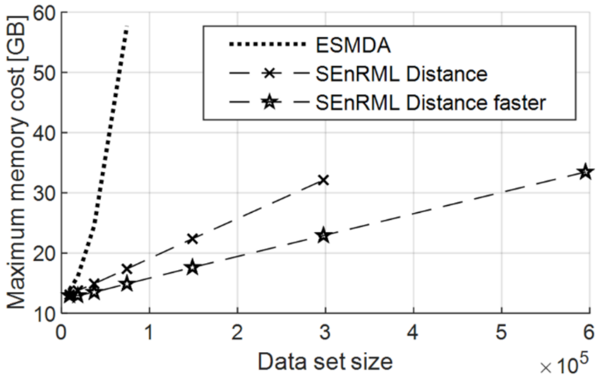


Figure 3.19: Required memory test with increased data set size.

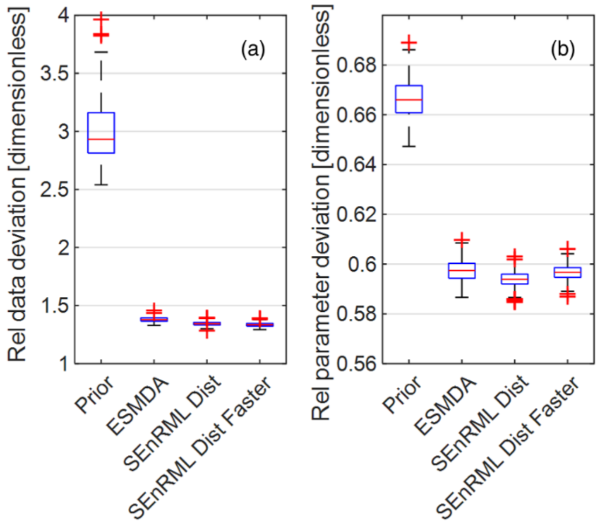


Figure 3.20: Field-scale case – comparison of the relative data deviation (a) and relative parameter deviation (b) of SEnRML with distance-based localization and the faster implementation of this method.

Figure 3.21 shows our CPU time test with the increasing amount of data. In this case, our initial configuration of SEnRML, with distance-based local analysis, was slower than ESMDA with Kalman gain localization. The main reason for this is that it is necessary to build matrices with a subset of the data and perform the SVD using them for each local group. Nevertheless, this computation time is not critical because the local analyses are independent, and one could fully parallelize them if needed. Furthermore, in our tests, the total CPU time of all local updates in serial computation was just a fraction of the time needed to run the forward reservoir simulations. Nevertheless, our method allows for performing a tradeoff between computation costs and accuracy by selecting fewer local groups and reducing the size of the E matrix. By using this faster configuration, we obtained significantly lower CPU time, outperforming ESMDA, even for relatively small data sets.

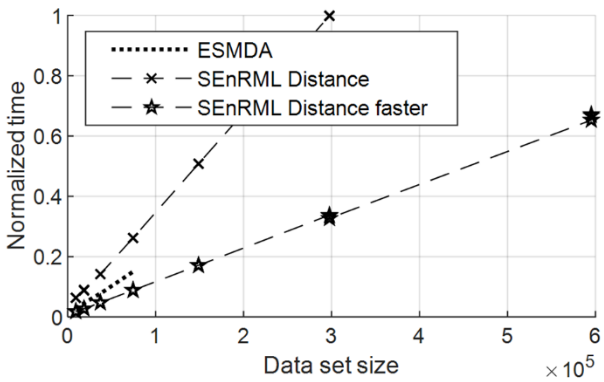


Figure 3.21: CPU time test with increased data set size.

Although it is not evident in Figure 3.21, the ESMDA CPU time tends to increase with the square of the data size. To show this, we registered the CPU time to initiate the row updates in Figure 3.22. This time starts short but can increase quickly with the data size if it reaches more than 10^5 data points.

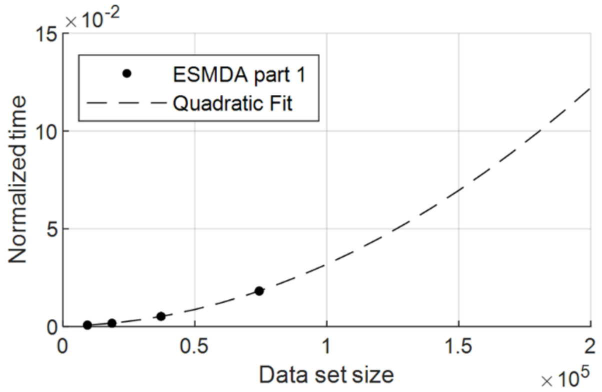


Figure 3.22: CPU time test of the first part of the ESMDA algorithm.

3.8 Summary and conclusions

In this work, we propose a way to use local analysis with the Subspace Ensemble Randomized Maximum Likelihood (SEnRML) method to improve time-lapse seismic data assimilation. The local analysis can mitigate the influence of spurious correlations and increase the degrees of freedom. We apply the method in two synthetic cases involving the assimilation of time-lapse seismic data.

The specific conclusions of the current work are:

- We successfully assimilated time-lapse seismic data using an efficient implementation of the Subspace Ensemble Randomized Maximum Likelihood method with local analysis.
- Although our test comprised relatively small reservoir models, we showed that this method is promising for reservoir models with big data sets because the memory requirement and CPU time increases linearly with the data size. These results occur because the equations scale linearly with the size of the data set. This characteristic is a significant improvement from previous methods, in which the computational costs scale with the square of the data set size, hindering applications with big data sets.
- The data mismatch and the parameter mismatch results, using both the distance-based and the correlation-based local analysis implementations, were comparable to ESMDA with Kalman gain localization. These results indicate that it is possible to obtain a satisfactory data match using our implementations of SEnRML with local analysis. This fact corroborates that the data reduction obtained by projecting it onto the subspace

created by the deconditioned ensemble anomalies did not cause a significant information loss that would jeopardize the quality of data assimilation.

- The automatic local analysis scheme implemented with SEnRML can lead to similar results as the distance-based methods for time-lapse seismic data assimilation.
- We proposed a different formulation for the automatic local analysis scheme that improved the final ensemble's variability results.
- In the automatic tapering case, it was necessary to choose the tapering vector for each group using a percentile of the pseudo-distance distribution among the group parameters. The choice of this percentile may be case-dependent. However, one can define it without running the full data assimilation process using different values.

Future studies intend to test the SEnRML method with local analysis in more complex cases, including bigger data sets and production data. Furthermore, it would be interesting to investigate more flexible ways to define the local groups and improve the formulation of the pseudo-distance variable in different situations.

3.9 Nomenclature

Variables

$\mathbf{1}$	Vector whose elements are equal to one
a_s	Saturation anomalies weight
$\boldsymbol{d}$	Data vector
$\boldsymbol{g}$	Forward model
p	Pressure
m	Number of datapoints
n	Number of parameters
n_p	Number of different types of parameters, e.g., net-to-gross, permeabilities at each direction, porosity
r	Correlation coefficient
$\boldsymbol{w}$	Vector of coefficients of parameters updates
$\boldsymbol{x}$	Vector of parameters
z	Element of pseudo-distance dummy variable
$\bar{\boldsymbol{z}}$	Vector of pseudo-distance dummy variable
$\boldsymbol{A}$	Matrix of ensemble anomalies
$\boldsymbol{C}_{dd}$	Covariance matrix of measurement errors
$\boldsymbol{D}$	Matrix of data realizations
$\boldsymbol{E}$	Matrix of measurement perturbations
$\boldsymbol{H}$	Matrix of innovations
$\widehat{\boldsymbol{H}}$	Tapered matrix of innovations
$\boldsymbol{I}$	Identity matrix
$\boldsymbol{\mathcal{J}}$	Cost function
N	Ensemble size

N_E	Number of columns of $\boldsymbol{E}$
N_{el}	Number of points
$\boldsymbol{R}$	Correlation matrix between the prior realization of parameters and simulated data
$\boldsymbol{S}$	Matrix of predicted and deconditioned ensemble anomalies
$\widehat{\boldsymbol{S}}$	Tapered matrix of predicted and deconditioned ensemble anomalies
S_w	Water saturation
$\boldsymbol{U}$	Matrix of the left singular vectors of $\boldsymbol{S}$
$\boldsymbol{V}$	Matrix of the right singular vectors of $\boldsymbol{S}$
$\boldsymbol{W}$	Matrix of coefficients of parameters updates
$\boldsymbol{X}$	Matrix of the ensemble of parameters
$\boldsymbol{Y}$	Matrix of predicted ensemble anomalies
$\boldsymbol{Z}$	Matrix of eigenvectors of the modified low-rank representation of the covariance matrix of measurement errors
$\boldsymbol{\delta}$	Relative deviation
γ	Step-length parameter
$\boldsymbol{\varepsilon}$	Measurement errors
$\boldsymbol{\zeta}$	Simulated value
$\boldsymbol{\eta}$	Normalization factor
θ	Correlation threshold
$\boldsymbol{\xi}$	Reference value
$\boldsymbol{\rho}$	Tapering vector
σ_ϵ	Standard deviation of the spurious correlations
ΔI_{PNORM}	Normalized acoustic impedance variation (dimensionless)
$\boldsymbol{\Lambda}$	Matrix of eigenvalues of the modified low-rank representation of the covariance matrix of measurement errors
$\boldsymbol{\Pi}$	Projector that subtracts the mean and normalizes parameters or data matrices
$\boldsymbol{\Sigma}$	Matrix of the singular values of $\boldsymbol{S}$

Subscripts

c	Datapoint
i	Initial value
j	Ensemble-member
k	Local analysis group
l	Parameter elements in the local analysis group
n	Points in a vector of data or parameters
N	Indicates that the size is equal to the ensemble size
N_Λ	Indicates that the size is equal to the number of remaining eigenvalues

Superscripts

a	Updated, also known as posterior or analysis result
f	Prior, also known as background
i	Iterations
obs	Observed data

sim	Simulated data
$true$	“True” value of the parameters

3.10 Acknowledgments

We conducted this work with the support of Libra Consortium (Petrobras, Shell Brasil, Total, CNOOC, CNPC, and PPSA) within the ANP R&D tax as “commitment to research and development investments” and Energi Simulation. We also acknowledge the support from the Research Council of Norway through the Petromaks-2 project DIGIRES (RCN no. 280473) and the industrial partners AkerBP, Wintershall DEA, Vår Energi, Petrobras, Equinor, Lundin, and Neptune Energy. The authors are grateful for the support of the Center of Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil), Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil), and NORCE Norwegian Research Centre. Also, a special thanks to CMG and Schlumberger for software licenses. The authors would also like to thank Patrick Raanes and Xiaodong Luo for the technical discussions that contributed to the enrichment of this work.

4 IMPROVING FLUID MODELING REPRESENTATION FOR SEISMIC DATA ASSIMILATION IN COMPOSITIONAL RESERVOIR SIMULATION

Authors

Gilson M. Silva Neto

Victor de Souza Rios

Alessandra Davolio

Denis J. Schiozer

4.1 Abstract

There is a growing interest in applying quantitative methods to adjust reservoir flow models using time-lapse seismic data. The most common approach relies on a petroelastic model to convert the flow simulator outputs into acoustic impedance. The comparison of this simulated data with the observed time-lapse seismic anomalies enables the computation of changes in the reservoir models' parameters, reducing the uncertainty, and improving the reservoir characterization. Among other properties, the petroelastic model requires fluid models capable of forecasting the speed of sound. This fact becomes more challenging when the oil is volatile and contains a significant amount of CO_2, which is the case in some reservoirs in the Brazilian pre-salt region. In this situation, some classical models fail to predict the speed of sound in the oil phase within reasonable accuracy. Other models require testing for specific fluids or are not conveniently build to the integration with actual compositional reservoir simulators. Therefore, we propose the application of a calibrated cubic equation of state to represent the fluid behavior for both reservoir flow and petroelastic simulations. For this purpose, we describe a methodology in which the fluid model is progressively adjusted using reservoir engineering and speed of sound experimental data, depending on the available information. We applied this methodology using the well-known Peng-Robinson equation, but similar results could be obtained with other models of this class. We show that the match to the conventional pressure-volume-temperature data, a common practice in reservoir engineering,

can be enough to generate fluid models capable of forecasting the speed of sound. Furthermore, the speed of sound experimental data can improve fluid characterization without jeopardizing the previous fitted experiments. We tested our methodology with experimental data of a fluid of one reservoir in the Brazilian pre-salt region. Moreover, we compared the results obtained in the equation of state with other published correlations and simplified models. Synthetic reservoir models with different production strategies were applied in these comparisons.

Abbreviations:

AAD – average absolute deviation

AADS – average absolute deviation with sign

EOS – equation of state

GOR – gas-oil ratio

MMP – minimum miscibility pressure

PEM – petroelastic model

PR – Peng-Robinson

PVT – pressure-volume-temperature

SRK – Soave-Redlich-Kwong

Keywords:

Speed of sound; cubic equation of state; hydrocarbon fluid; petroelastic model

4.2 Introduction

4.2.1 Motivation

The reservoir flow models are a vital tool for the decision-making process in oil production projects. Different kinds of data are used to calibrate these models and reduce the uncertainties in the production forecasts (OLIVER, D. S.; REYNOLDS; LIU, 2008).

There is a growing interest in applying time-lapse seismic quantitatively to adjust the reservoir models (DA NÓBREGA; DE MORAES; EMERICK, 2018; EMERICK, Alexandre A., 2016; LE RAVALEC *et al.*, 2012; LYGREN *et al.*, 2003; MACBETH; FLORICICH; SOLDO, 2006; STEPHEN, Karl D. *et al.*, 2006; YIN; FENG; MACBETH, 2019). In usual data assimilation workflows, the simulated and observed seismic data are compared in the same domain, and the deviations are used to correct the models' parameters. For this purpose, it is common to apply a petroelastic model to convert the pressure and saturation fields to seismic attributes, such as acoustic impedance (FAHIMUDDIN; AANONSEN; SKJERVHEIM, 2010b; GOSSELIN *et al.*, 2003; SAGITOV; STEPHEN, 2013). On the other hand, the seismic signal is inverted to the same domain, enabling a direct comparison.

In addition to the mineral's distribution and rock properties, the information about the fluid properties, namely the density and the speed of sound or bulk modulus, is necessary to build a petroelastic model capable of predicting the impedance changes caused by the fluid movements inside the porous media (AVSETH; MUKERJI; MAVKO, 2005; MAVKO; MUKERJI; DVORKIN, 2009). Different correlations for the speed of sound in the oil phase were proposed for application in these models. Batzle and Wang (1992) proposed several correlations for the elastic properties of fluids, which are widely applied in time-lapse seismic analysis. Nevertheless, they were not developed for volatile oils with significant amounts of dissolved CO_2. Furthermore, Han et al. (2012, 2013) presented an empirical correlation for the speed of sound in hydrocarbon liquids with high CO_2 content. However, their correlation was not conveniently built to integrate with compositional reservoir simulators, and they require testing for application with specific oils.

Projects in some of the main reservoirs in the Brazilian pre-salt area often produce oils with large amounts of dissolved CO_2 and natural gas (DE SANT´ANNA PIZARRO; BRANCO, 2012; PETROBRAS, 2015). The classical models, such as the Batzle-Wang correlation, are not suited for this kind of fluid (ALTUNDAS, Bilgin *et al.*, 2017; TAHANI, 2012). Furthermore, in this situation, the use of compositional reservoir simulation is usually required. In such formulation, cubic equations of state are traditionally applied to represent the fluid behavior.

Since the famous van der Walls equation (1873), other authors proposed different cubic equations, mainly intending to improve liquid-phase molar volume predictions. Redlich and Kwong (1949) presented the first cubic equation with a global use and was considered by many authors the first modern equation of state, also known as RK. Soave (1972) proposed a modification on the original RK equation to improve the accuracy of pure-component vapor pressures, this new equation is usually referred to as Soave-Redlich-Kwong, or SRK equation. Peng and Robinson (1976) proposed their equation of state intending to correct the liquid-phase densities predicted using SRK, which are, in general, low. In 1978 they presented a modification of the original equation, which is the most frequently used equation of state, including in compositional reservoir simulation (ROBINSON; PENG, 1978).

Previous studies address the application of the Peng-Robinson equation of state to predict the speed of sound in the oil phase (DARIDON *et al.*, 1998; SALIMI; BAHRAMIAN, 2014; TAHANI, 2012). However, difficulties have been reported to represent the behavior of this variable in liquids, especially in heavy oils, (PICARD; BISHNOI, 1987; SALIMI; BAHRAMIAN, 2014). Some of these issues could be alleviated if experimental or field data were used to calibrate the EOS (equation of state) parameters. Furthermore, other equations of state, especially the SAFT-BACK, have achieved better performance in predicting the speed of sound in liquids (SALIMI; BAHRAMIAN, 2014; TAHANI, 2012). Nevertheless, they were not tested in this work, because this kind of EOS is still absent in popular reservoir engineering software, hindering field applications.

Other researches focus on the use of the speed of sound experimental data to calibrate the equation of state (SHABANI; RIAZI; SHABAN, 1998; YE *et al.*, 1991). Since there is a level of uncertainty in the EOS parameters of multicomponent mixtures, calibration using experimental data of the oil is usually necessary (PEDERSEN; CHRISTENSEN; SHAIKH, 2015). Here, we consider the speed of sound data as one of the sources of information to characterize the fluid for both reservoir engineering and seismic applications.

To our knowledge, there is no published work addressing the sequential improvements of the EOS using specific field data, full pressure-volume-temperature (PVT) data, and speed of sound experimental data for the reservoir flow simulation and acoustic impedance changes. One noteworthy effort was performed by Barreau et al. (1997), who tuned the Peng-Robinson EOS using part of experimental data on a condensate gas, and the authors were able to predict the speed of sound values with an average deviation of 2.8%.

In this work, we apply the same Peng-Robinson cubic EOS in both the reservoir and the petroelastic models. We show that, following a classical reservoir engineering workflow, it is possible to obtain an EOS capable of predicting the speed of sound in the oil phase with an accuracy comparable to some of the best results reported in the literature for this kind of fluid. Our results, regarding an oil similar to a specific pre-salt reservoir, demonstrate that it is possible to further improve the EOS by calibrating its' parameters using experimental speed of sound data, without impairing the PVT and specific field data matches. In our case, the speed of sound data act as supplementary information to the EOS calibration process. Therefore, we propose a straightforward methodology to develop a fluid model based on a cubic EOS to be applied to both reservoir flow and acoustic impedance simulations.

We address the practical implementation of the De-hua Han correlation and the Peng-Robinson EOS in a reservoir engineering workflow using commercial tools (CMG, 2017, 2015). Some simplified models were also tested. Since the Peng-Robinson EOS is already implemented in the commercial flow simulators, the rigorous equations may be readily coupled with their outputs to represent the impedance changes, due to the fluids' movement in the porous media. We compared the EOS' results with experimental data and other models, using simple synthetic reservoir models.

4.2.2 Objectives

The objectives of this work are:

- Show that it is possible to use the EOS model obtained after the conventional reservoir engineering calibration in the petroelastic models of pre-salt projects.
- Show that it is possible to improve the EOS by calibration with experimental speed of sound data, without impairing other data matches.
- Test the application in simplified models.
- Propose a practical methodology to simulate both the fluid flow and the impedance changes in the reservoir, using commercial tools and the Peng-Robinson EOS, calibrated to the available data in pre-salt projects.

4.2.3 Organization

The remainder of the text is organized as follows. The second section presents, in brief, the methods that we tested for computing the speed of sound in the oil phase. In the third section, we explain our methodology to model the fluid for both reservoir flow and petroelastic simulations. The fourth section includes information about the fluid that we studied and the available data. Due to confidentiality clauses, we show limited information about the latter. Section five contains a description of our test cases, whereas section six presents our results and some discussions. The conclusions of this work are in section seven. A list of symbols is shown after that.

4.3 Methods to represent the speed of sound in oil phase

We analyzed three of the available methods to represent the speed of sound in the oil phase. The first was the direct derivation of the speed of sound using the Peng-Robinson EOS (PENG; ROBINSON, 1976; ROBINSON; PENG, 1978). The second was the De-hua Han correlation (HAN; SUN; LIU, 2012, 2013), based on experimental data using fluids that are similar to the one applied in this work. The last was the Batzle-Wang correlation (BATZLE; WANG, 1992), one of the most popular models used to simulate the fluid behavior in time-lapse seismic studies (AVSETH; MUKERJI; MAVKO, 2005).

One could adopt some simplifications of the previously cited models, aiming at facilitating their implementation based on the outputs of the compositional reservoir simulator, at the cost of increased deviations. The general hypothesis to simplify the models is that the compositional changes in the porous fluids will not impact their behavior significantly during the simulation. By assuming that, some variables that would demand extra computations become approximately constant. The drawback will occur when the reservoir fluids' composition changes dramatically. We test two possible simplifications: the constant γ formulation and the constant gas content in De-hua Han correlation. We present more information about these models in sections 4.3.2 and 4.3.4, respectively. It is valid to highlight that the use of these simplifications is optional, and the full-fledged models should be applied when the accuracy of the simplified versions is not confirmed to the particular case.

We describe the formulation of the applied methods in the following subsections. We do not intend to provide a comprehensive explanation of each technique. Only the general aspects and equations will be shown to provide a basis for specific discussions. The definition of each symbol is presented at the end of the manuscript.

4.3.1 Peng-Robinson EOS

In this section, we show the equation of state formulation we have applied to obtain the speed of sound forecasts. We chose to work with the Peng-Robinson EOS (PENG; ROBINSON, 1976; ROBINSON; PENG, 1978), but similar results should be accomplished using other cubic equations, such as Soave–Redlich–Kwong (SOAVE, 1972).

The Peng-Robinson equation of state considering volume translation (PÉNELOUX; RAUZY; FRÉZE, 1982) is shown in equation (4.1).

$$P = \frac{RT}{(\bar{V}+c-b)} - \frac{a\left[1+m\left(1-\frac{T^{0.5}}{T_c^{0.5}}\right)\right]^2}{(\bar{V}+c)(\bar{V}+c+b)+b(\bar{V}+c-b)} \tag{4.1}$$

Each parameter in the equation is defined for a component in equations (4.2), (4.3), (4.4), and (4.5).

$$a = 0.45724\frac{(RT_c)^2}{P_c} \tag{4.2}$$

$$b = 0.0778\frac{RT_c}{P_c} \tag{4.3}$$

$$m = 0.37646 + 1.54226\omega - 0.26992\omega^2 \quad ; \quad \omega \leq 0.49 \tag{4.4}$$

$$m = 0.379642 + 1.485030\omega - 0.164423\omega^2 + 0.0116666\omega^3 \quad ; \quad \omega > 0.49 \tag{4.5}$$

The speed of sound in the fluid is obtained using equation (4.6).

$$v_{PO} = \sqrt{\frac{C_P}{C_V \rho c_T}} \tag{4.6}$$

The specific heat ratio, γ is defined as (4.7).

$$\gamma = \frac{C_P}{C_V} = 1 - \frac{T}{C_V}\frac{\left(\frac{\partial P(T,\bar{V})}{\partial T}\right)_{\bar{V}}^2}{\left(\frac{\partial P(T,\bar{V})}{\partial \bar{V}}\right)_T} \tag{4.7}$$

The heat capacity at constant volume is obtained from (4.8).

$$C_V = \left(\frac{\partial H^{id}}{\partial T}\right)_P - R - \frac{ma(1+m)}{4\sqrt{2}T^{0.5}T_c^{0.5}b}\ln\left\{\frac{\bar{V}-\left[\left(-1+\sqrt{2}\right)b-c\right]}{\bar{V}-\left[\left(-1-\sqrt{2}\right)b-c\right]}\right\} \tag{4.8}$$

Since we represent the oil by a mixture of pseudo-components, we must apply mixing rules (PEDERSEN; CHRISTENSEN; SHAIKH, 2015) to compute the parameters of the equations (4.1), (4.6), (4.7) and (4.8). Note that the critical properties and molar weights of the pseudo-components and the interaction coefficients of the mixing rules exhibit a level of uncertainty. This uncertainty justifies the calibration of the model using experimental and field data.

4.3.2 Constant $\boldsymbol{\gamma}$ simplification

To avoid computing equations (4.7) and (4.8), in practical applications, one could adopt the simplifying assumption that the specific heat ratio, γ, is constant. Thus, the two variables of equation (4.6), namely the oil density and the oil compressibility at a constant temperature, are the only necessary information to obtain the speed of sound in the corresponding phase for the whole reservoir grid. This assumption greatly simplifies the computations based on commercial reservoir simulators, since ρ and c_T are two common outputs to these programs. However, it is well known that γ will vary, especially when the composition of the fluid changes in the porous media. We show the impact of this on the time-lapse seismic response in section 4.8.1.3.

4.3.3 De-hua Han correlation

Han, Sun and Liu (2012) measured the changes in the density and speed of sound in the fluid caused by CO_2 in a mixture with oil over a pressure range of $20MPa \leq P \leq 100MPa$ and a temperature range of $40°C \leq T \leq 100°C$. They proposed in (HAN; SUN; LIU, 2013) an empirical model for the speed of sound that is represented by the equation (4.9).

$$
\begin{aligned}
v_{PO} &= v_{oil+gas} - \Delta v_{oil+CO_2} + C_{oil+gas+CO_2} \\
&= v_{oil+gas} - \left(v_{dead\ oil} - v_{oil+CO_2}\right) + C_{oil+gas+CO_2}
\end{aligned}
\tag{4.9}
$$

In equation (4.9), $v_{oil+gas}$ is the speed of sound in the oil with dissolved hydrocarbon gas, $v_{dead\ oil}$ pertains to the gas-free oil, v_{oil+CO_2} pertains to the oil with dissolved CO_2 and $C_{oil+gas+CO_2}$ is a correlation variable defined in equation (4.10).

$$
\begin{aligned}
C_{oil+gas+CO_2} &= C_{\rho_0} C_{GOR} \\
&= (1.6457\rho_0 - 1.3174)\left[\frac{GOR_{CO_2}}{600}\left(1 + \frac{500}{GOR_{CO_2} + 1}\right)\right]
\end{aligned}
\tag{4.10}
$$

Among the necessary data to compute the variables in equations (4.9) and (4.10), we highlight the Gas-Oil Ratio (GOR), the CO2-Oil ratio, the specific gravity of oil, and the specific gravity of hydrocarbon gas. These properties change with composition and are not necessarily computed in the compositional simulation of fluid flow in the reservoir. This fact may add some difficulties in coupling its results with the petroelastic model built based on the so-called De-hua Han method.

4.3.4 Constant gas content and composition in De-hua Han method simplification

Aiming at avoiding the tedious external computations that may be necessary to use the De-hua Han method with commercial compositional reservoir simulation tools, one could assume the simplifying hypothesis of constant gas content and composition in the reservoir. Thus, the GOR and specific gravities could be estimated at a particular condition and would remain unchanged at different times and positions. Nevertheless, when the composition of the oil changes in the reservoir, the amounts of gas in solution and its properties will vary. We show the impact of this model imperfection in the test of section 4.8.1.3.

4.3.5 Batzle-Wang equations

Batzle and Wang (1992) proposed correlations to compute the elastic properties of fluids that are widely applied in time-lapse seismic studies. Here, we use their method to estimate the speed of sound in the oil phase to compare with the other formulations and the experimental data. In this work, we call it the Batzle-Wang model.

Accordingly to Batzle and Wang (1992), the speed of sound in the oil phase can be computed using equation (4.11), where the pressure is in MPa, and the temperature is in °C.

$$\begin{aligned} v_{PO} = {} & 2096\left(\frac{\rho'}{2.6-\rho'}\right)^{0.5} - 3.7T + 4.64P \\ & + 0.0115\left[4.12\left(\frac{1.08}{\rho'}-1\right)^{0.5} - 1\right]T \times P \end{aligned} \tag{4.11}$$

The so-called pseudo-density, ρ', is a correlation variable defined in equation (4.12), where ρ_0 is the specific gravity of the oil, B_o is its formation volume factor, and R_s is its gas solution ratio (STANDING, 1962; apud BATZLE; WANG, 1992), which are dimensionless. These variables also change with composition and require an additional external calculation to adapt the output from the compositional reservoir simulator to the petroelastic model.

$$\rho' = \frac{\rho_0}{B_o}(1 + 0.001R_s)^{-1} \tag{4.12}$$

4.4 Fluid characterization methodology

Our methodology has a general aim of defining a single equation of state to be used in the reservoir flow simulation and to represent the speed of sound in the hydrocarbons. We focused our analysis on the oil phase because it is considered more challenging to characterize using this kind of model when compared to the gas phase in the porous media. The methodology comprises four steps, namely: equation of state and pseudo-components choice, data assimilation, PVT and flow behavior analysis, and speed of sound simulation analysis. The general methodology is shown in Figure 4.1. Each of these steps is described in the following subsections.

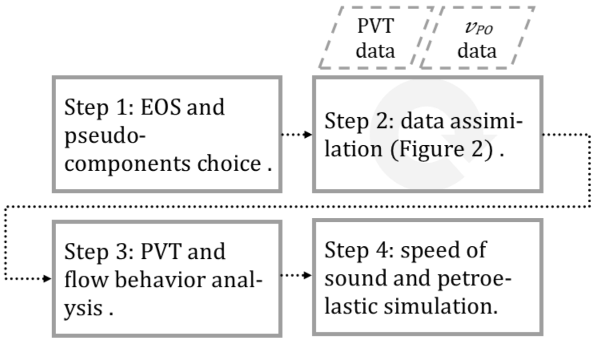


Figure 4.1: General fluid modeling methodology. The equation of state and the pseudo-components are chosen in step 1. The data assimilation process is performed in step 2. The PVT and flow behavior are simulated and analyzed in step 3. The speed of sound and petroelastic simulations are done in step 4.

4.4.1 Equation of state and pseudo-components choice

A proper phase behavior prediction of oil and gas mixtures is an important task for effectively characterizing the reservoir fluids. In the petroleum industry, most of the calculations of pressure-volume-temperature (PVT) relation are based on Peng-Robinson cubic equation of state with Péneloux's volume translation. However, the commercial tools commonly provide options to work with other cubic equations, such as the SRK equation. It may be considered equivalent to the former for reservoir engineering purposes.

A reservoir fluid is a complex mixture of hydrocarbons and other components, such as N_2, CO_2, and H_2S and its complete characterization can consist of more than 80 components and pseudo-components (PEDERSEN; CHRISTENSEN; SHAIKH, 2015). Thus, to perform phase equilibrium calculations it is necessary to reduce this number of components. This process is usually known as lumping or pseudoization and requires a proper decision of how to group the components. In compositional reservoir engineering simulation studies, the lumping procedure is highly dependent on the dynamic evaluations to be performed and a balance between computational performance and fluid representation is mandatory.

In general, when gas is injected into the reservoir, a complex mass transfer occurs between oil and gas. Thus, the pseudoization process needs to be carried out without oversimplification of the light components, since this can jeopardize the expected interaction between injected gas and reservoir oil. On the other hand, if an immiscible fluid is injected and the reservoir pressure can be maintained above the saturation pressure, a more aggressive lumping scheme can be adopted for the same reservoir fluid.

As a practical workflow for establishing an effective fluid pseudoization, we can begin with an aggressive simplification as a starting point, which represents the maximum performance solution for the problem. Then, after checking the PVT experimental results consistency, numerical reservoir simulation in the same conditions, to be expected in the development plan, can be conducted. In sequence, a more moderate lumping scheme can be tested, and the same checks performed until no significant changes in the numerical results are observed. By doing so, we can choose the number of pseudo-components that tends to balance both performance and compositional representation.

4.4.2 Data assimilation

In reservoir engineering applications, the equation of state is usually calibrated using experimental and field data before it is incorporated into flow simulators. During an oil field production life, increasing amounts of information become available and can be used to reduce the uncertainty in the fluid models. In this study, we consider the application of the EOS with four levels of accuracy: without experimental information, isolated information, full PVT analysis, and full PVT with speed of sound data. The three former cases are well established in reservoir engineering practice, and we briefly cite the definition and calibration processes. In our methodology, the analysis of the four cases is useful to determine at which point the fluid model is already suitable for speed of sound calculations.

In the former case, while no experimental data is ready, tabulated data available in the literature (FIROOZABADI, 1999) are used to compute the EOS parameters. After that, in the initial stages of a field study, including the exploratory phase, some isolated information regarding the fluid behavior may become available before the full PVT analysis is performed. These data include saturation pressure and GOR, which may be used to improve the fluid model. Later, a full PVT dataset is ready to use. This is commonly available in the development and management phases of oil production projects. Therefore, in a conventional reservoir engineering workflow, the EOS is calibrated with this amount of information.

The reservoir engineering software provides tools to use the data mentioned so far to calibrate the equation of state parameters. The engineer must select the uncertain variables to determine, which usually include binary interaction coefficients and the properties of the heavier pseudo-components, such as molar weights and critical points. After selecting the parameters, the commercial tool performs a multivariable regression. Then, before the model is considered acceptable, the engineer checks the match's quality and the admissibility of the changes. For more details about this process, please refer to (PEDERSEN; CHRISTENSEN; SHAIKH, 2015).

It is possible to improve the fluid models, especially for time-lapse seismic studies, using speed of sound experimental data. In a reservoir engineering perspective, when compared to PVT data, this information is uncommon. However, we show that it can be applied to improve the speed of sound predictions and reduce the uncertainty in the EOS parameters. Since the speed of sound is not a standard data, this calibration is not included in the reservoir engineering commercial tools. Therefore, we propose to perform this calibration in the iterative process shown in Figure 4.2 and described by the following steps:

1. Generate multiple EOS varying the uncertain parameters.
2. Calibrate each new EOS using the available PVT data. Since this is an inverse problem, this will lead to a set of different EOS that reasonably matches the experimental data.
3. For each model, simulate the speed of sound experiment.
4. The simulations of the speed of sound are compared with the speed of sound observed data. If the error level is ok, store the best EOS and end the process. If not, proceed to step 5.
5. Select the uncertain parameters with the highest variance and the highest correlation with the speed of sound results.
6. Estimate the parameter values to set the equation (A.2) to zero and go back to step 2.

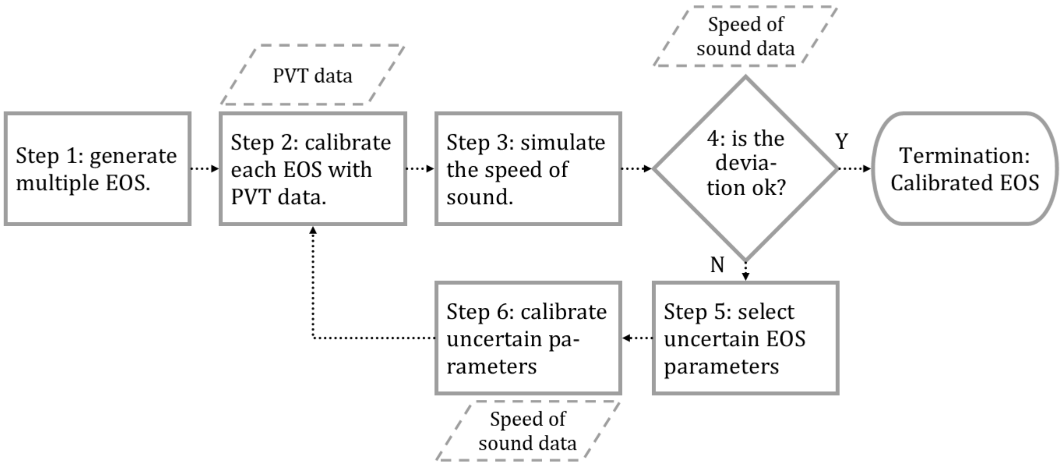


Figure 4.2: Process of calibration with speed of sound data. An ensemble of EOS with different initial parameters are generated in step 1.

Each EOS is calibrated using PVT data in step 2. These EOS are used to simulate the speed of sound in step 3. The previous results are compared with the speed of sound data in step 4. If the error is within the tolerance range for at least one EOS, the process is finished. If not, the uncertain parameters of the EOS are selected in step 5. These parameters are calibrated using the speed of sound data in step 6. The process then returns to step 2.

$$AADS = \frac{1}{N_d}\left(\sum_{i=1}^{N_d}\frac{\left|d_{sim,i} - d_{obs,i}\right|}{d_{obs,i}}\right) \times \left[\frac{\sum_{i=1}^{N_d}\left(d_{sim,i} - d_{obs,i}\right)}{\sum_{i=1}^{N_d}\left|d_{sim,i} - d_{obs,i}\right|}\right] \tag{4.13}$$

Note that the definition of Average Absolute Deviation with Sign, AADS, in equation (A.2) is very similar to AAD (Average Absolute Deviation), which is commonly used to quantify the deviation between the model and the experimental data. However, we opted to differentiate models that tend to overestimate or underestimate the speed of sound in the oil phase, by setting positive or negative values to AADS, respectively. Moreover, by using AADS, we highlight the close to linear behavior of our problem in the working range. Nevertheless, one could opt to minimize the AAD, which is equivalent to setting the AADS to approximately zero, and perhaps this is the more conventional way to formulate the problem.

The objective of the workflow shown in Figure 4.2 is to generate a fluid model that is capable of representing all the available data: the full PVT data set and the speed of sound measurements. A direct multivariate regression could be applied to obtain models calibrated to these data. Nevertheless, we opted to adapt the methodology to use commercial tools that lack this option because we believe that this would facilitate immediate field applications.

Once the calibrated EOS is available, the fluid model can be tested in both the reservoir flow simulator and the petroelastic model. These tests will be covered in the following subsections.

4.4.3 PVT and flow behavior analysis

After the fluid model calibration, it is important to check the reservoir production forecast behavior. The specific objective of this step is to detect significant differences in fluid behavior and analyze their plausibility. If this is part of a practical project, the reservoir flow model may be used as a framework to test the fluid model behavior. One way to perform this is to analyze the oil, gas, and water production forecasts using each fluid model. If there is production data (history), it may be used to validate the models. If the equations of state were calibrated using PVT or PVT with speed of sound data, one should expect low model deviations between the model and the field data, if the errors of the reservoir model are neglectable. However, major differences may be observed due to fluid models calibrated using only isolated data or no data at all.

When the development plan project considers miscible gas injection as a recovery method, it is crucial to check the miscibility behavior, apart from the regular PVT information. For this evaluation, the Minimum Miscibility Pressure (MMP) is an important parameter. This is the lowest pressure, at a fixed temperature, in which gas and oil in contact with each other achieve miscibility, no matter the proportion.

Numerical evaluations were carried out reproducing the slimtube experiment (RIOS, VS; SANTOS; ESPÓSITO, 2016) to investigate the different fluid models regarding miscible behavior. There is no universally accepted method to define the MMP from slimtube experiments, but a widely adopted criterion for determining the MMP is the construction of a graph of oil recovery factor after the injection of 1.2 pore volumes (PV) of gas against the injection pressure. The MMP is the pressure value at which oil recovery factor reaches 90–95% (RIOS, V.S. *et al.*, 2019).

After validating the fluid model in the flow simulation, it is ready to be used to compute speed of sound and acoustic impedance variations, which will be addressed in the following subsection.

4.4.4 Speed of sound and acoustic impedance simulation

At this step, the PEM is used to compute the compressional wave velocity and acoustic impedance changes in the reservoir, due to the pressure changes and fluid movement.

In this work, for comparison and validation purposes, we performed this step using different models: EOS, De-hua Han, simplified De-hua Han, and constant γ. The Batzle-Wang model was not applied in impedance simulations, because its results in the initial speed of sound tests were not satisfactory.

In the following subsections, we briefly describe how each speed of sound method can be computed from the output of the reservoir simulator. Notice that we used WinProp, Fluid Property Characterization Tool version 2015.10 (CMG, 2015), and GEM, Compositional & Unconventional Simulator version 2017.10 (CMG, 2017), both from Computer Modelling Group Ltd. Nevertheless, equivalent simulation studies could be performed using other commercial tools.

4.4.4.1 Obtaining the speed of sound using Peng-Robinson EOS

To obtain the speed of sound directly from the cubic equation of state that is used in the reservoir simulation, it is possible to apply Equation (4.6). Both the fluid density, ρ, and the fluid compressibility at a constant temperature, c_T, are outputs of the reservoir simulator. However, the specific heat ratio γ must be calculated using Equation (4.7), whose derivatives may be obtained analytically from Equation (4.1). The heat capacity at constant volume is obtained using Equation (4.8), in which the ideal enthalpy derivative may be computed from the equation provided by the fluid simulator. Mixing rules are applied to obtain the parameters for each equation. Furthermore, unit conversions are usually needed.

4.4.4.2 Obtaining the speed of sound using De-hua Han correlation

It is necessary to estimate GOR and specific gravities of the fluids to compute the speed of sound using the De-hua Han correlation. Nevertheless, these are not variables necessarily calculated during a compositional reservoir simulation. It is possible to export these results from individual sectors of the reservoir. Still, it becomes time demanding to perform this in full-field reservoir simulations, particularly for giant models, reaching hundreds of thousands or millions of active cells. Another possible solution is to obtain these variables externally, by computing phase equilibria in the petroelastic model code, but this significantly increases its complexity. It is interesting to note that similar issues would be observed in the application of Batzle-Wang model from the output of compositional reservoir simulations.

In our applications, we only calculated the rigorous implementation of the De-hua Han correlation for isolated fluid simulations, without requiring the reading of reservoir grid properties. We managed to compute the speed of sound using the full-fledged De-hua Han correlation in a test to compare it with the experimental data. The results are discussed in section 4.8.1.1.

4.4.4.3 Obtaining the speed of sound using De-hua Han correlation with constant gas content and composition simplification

By considering a constant gas content and composition simplification, it is possible to compute the speed of sound using the De-hua Han correlation, assigning constant values to the GOR and specific gravities. For instance, one can use the initial conditions of the reservoir simulation, which are commonly known. The simplification dramatically simplifies the coupling of the results of the reservoir flow simulation to the petroelastic model. Still, it can incur higher deviations if compositional changes occur in the porous media.

4.4.5 Obtaining the speed of sound using the constant γ simplification

The direct application of Equation (4.6) is possible since all the terms are either considered constant or obtained from the outputs of the compositional simulation. The value of the γ can be estimated using a specific condition, for instance, the initial values in the reservoir.

4.5 Fluid information

We performed simulation tests of the fluid models using the available information about oil from one reservoir in the Brazilian pre-salt region. In this section, we briefly describe this fluid and available data.

4.5.1 General fluid characteristics

We applied our methodology to a fluid whose characteristics are presented in Table 4.1 and were obtained from the unclassified report (PETROBRAS, 2015). The main features of this fluid to our application is the volatility and high CO_2 content at reservoir conditions.

Table 4.1: Reservoir fluid properties at the initial conditions.

Property	Value
API	27°
GOR	415 L/L
Fluid temperature at reservoir conditions	90 °C
CO_2 content in gas phase	44%
CO_2 content in reservoir fluid	37%
Oil formation volume factor	≈ 2 L/L

4.5.2 Experimental data and methods

This study employed two experimental data sets provided by an oil field operator in Brazil. The first one comprises standard PVT measurements using reservoir oil samples. Classical PVT experiments, such as constant composition experiment, differential liberation and separator tests were performed. For details about these experiments, we refer to the third chapter of (PEDERSEN; CHRISTENSEN; SHAIKH, 2015).

The second data set comprised the speed of sound in the oil phase with dissolved gas, under pressure and temperature conditions similar to the reservoir. The test consists of measuring the transit times of a known ultrasound pulse in a cell, where the fluid is kept under different pressures and at the reservoir temperature. This study considers only data measured under pressure above the bubble point to avoid changing the amount of gas content in the oil phase and the influence of free gas in the experiment. For more details about this kind of experimental equipment and procedure, we refer to (WANG; NUR; BATZLE, 1988).

The experimental data used in this study are confidential. Therefore, the data are not presented, and the graphs where they appear have linearly transformed axes to vary in the interval [0,1]. All axes whose data have the same dimension, for instance, speed of sound, were transformed using the same linear relation.

4.6 Simulation studies

We performed two tests using the fluid models: one simulation of the speed of sound experiment and a 2D numerical example. In the latter, we were able to analyze the distortions on the time-lapse impedance variation caused by the fluid models. In order to compute the impedance values, a petroelastic model compatible with the information available regarding one pre-salt reservoir was built. In the following subsections, we provide some details about the reservoir flow model and the petroelastic model.

4.6.1 Case studies description

Our first case study was the simulation of the pressure and temperature conditions during the speed of sound experiment. Our objective was to determine the average absolute deviation of each method in representing the experimental data.

The second case study was a reservoir simulation, which was performed to evaluate possible distortions in the impedance variations estimation caused by the fluid models. In a field application, this analysis could be performed using a 3D reservoir model, representing the whole complexity of the project. Nevertheless, the objective of our test is to capture the impact of the fluid properties in the impedance variations and show it visually. Therefore, the utilization of a simplified 2D model will not affect the conclusions.

Therefore, we applied the 37°th layer of the widely known SPE10 problem (CHRISTIE; BLUNT, 2001) to our analysis and upscaled the model with a flow-based algorithm (CHRISTIE, 1996; KUMAR *et al.*, 1997), to obtain $15 \times 55 \times 1$ cells with the dimensions $24.348 \times 12.192 \times 0.6096\ m$. The model's porosity field is shown in Figure 4.3, where four producers are located at the corners, and one injector is at its center. The temperature and pressure conditions of the model were altered to represent similar conditions of the fluid in the pre-salt reservoir. Moreover, the relative permeability curves and rock compressibility were chosen to represent the behavior of a carbonate reservoir.

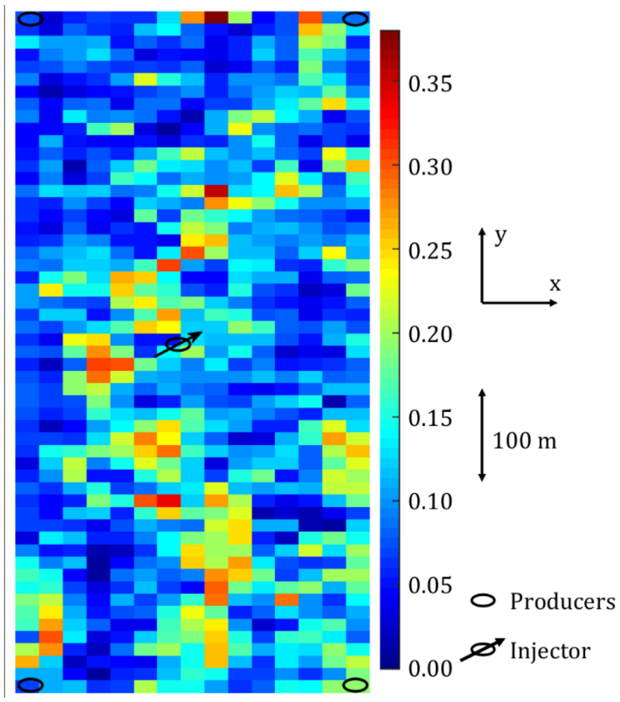


Figure 4.3: Porosity horizontal map and well locations of SPE10 2D model.

We used the model shown in Figure 4.3 to simulate three different production strategies: waterflood, gas reinjection, and simple depletion.

4.6.2 Petroelastic model

To perform numerical examples using different fluid models, we built a petroelastic model compatible with the information available regarding one pre-salt reservoir. It reads the data from the reservoir compositional flow simulator and computes the acoustic impedance and other seismic attributes at any simulation time. Our PEM used the average of the Hashin-

Shtrikman bounds (HASHIN; SHTRIKMAN, 1963; apud AVSETH; MUKERJI; MAVKO, 2005; MAVKO; MUKERJI; DVORKIN, 2009) to compute the effective elastic properties of the mixture of minerals. The dry rock properties as functions of porosity were modeled based on available well data. Furthermore, the pressure dependence was obtained from published results for similar reservoirs (COSTA *et al.*, 2016). Finally, we determined the saturated rock properties using the classical Gassmann equation (GASSMANN, 1951).

We acknowledge that measured data may diverge from the Gassmann theory. One of the reasons for this behavior is that Gassmann model assumes that the rock is relaxed within a half cycle of the seismic wave. The equilibrium assumption may not hold, mainly when one or more of the following conditions occur: tight porous media, rocks saturated with high viscosity fluids, and when ultrasound frequencies are applied (BA *et al.*, 2016, 2017; CHENG *et al.*, 2019; XU; PAYNE, 2009). Some models are aiming at handling this effect (BA *et al.*, 2017; XU; PAYNE, 2009). Other limitations to Gassmann's equations are homogeneous and isotropic medium with well-connected pores (MAVKO; MUKERJI; DVORKIN, 2009) and no rock-fluid chemical interactions. Nevertheless, there are results regarding the type of reservoir that we are investigating that support the application of the more straightforward Gassmann equation to model the time-lapse seismic response (SILVA *et al.*, 2020).

4.7 Results and discussions

In this section, we show the main results that were generated to support our study. They are presented in the order of our proposed methodology, showing the reader the steps that can be performed in a field application. However, we also included additional comparisons with other methods in order to validate our main contributions.

4.7.1 Equation of state and pseudo-components choice

The first step in a fluid characterization process is to select the equation of state to be considered in the phase behavior modeling. Then, the pseudo-components selection is a crucial step to build a compositional numerical model, considering the fluid characteristics and dynamic evaluations. In this step, the objective is to find a harmonic balance between computational performance and compositional representation.

As previously mentioned, we consider PR with Péneloux's volume translation as our cubic equation of state, since it is good at predicting both liquid and gas phases and is the most frequently used equation. Despite that, it is worth mentioning that our methodology can be equally applied to other cubic equations, such as SRK.

Our reservoir fluid is a volatile oil characterized by a high CO_2 fraction. Also, the development plan establishes that the totality of the produced gas needs to be reinjected into the reservoir. These two aspects provide important information for the selection of the pseudo-components.

The most aggressive lumping procedure needs to consider that it is suggested to make CO_2 and methane (usually combined with a small fraction of N_2) as single components since they are the most representative ones. The other light components up to C5 can be combined into one pseudo-component. The pseudo-components from C6 to C20+ (this heavy component can vary depending on the laboratory characterization) can be grouped in two. This configuration tends to be efficient in performance (since it presents only 5 components) and still preserves the capability of reproducing the experiments. On the other hand, if the injected fluid is changed and detailed control of the C2-C5 group is required, this approach can be overly simplified.

To overcome the limitation mentioned above, a new lumping scheme can be considered, in which the light components from C2 to C5 are grouped into two pseudo-components. Also, the medium and heavy components can be grouped into four pseudo-components. This division, with a total of eight pseudo-components, can be more flexible to predict phase behavior of a more detailed fluid composition but will have an impact on the numerical performance.

We started our analysis with the former lumping strategy and then verified the impact of increasing the complexity of the model in section 4.8.2.

4.7.2 Data assimilation

Using the simplest pseudo-components choice described in section 4.7.1, we started the data assimilation process. Since we wanted to compare the quality of the speed of sound forecasts in different situations, we performed the data assimilation in three stages. The EOS started with all parameters defined with tabulated values from the literature, which we call the EOS5CV0 case. Using this model, we assimilated specific field data, including oil saturation pressure at the reservoir temperature and the Flash GOR. This was performed using the commercial reservoir engineering software Winprop version 2015.10 (CMG, 2015) and the result is here called the EOS5CV1. In the subsequent stage, a full PVT dataset was assimilated using the same commercial software, leading to the EOS5CV2 case. This information content is a common situation in the development and management phases of a project. Finally, the EOS5CV3IT4 case was obtained using the previous information and speed of sound experimental data of the same oil, after four data assimilation iterations (see Figure 4.2). In Table 4.2, we show a list of the EOS cases and the information that was available at each step.

Table 4.2: Equation of state cases with 5 pseudo-components.

Model	Information
EOS5CV0	None
EOS5CV1	Saturation pressure and Flash GOR
EOS5CV2	Full PVT dataset
EOS5CV3IT4	Full PVT dataset and speed of sound experimental data

In Figure 4.4, we show the comparison of the last three fluid models, using 5 pseudo-components, in terms of PVT behavior. The EOS5CV0 results are not included in the figure, because they are too different from the calibrated models, and the fluid simulator failed to converge when calculating the experiments at some conditions. We noticed that the equation of state models from version 1 provides a reasonable representation of the fluid PVT behavior for reservoir engineering purposes. If a stricter evaluation is performed, one could point the exception of EOS5CV1, which incurs an AAD of 17% in forecasting the oil density, while EOS5CV2 and EOS5CV3IT4 exhibit 0.5% and 2.8% respectively. This points to the importance of using the full PVT data set to characterize the fluid. The EOS5CV1 still has high uncertainty, particularly in the properties of the heavy pseudo-components, which leads to higher deviation levels in the oil density.

The results in Figure 4.4 show that both the EOS5CV2 and the EOS5CV3IT4 reasonably represent the PVT behavior of the fluid. This suggests that, by applying the workflow presented in Figure 4.2, we were able to match both the PVT data and the speed of sound data. Therefore, the additional data assimilation did not jeopardize the classic fluid characterization process.

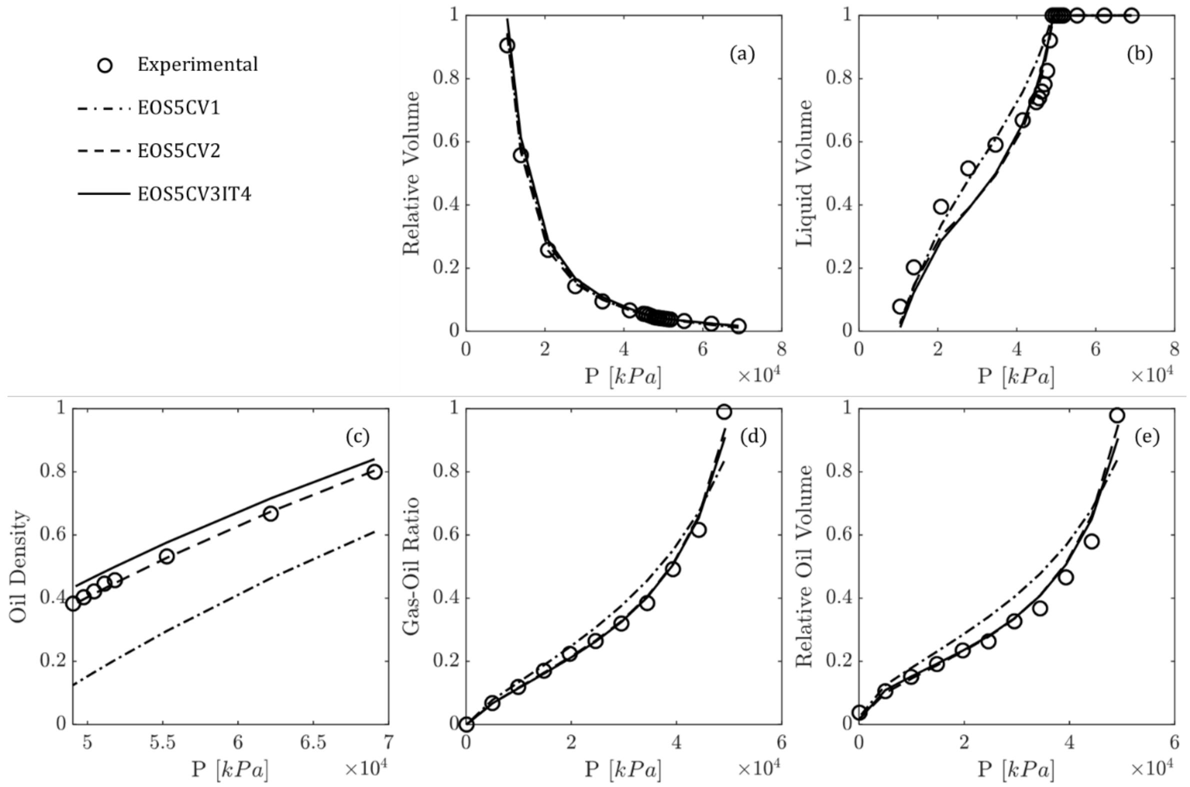


Figure 4.4: Volumetric behavior of the equations of state, based on different amounts of experimental data used in the calibration. All vertical axes are linearly transformed to vary in the interval [0,1] due to confidentiality clauses. Therefore, the vertical axes are dimensionless.

Figure 4.5 presents the results of the speed of sound experimental data assimilation. This was computed outside the commercial reservoir engineering tool because this data is rarely available for reservoir engineers, and the most known commercial tools lack options to deal with it. The parameters changed at this stage were chosen based on the variability after the PVT data assimilation and the correlation with the speed of sound values. Therefore, we selected the parameters whose uncertainty remained relatively high after the calibration using PVT data and that were correlated to the speed of sound. In our case, the chosen parameters were the binary interaction coefficients between the third and the first, the third and the second, the fifth and the first, and the fifth and the second pseudo-components.

From Figure 4.5, it is possible to notice that the relations between the four selected parameters and the speed of sound average deviation are almost linear. This facilitates the assimilation process, which reached the desired level of deviation after four iterations. The average relative deviation of the final model, EOS5CV3IT4, was 0.35%.

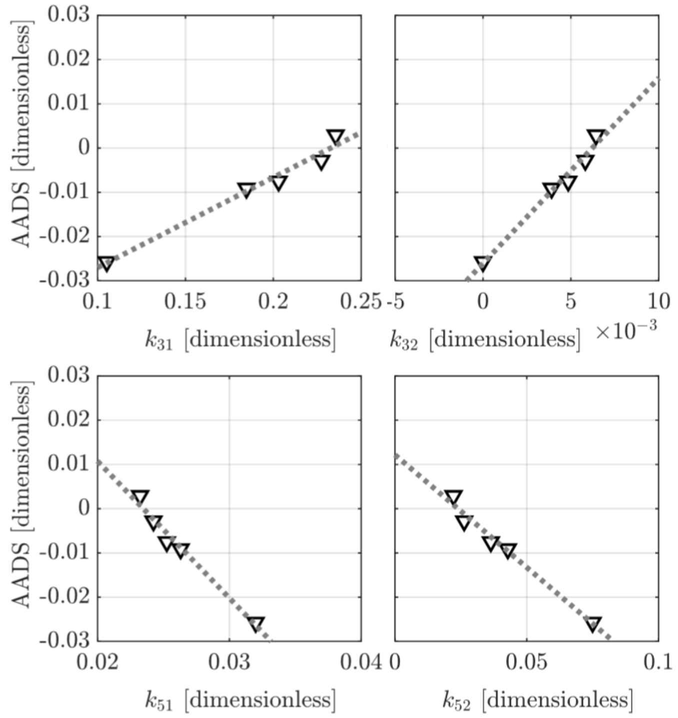


Figure 4.5: Equation of state parameters calibration using speed of sound experimental data.

The speed of sound variation with the pressure is shown in Figure 4.6. The results of the models are compared to the experimental data. The EOS5CV1 led to speed of sound values lower than the observed data, with an average absolute deviation of roughly 8%. Relatively high AAD values were expected for this model since only isolated information regarding saturation pressure and GOR were applied in the calibration process. This deviation decreases to values limited to 2.7% for EOS5CV2 when including a set of PVT experiments to estimate the parameters. Finally, after the iterations of speed of sound incorporation shown in Figure 4.6, the EOS5CV3IT4 exhibits better accuracy, with average absolute deviations of roughly 0.35%.

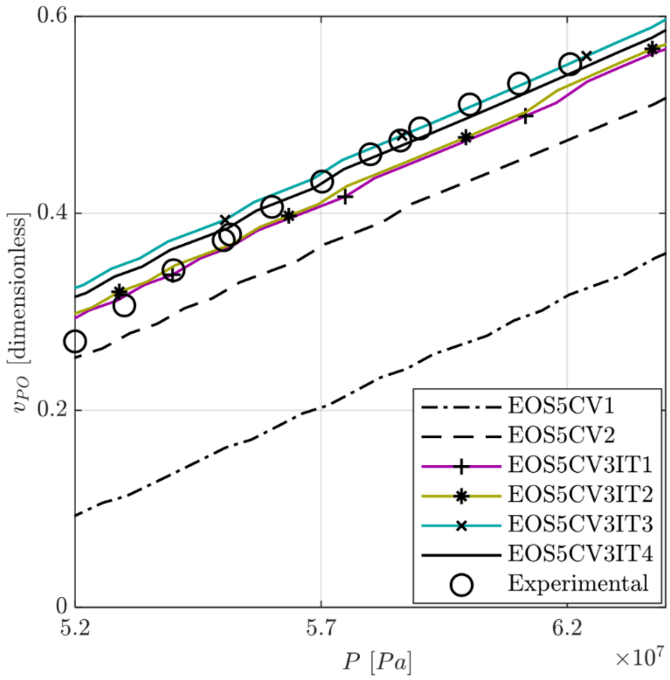


Figure 4.6: Speed of sound data match. The vertical axis is linearly transformed to stay in the interval [0,1] for confidentiality.

At the end of the data assimilation process, we considered the EOS5CV3IT4 to be our best model, using the Peng-Robinson formulation, in a sense that it represents our empirical knowledge about the fluid. In the next section, this model and the intermediate versions are compared in terms of production the forecast in a reservoir simulator.

4.8 PVT and flow behavior analysis

After performing the complete data assimilation process, it is important to check the fluid flow behavior in the reservoir and how it affects the production forecasts. Intending to capture different aspects of the fluid behavior, we performed three simulation examples: waterflood, gas injection, and depletion recovery processes. In all tests, we applied a modified version of the SPE-10 model described in section 4.6.1.

In Figure 4.7, we show the reservoir forecast in the waterflood experiment. We analyzed the cumulative oil production (Figure 4.7 a), the ratio between the gas and oil rates at surface conditions (Figure 4.7 b), the average reservoir pressure (Figure 4.7 c), and the cumulative water production (Figure 4.7 d). In this case, the well in the center injects water to maintain the pore-pressure. All the producers operate at a constant bottom-hole pressure of $50000\ kPa$, while the injector exhibits a bottom-hole pressure of roughly $78500\ kPa$ and a maximum water rate at surface conditions of $0.6\ m^3\ std\ /\ d$. Despite the EOS5CV0, the remaining models that were calibrated in the previous step exhibited similar behavior in terms of reservoir rates and pressure forecasts. Moreover, the EOS5CV1 presents minor deviations in the pressure forecast in comparison to the other models (Figure 4.7 c), due to the lack of information applied in the data assimilation process. The cause of this behavior was a slight difference in the compressibility of the fluid in the case where only isolated field data was incorporated.

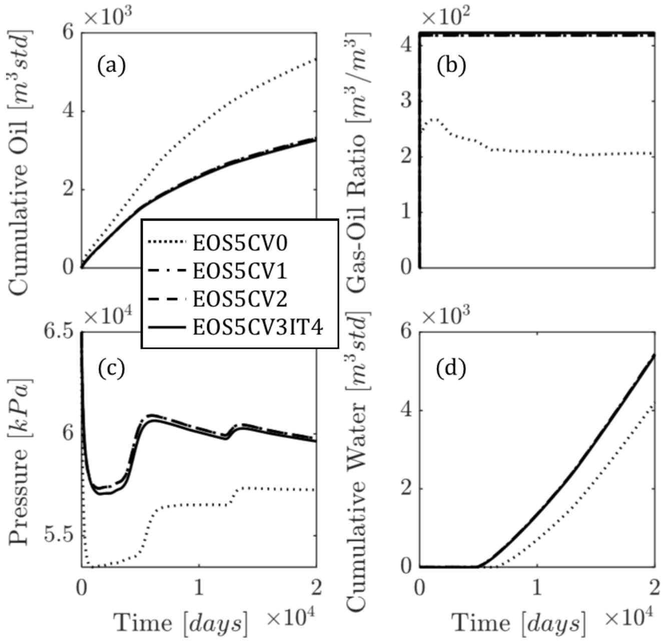


Figure 4.7: Reservoir forecast in the waterflood experiment.

In the following step, we performed a gas injection experiment to capture compositional changes and mass transfer interaction between the injected fluid and the reservoir fluid, as in the behavior of the injected fluid itself. The well in the center of the reservoir injected gas with the same composition of the produced gas and with a maximum bottom-hole pressure of roughly $78500\ kPa$.

The results of the gas injection simulation are shown in Figure 4.8, where the cumulative water production was omitted because it is nearly zero in this kind of simulation. Again, as expected, the EOS5CV0 exhibits significant differences when compared to the others because no calibration was computed in it. Moreover, we were able to notice higher deviations among the other models when compared to the waterflood experiment, especially in the GOR of EOS5CV1 (Figure 4.8 b). This was expected for this latter model because the oil saturation pressure and the initial GOR are not enough to characterize the oil and gas behavior in the reservoir.

Comparing the two remaining models, EOS5CV2 and EOS5CV3IT4, we see similar results, which indicates that the PVT dataset is enough to model the fluid behavior for reservoir flow simulation purposes. The only noticeable difference occurs in the average reservoir pressure (Figure 4.8 c), which motivated another comparison of these models in a simple depletion simulation.

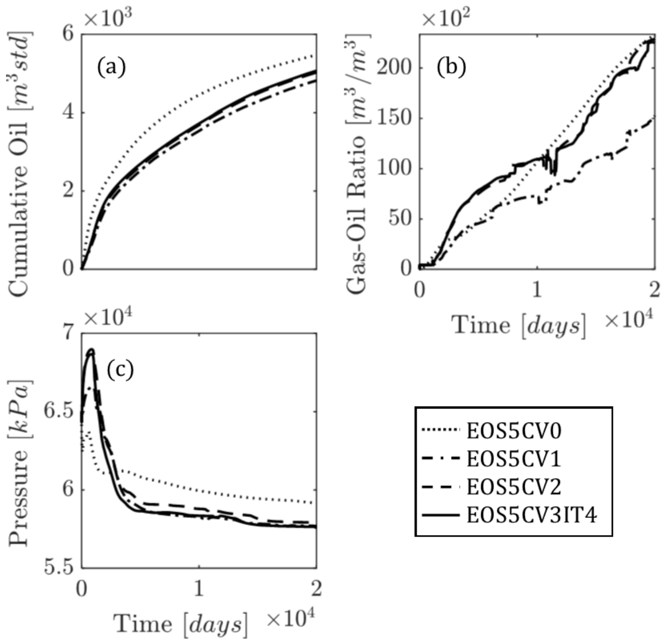


Figure 4.8: Reservoir forecast in the gas injection experiment.

In the following reservoir simulation test, we computed a simple depletion case, by closing the injector well located at the center of the reservoir. Since our objective at this point was to investigate the pressure differences observed in Figure 4.8 c, only EOS5CV2 and EOS5CV3IT4 simulation models were run. We found slight variations in the cumulative oil production (Figure 4.9 a), GOR (Figure 4.9 b), and average reservoir pressure (Figure 4.9 c). This was caused mainly because of the differences in the gas and oil compressibilities at a constant temperature, shown in Figure 4.9 d. This indicates that the speed of sound data may help calibrate the compressibilities that remain somewhat uncertain after the PVT data assimilation. In our case, the incorporation of the speed of sound experimental data reduced the oil compressibility at constant temperature and increased the gas compressibility at a constant temperature, when compared to the previous model. Therefore, the average reservoir pressure remained slightly lower when using EOS5CV3IT4 at the early times of production, while the porous media is saturated mainly with oil. At later times, when more gas appears inside the reservoir, the pressure in the simulation with EOS5CV3IT4 remains slightly higher than in the simulation with EOS5CV2. It is worth mentioning that this was expected because the pressure decline rate is inversely proportional to the overall reservoir compressibility. Nevertheless, despite it being considered beneficial to the reservoir characterization process, this compressibility change may be secondary in a production forecast point of view, because the effects are minor.

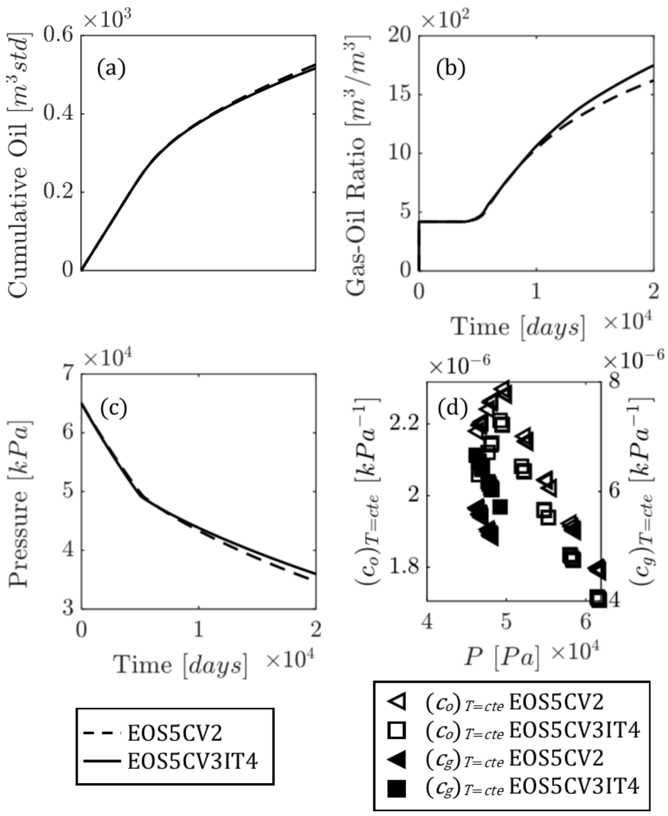


Figure 4.9: Reservoir pressure and compressibility behavior in a depletion experiment.

In projects that apply miscible recovery strategies, it is essential to check if the fluid model represents the miscibility conditions adequately. Therefore, in the final simulation test, we analyzed the miscibility between the injected gas and the reservoir oil in a slimtube test. The results are highlighted in Figure 4.10, where a classical slimtube plot is presented. Again, EOS5CV0 was not considered due to its high distortion in the results. We can observe that all three models presented MMP in the investigated pressure range. However, EOS5CV1 predicted a lower MMP. EOS5CV2 and EOS5CV3IT4 showed similar results, as also observed in the previous investigations. It is essential to highlight, however, that experimental slimtube results would be necessary to allow a more accurate and quantitative evaluation.

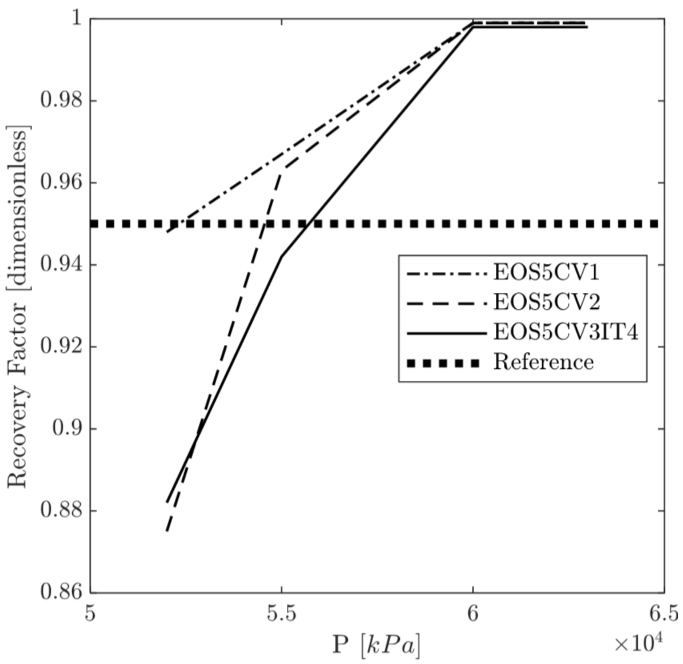


Figure 4.10: Miscibility simulation using different EOS versions.

The results reported in this subsection indicates that the two last models, namely the EOS5CV2 and EOS5CV3IT4 produce similar results when used in a reservoir flow simulator. Moreover, since they also reasonably match the PVT data in Figure 4.4, and the speed of sound experimental data in Figure 4.6, both are appropriate for reservoir production forecast and time-lapse seismic analysis. The results of this latter application will be addressed in the next subsection.

4.8.1 Speed of sound and acoustic impedance simulation

In this section, three simulations are presented to illustrate the application of the proposed fluid model in time-lapse seismic analysis. We compare the results obtained with the EOS5CV2 and the EOS5CV3IT4 models to two correlations from the literature: the classical Batzle-Wang – BW correlation (BATZLE; WANG, 1992) and the De-hua Han correlation (HAN; SUN; LIU, 2012, 2013). In the first simulation, the speed of sound experiment was reproduced to compare the models with the observed data. In the second and third, time-lapse seismic data from the waterflood and the gas injection applications were computed, respectively. The SPE-10 model, described in section 4.6.1, was used in these two latter tests.

4.8.1.1 Pressure dependence experiment

We simulated the speed of sound in the oil phase experiment using the four models: Batzle-Wang, De-hua Han, EOS5CV2, and EOS5CV3IT4. The results are shown in Figure 4.11, and the average absolute deviations (AAD) of each model are presented in Table 4.3.

Based on these results, we noticed that, as expected, the Batzle-Wang model is not recommended for fluids similar to ours. It is volatile and contains high amounts of carbon dioxide, resulting in an AAD of roughly 11.3%. This finding is in agreement with previous studies (ALTUNDAS, Yusuf B.; CHUGUNOV; RAMAKRISHNAN, 2013; TAHANI, 2012).

The De-hua Han model and the equation of state calibrated with PVT data reached similar performance, resulting in AAD between 2.0 and 2.5%, respectively. Therefore, both are recommended for future applications with this fluid. However, it is essential to remember that the equation of state approach is much less demanding in terms of straightforward coupling with the commercial compositional reservoir simulator, which already applies this kind of model. Based on these results, our best model is the EOS5CV3IT4. This finding was expected since this EOS was calibrated using the same experimental data shown in Figure 4.11 and managed to maintain the PVT match, as was presented in Figure 4.4.

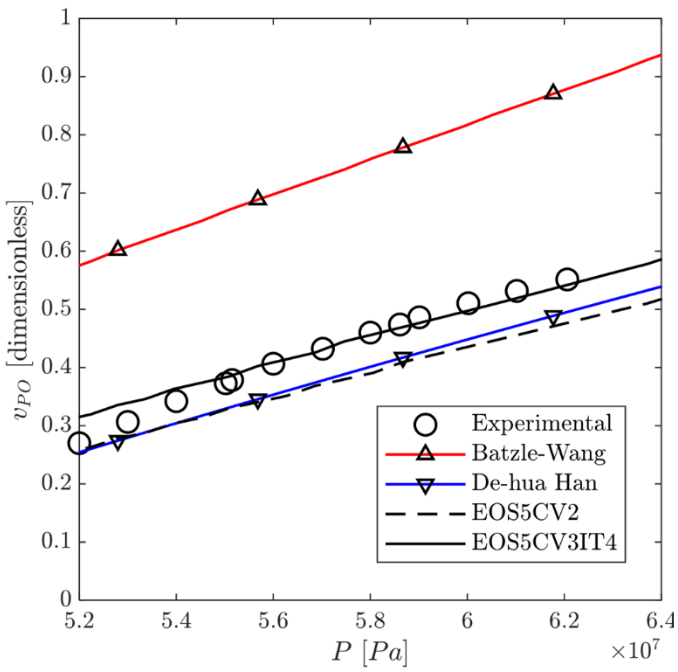


Figure 4.11: Batzle-Wang, De-hua Han and calibrated EOS models comparison with experimental data in a depletion experiment. The vertical axis is linearly transformed to stay in the interval [0,1] for confidentiality.

Table 4.3: Average absolute deviations of the models of speed of sound in the oil phase.

Model	AAD
Batzle-Wang	11.28%
De-hua Han	2.00%
EOS5CV2	2.49%
EOS5CV3IT4	0.35%

Except for the Batzle-Wang correlation, the deviation results of the models presented in Table 4.3 are compatible with other studies regarding speed of sound in reservoir fluids. For instance, Daridon et al. (1998) reported an AAD of 2.5% in light oil using Peng-Robinson equation with volumetric translation, Tahani (2012) presented AAD of less than 1% in reservoir oils using SAFT-BACK equation, and Dashti and Riazi (2014) reported AAD from 1.9% to 3.8% in crude oils, at varying temperatures and pressures, using a model based on the extended principle of corresponding states.

After this initial comparison of the models, we performed time-lapse seismic simulations using the modified SPE-10 model with two different production strategies: waterflood and gas injection. These results are presented in the following subsections.

4.8.1.2 Waterflood SPE-10 experiment

In this simulation, two acoustic impedance maps were generated using the PEM described in section 4.6.2, one before beginning the oil production and one after 2500 days, or 8.3% of the pore volume injected. Three fluid models were considered for comparison purposes: the EOS5CV3IT4, the De-hua Han, and the simplified model with constant γ. It is important to mention that only the simplified version of the De-hua Han model was implemented in a petroelastic model, coupled with the reservoir compositional flow simulator in our studies. This was because of the demanding calculations required to simulate a rigorous version of the correlation, which requires information about GOR and gas phase composition at surface conditions for the fluid in each reservoir cell. With our commercial compositional reservoir simulation, this would require time-consuming exporting and importing of grid properties or phase equilibria computations outside the reservoir simulator. The simplification adopted here was constant GOR and gas phase composition at surface conditions throughout the simulation. This is met in this waterflood experiment because the reservoir pressure was kept above the saturation point at all times, and the fluid compositions are constant inside the porous media. Thus, for comparison purposes, we can treat the results of the simplified version of the De-hua Han model, in this simulation, as equal to the original correlation.

The results in Figure 4.12 show that the differences between the De-hua Han and EOS5CV3IT4 models are neglectable in terms of impedance variations forecasts in a waterflood simulation. Both Figures (a and b) are nearly identical. Moreover, the absolute deviation values are less than 5 $\frac{g}{cm^3}\frac{m}{s}$, while the relative deviations are less than 2% at most of the points, except in a few cells where the low impedance variations cause high relative errors of roughly 10%. Since the impedance variation can exhibit values of zero, we used the definition of relative absolute deviation (RAD), defined in equation (4.14).

$$RAD = \frac{2\left|d_{sim,i} - d_{obs,i}\right|}{\left|d_{sim,i}\right| + \left|d_{obs,i}\right| + 10^{-6}} \tag{4.14}$$

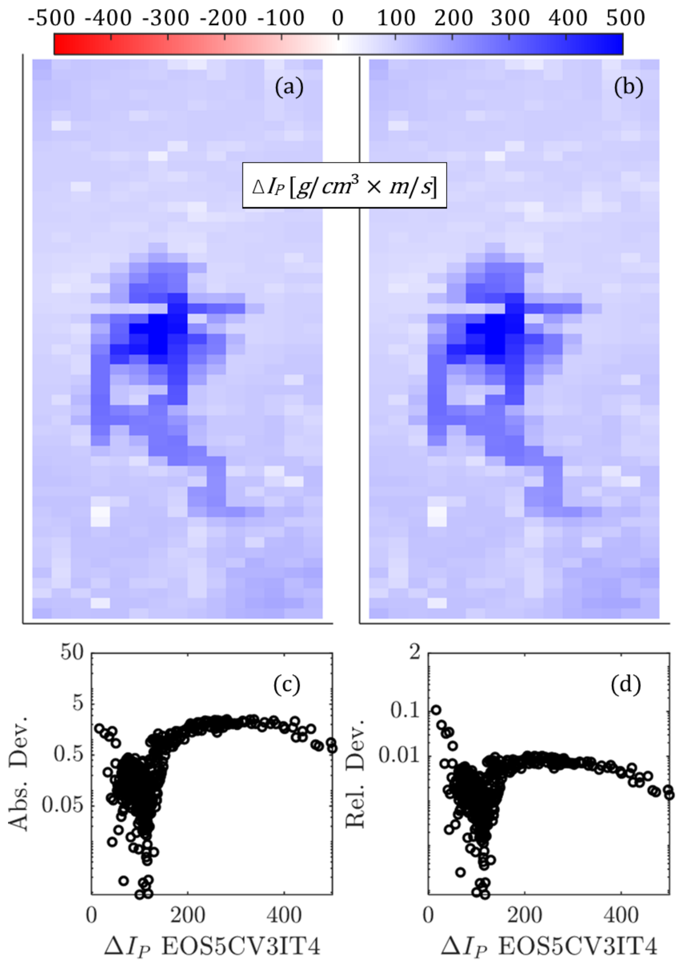


Figure 4.12: Comparison of the De-hua Han and EOS5CV3IT4 models in the waterflood simulation: (a) De-hua Han ΔI_P, (b) EOS5CV3IT4 ΔI_P, (c) absolute deviation, (d) relative absolute deviation.

In Figure 4.13, we compare the constant γ simplification with the EOS5CV3IT4 model. Again, the constant composition condition in the simulation is consistent with the premise of the simplification. Slight variations in γ occur due to pressure variations, but they have minor effects on the speed of sound and impedance results. Figures (a) and (b) are almost identical, and the absolute deviations are less than 5 $\frac{g}{cm^3}\frac{m}{s}$ and the relative absolute deviations are less than 2% in most cells. The only exceptions to the latter are in the cells where the impedance variations are nearly zero.

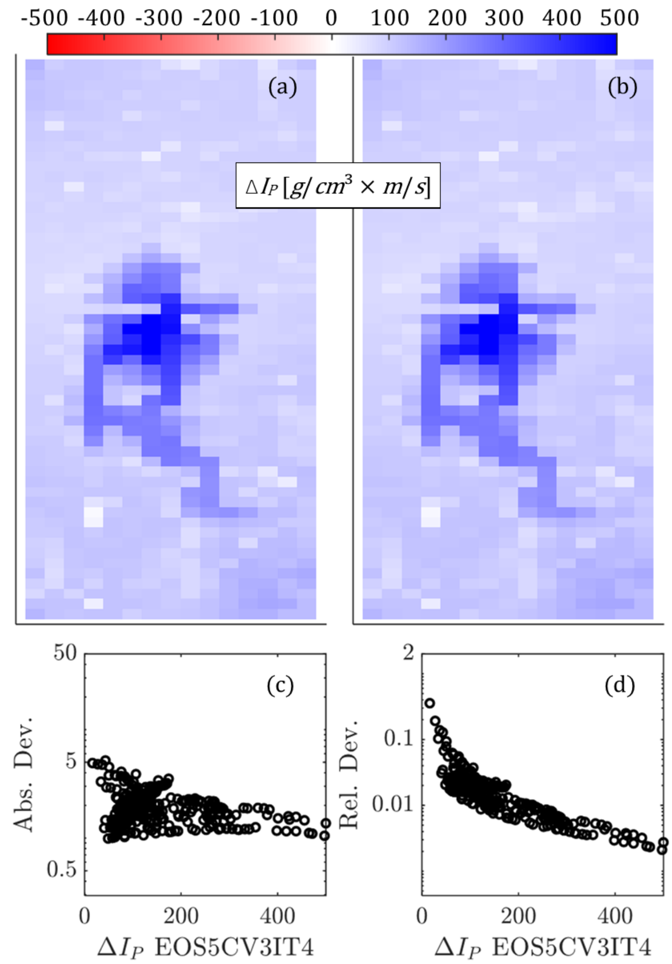


Figure 4.13: Comparison of the constant γ and EOS5CV3IT4 models in the waterflood simulation: (a) constant γ ΔI_P, (b) EOS5CV3IT4 ΔI_P, (c) absolute deviation, (d) relative absolute deviation.

The results of the waterflood simulations indicate that the models De-hua Han, constant γ, and EOS5CV3IT4 are capable of representing the fluid behavior in this test. The neglectable compositional variations inside the porous media during these simulations influences this result. However, this condition does not hold in the case of miscible or nearly miscible gas injection. Therefore, we performed the same test, changing the oil recovery strategy to gas injection. The results are shown in the next subsection.

4.8.1.3 Gas injection SPE-10 experiment

The gas injection simulation was performed throughout 2500 days of production when 45.5% of the porous volume was injected in the reservoir. The impedance variations from the beginning of the simulation until day 2500 were computed using the petroelastic model. The same fluid models were tested: the EOS5CV3IT4, the De-hua Han with constant gas content and composition, and the simplified model with constant γ. However, in this case, the gas injection caused the hydrocarbon fluid composition to change regionally and over time. Therefore, we expected significant deviations between the models.

We compare the De-hua Han model with constant gas content and composition with EOS5CV3IT4 in Figure 4.14. Even though figures a and b are visually similar, there are important differences, shown in c, caused by the simplification in the fluid model. Analyzing the difference between figures a and b, we notice that major deviations of roughly 50 $\frac{g}{cm^3}\frac{m}{s}$ occur at the gas front, where the oil saturation is still high, and the composition is changing. Another region in which significant differences occur is behind the gas front, where the gas has already replaced the oil. These differences, which vary between 9 and 10 $\frac{g}{cm^3}\frac{m}{s}$, are due to deviations in the baseline map. They are less intense in the waterflood case because error cancelation occurs when ΔI_p is computed with oil existing in both times. The systematic deviations present in the gas injection case, using the De-hua Han model, with constant gas content and composition, could cause bias in a data assimilation process and, when possible, should be avoided.

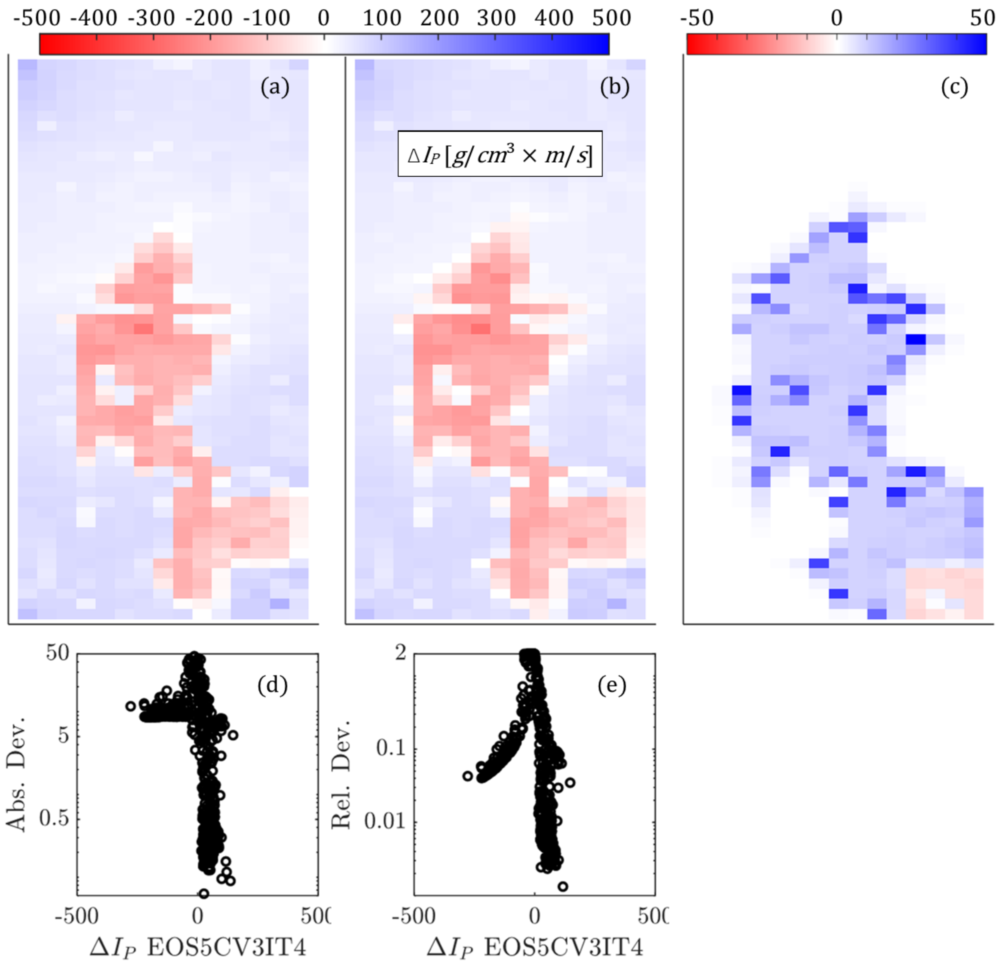


Figure 4.14: Comparison of the De-hua Han with constant gas content and composition and EOS5CV3IT4 models in the gas injection simulation: (a) De-hua Han ΔI_P, (b) EOS5CV3IT4 ΔI_P, (c) difference of images, (d) absolute deviation, (e) relative absolute deviation.

In Figure 4.15 we present the same analysis with the constant γ model, compared to EOS5CV3IT4. In this case, the differences are lower than in Figure 4.14. Nevertheless, significant deviations of roughly 10 $\frac{g}{cm^3}\frac{m}{s}$ occur at the gas front, where remaining oil with altered composition saturates the porous media. It is important to notice that we considered an almost exact $\gamma_i = 1.167$ (dimensionless) in the simulation, leading to no error in the baseline map. One should be cautious using this model if there is relevant uncertainty in determining the representative value of γ for baseline and monitor times.

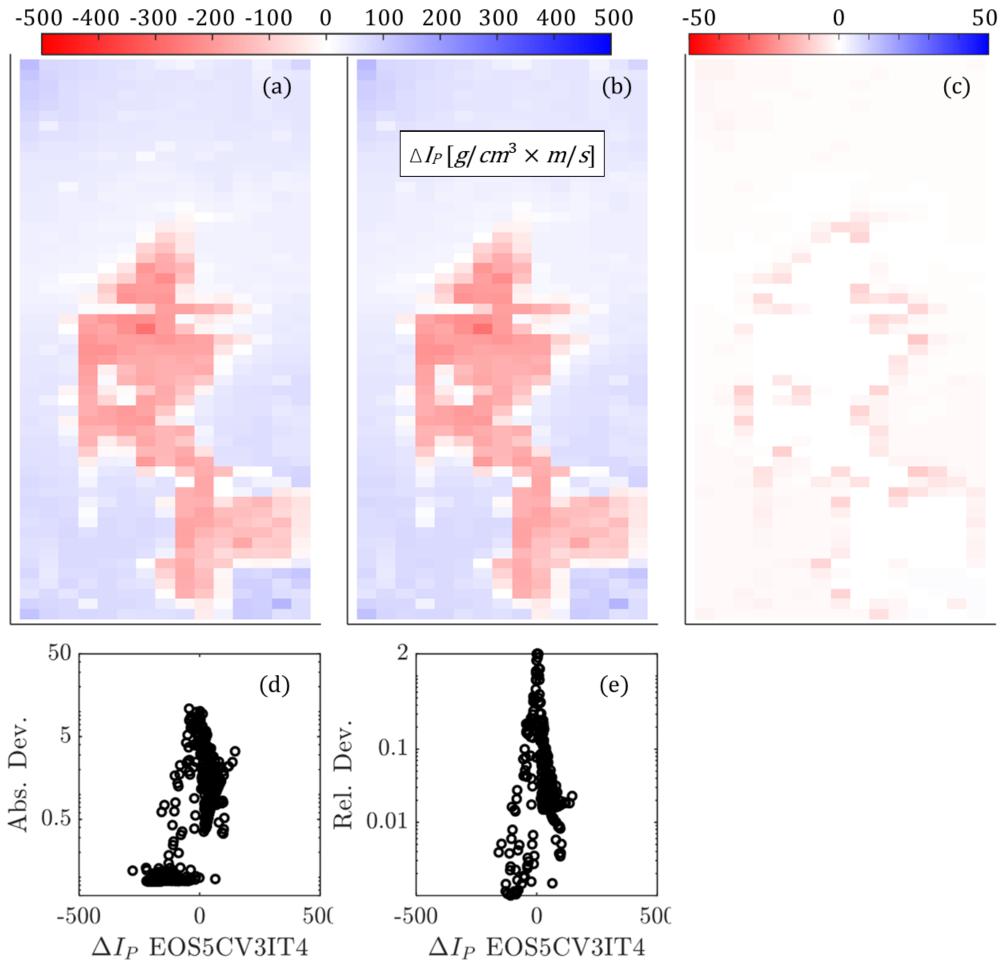


Figure 4.15: Comparison of the constant γ and EOS5CV3IT4 models in the gas injection simulation: (a) constant γ ΔI_P, (b) EOS5CV3IT4 ΔI_P, (c) difference of images, (d) absolute deviation, (e) relative absolute deviation.

We tested the same fluid models in the gas injection experiment using the Elastic Impedance attribute defined in equation (4.15) (AVSETH; MUKERJI; MAVKO, 2005; CONNOLLY, 1998; MAVKO; MUKERJI; DVORKIN, 2009; MUKERJI *et al.*, 1998). In this equation, $K = (v_S/v_P)^2$, the variables v_{Po}, v_{S0} and ρ_0 are normalization constants and were defined as the mean values of the baseline map using EOS5CV3IT4. The incidence angle, θ, was 30°. We applied the normalization proposed by Whitcombe (2002) to allow the direct comparison with the acoustic impedance images. The objective of this analysis is to confirm the results presented in Figure 4.14 and Figure 4.15 using an attribute that can be more sensitive to changes in porous fluids.

$$I_e(\theta) = v_{Po}\rho_0 \left(\frac{v_P}{v_{P0}}\right)^{1+(\tan\theta)^2} \left(\frac{\rho}{\rho_0}\right)^{1-4K(\sin\theta)^2} \left(\frac{v_S}{v_{S0}}\right)^{-8K(\sin\theta)^2} \qquad \textbf{(4.15)}$$

In Figure 4.16, we compare the De-hua Han model with constant gas content and composition with EOS5CV3IT4 using the Elastic Impedance attribute. The results are similar to Figure 4.14, except that the new attribute highlights the fluid-related anomalies. Furthermore, the absolute value of the difference of the images shown in Figure 4.16 (c) is also more significant, reaching values higher than $50 \frac{g}{cm^3} \times \frac{m}{s}$ at the gas front and around $12 \frac{g}{cm^3} \times \frac{m}{s}$ behind it. The same effect was observed in the comparison between the constant γ and the EOS5CV3IT4 in Figure 4.17. The fluid-related anomalies are more intense when compared to the pressure-related changes. Moreover, the absolute difference of the maps slightly increased, reaching values higher than $12 \frac{g}{cm^3} \times \frac{m}{s}$ at the gas front.

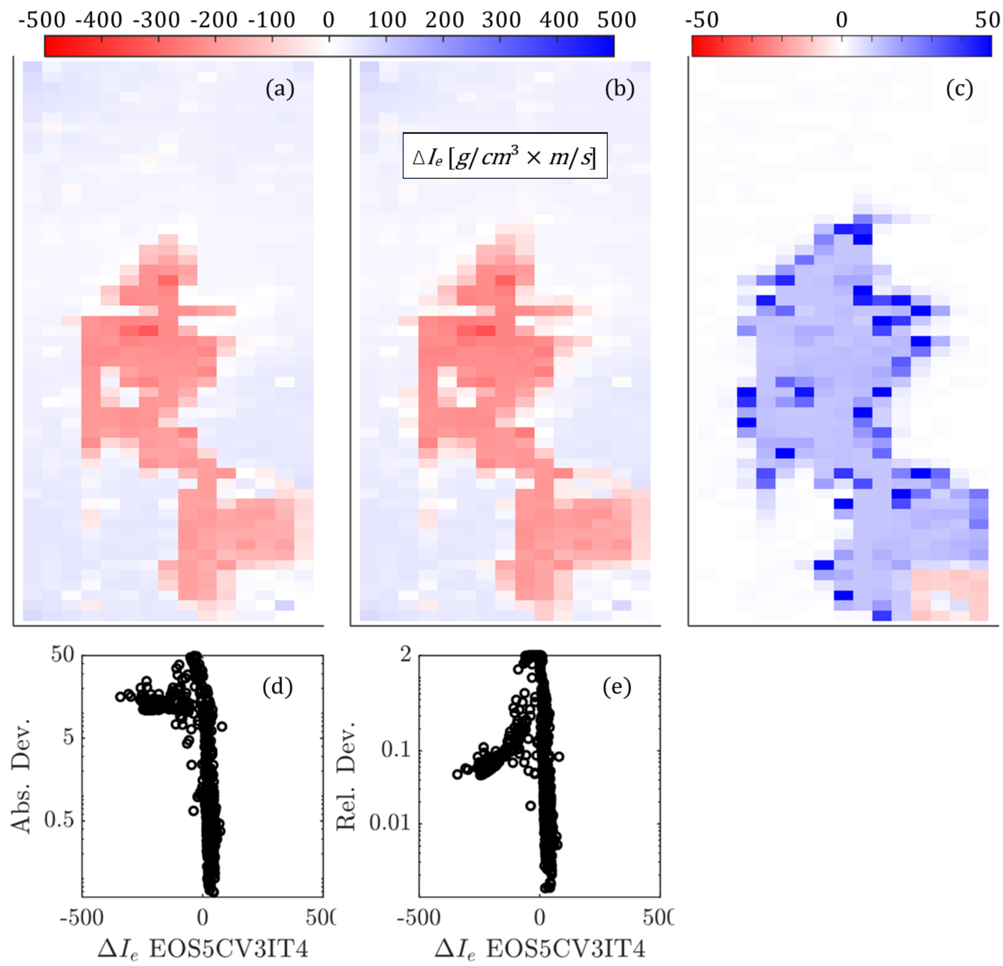


Figure 4.16: Comparison of the De-hua Han with constant gas content and composition and EOS5CV3IT4 models in the gas injection simulation using the Elastic Impedance: (a) De-hua Han ΔI_e, (b) EOS5CV3IT4 ΔI_e, (c) difference of images, (d) absolute deviation, (e) relative absolute deviation.

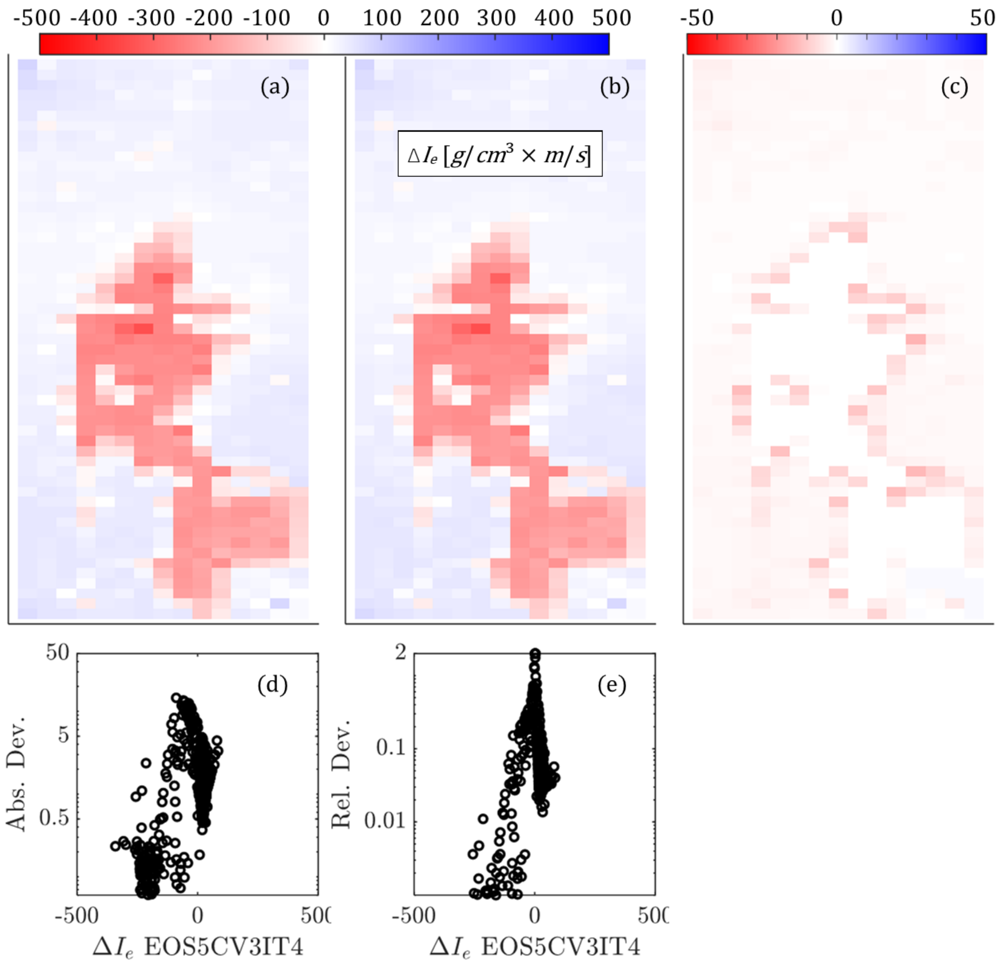


Figure 4.17: Comparison of the constant γ and EOS5CV3IT4 models in the gas injection simulation using the Elastic Impedance: (a) constant γ ΔI_e, (b) EOS5CV3IT4 ΔI_e, (c) difference of images, (d) absolute deviation, (e) relative absolute deviation.

The results of this subsection suggest that the simplified models of De-hua Han with constant gas content and composition, and constant γ, should be avoided when significant compositional changes may occur during the simulation. Despite the latter exhibiting deviations on a lower scale, there is a risk of higher error if the estimated γ value does not represent the fluid in the baseline or monitor conditions.

When we perform the gas injection simulations, the amount of light components in the oil phase increases in regions when it is in contact with the injected fluid. This situation is more complicated than our experimental data about the speed of sound in the oil phase since a constant gas content was imposed during the measurements. Therefore, we expect that all of our models would exhibit more significant deviations than the ones that we reported here if they were compared with comprehensive data with varying gas contents. Furthermore, we expect that the EOS5CV3IT4 model would perform better than the others in this situation for two reasons. First, it is more consistent with thermodynamic principles than the empirical correlations. Second, it is calibrated with all the available speed of sound data and PVT data. It is important to notice that this latter considers the variation of the volatile components' contents during the experiments.

If it is possible to obtain detailed information about the speed of sound in the oil phase with different gas content, we recommend applying our methodology, including all the compositions in steps 3 to 6 of Figure 4.2. Thus, one would obtain a model that is fully capable of representing the real fluid in miscible gas injection projects.

4.8.2 Additional analysis

In order to check the applicability of our methodology with more complex models, we decided to repeat it selecting 8 pseudo-components instead of 5. After that, we performed the data assimilation process, including PVT and PVT with speed of sound data. The results of this process are shown in Figure 4.18, where we can notice that the final simulated values of the speed of sound are approximately the same. Nevertheless, it is important to mention two differences in this case.

Firstly, by choosing to represent the fluid by 8 pseudo-components, we increased the number of uncertain parameters. Therefore, we needed to select more parameters to calibrate with the speed of sound data in the iterative process. Moreover, due to the higher complexity, the total number of iterations increased from 4, with 5 pseudo-components, to 6. However, this is not considered a critical issue.

Secondly, the higher degrees of freedom of the model with 8 pseudo-components led to higher variability of the speed of sound results of the 2 EOS version, which are calibrated only with the PVT data. Therefore, we decided to use the mean result of a set of calibrated equations instead of working with individual results.

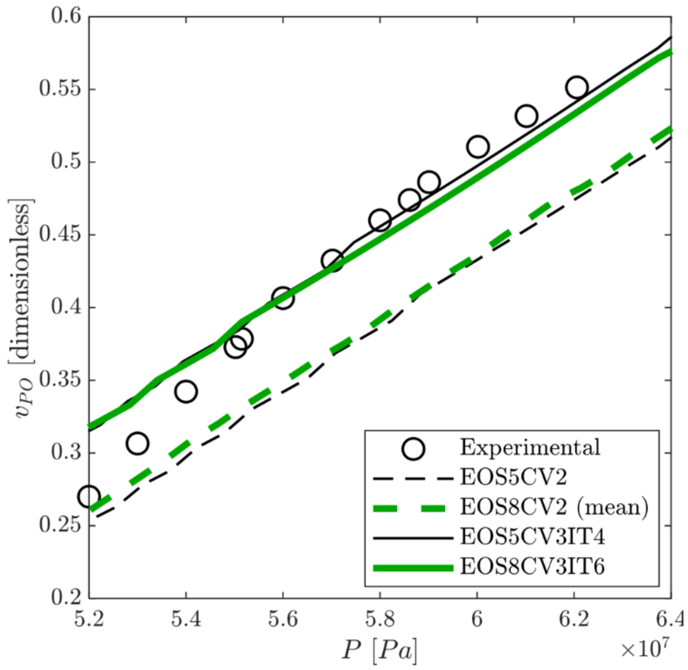


Figure 4.18: 5 and 8 pseudo-components EOS comparison. The vertical axis is linearly transformed to stay in the interval [0,1] for confidentiality.

4.9 Conclusions

In this work, we propose a methodology to characterize oil for time-lapse seismic analysis using Peng-Robinson equation of state calibrated with reservoir engineering and speed of sound data. We show that, by following this workflow, it is possible to obtain satisfactory models for volatile oils with high CO_2 content, which are characteristics of some pre-salt reservoir fluids. We compared our model with experimental data and correlations from the literature and described some situations when each of them may be applicable.

The main advantages of applying the EOS to speed of sound computations are: (1) the possibility to improve the fluid model using different information sources and (2) the easiness in coupling the flow and the petroelastic models using commercial tools since the same equations are applied for the fluid behavior. In practice, the petroelastic model with the cubic equation of state as a fluid model may be implemented as a plugin of a post-processing reservoir simulation tool. This may lead to simulations of impedance changes with one-click, facilitating the geophysics and reservoir engineering integration.

The specific conclusions of our work are:

- Our results show that it is possible to apply the same EOS to simulate the flow in the reservoir and the speed of sound changes in the oil phase, after the calibration using PVT data (2.5% average absolute deviations).
- Despite the favorable results that were achieved using the EOS calibrated with PVT data, we advocate for the acquisition of speed of sound experimental data to improve fluid models for both reservoir flow and petroelastic simulations.
- Based on our PVT data assimilation results, it is possible to improve the EOS model using the speed of sound data, maintaining the PVT match.
- Some simplified fluid models that consider constant gas content and characteristics may be applied to simulations where only slight compositional changes are expected, such as waterflood projects.
- In cases with miscible gas injection with a significative CO_2 content, we recommend performing a rigorous implementation of De-hua Han correlation or using a calibrated equation of state.
- In some early analysis, only isolated field data is available, such as saturation pressure, oil specific gravity, and flash GOR. In this situation, we recommend applying the De-hua Han correlation to predict the speed of sound in oil phase in projects where the fluid characteristics are close to ours.
- The application of Batzle-Wang correlations in other cases where the fluids are volatile and with significant CO_2 content would require specific tests to check its accuracy. In our tests, these correlations poorly represented the speed of sound in the oil phase, leading to average absolute deviations about 11.28%.

We list the following future researches to continue developing this work:

- Test fluid models using more extensive experimental data, including different gas contents.
- Analyze the capability of the speed of sound experimental data to provide information regarding the fluid behavior in a gas injection flow simulation.
- Study the uncertainty quantification considering fluid model parameters as uncertain attributes with different amounts of information available.

4.10 Symbols and nomenclature

Variable	Definition
a and b	Equation of State parameters
B_o	Oil formation volume factor
c	Volume-shift parameter
c_T	Isothermal compressibility
C_P	Heat capacity at constant pressure
C_V	Heat capacity at constant volume
d_{sim}	Simulated data
d_{obs}	Observed data
GOR	Gas-oil ratio
GOR_{CO_2}	Gas-oil ratio of CO_2
I_e	Elastic impedance
I_P	Acoustic impedance
P	Pressure
P_c	Critical pressure
R	Gas constant
R_s	Gas solution ratio in the oil
T	Temperature
T_c	Critical temperature
$v_{dead\ oil}$	Speed of sound in the oil without dissolved gas
v_{oil+CO_2}	Speed of sound in the oil with dissolved CO_2
$v_{oil+gas}$	Speed of sound in the oil with dissolved hydrocarbon gas
v_P	P-wave velocity
v_{Pe}	P-wave velocity normalization constant
v_{PO}	Speed of sound in the oil phase
v_S	S-wave velocity
v_{Se}	S-wave velocity normalization constant
$\bar{V}$	Molar volume
ΔI_e	Elastic impedance difference between monitor and base times
ΔI_P	Acoustic impedance variation between monitor and base times
Δv_{oil+CO_2}	Variation of the speed of sound in the oil when CO_2 is dissolved
γ	Specific heat ratio
θ	Incidence angle
ρ	Density
ρ_0	Density of the oil without dissolved gas
ρ_e	Density normalization constant
ω	Acentric factor

4.11 Acknowledgments

This work was conducted with the support of Libra Consortium (Petrobras, Shell, Total, CNOOC, CNPC and PPSA) within the ANP R&D tax as "commitment to research and development investments" and Energi Simulation. The authors are grateful for the support of the Center of Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil) and Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil). In addition, special thanks to CMG and Schlumberger for software licenses. The authors would also like to thank Guilherme Fernandes Vasquez and Marcos Sebastião dos Santos for the technical discussions that contributed to the enrichment of this study.

5 ASSIMILATING WELL AND TIME-LAPSE SEISMIC DATA IN A CHALLENGING PRE-SALT-LIKE CASE USING AN ITERATIVE ENSEMBLE SMOOTHER FOR BIG DATA SETS

Authors

Gilson Moura Silva Neto

Vinicius de Souza Rios

Célio Maschio

Alessandra Davolio

Denis José Schiozer

Declarations of interest: none.

5.1 Abstract

Time-lapse seismic and well data are complementary information sources to calibrate reservoir models, enabling them to provide useful production forecasts. Iterative ensemble smoothers are a typical class of methods to solve this kind of problem. Nevertheless, the data assimilation workflows, including both types of data, are commonly challenging.

Brazilian pre-salt, located in Santos and Campos sedimentary basins, is currently the most significant oil field province in Brazil regarding production rates and reserves. Besides the high productivity and large volumes, there are relevant characterization challenges for this province's reservoirs. We highlight the complex physics and the big data sets among the known challenges for well and time-lapse seismic data assimilation. The former is related to the heterogeneous porous media containing volatile fluids with high CO_2 contents and miscible gas injection, while the latter is associated with the big reservoir models and seismic monitoring projects.

With this background, this work presents a synthetic case study of a reservoir flow model calibration using well and time-lapse seismic data, employing two methods. The first is a compositional fluid model for time-lapse seismic simulation, enabling the consistent representation of the water alternating gas recovery strategy. The second is the Subspace Ensemble Randomized Maximum Likelihood method, with local analysis, an iterative ensemble smoother that is suitable for big reservoir models and big data sets. Using the realistic synthetic UNISIM-III benchmark, we show that these methods provide a viable solution to assimilate well and seismic data in a challenging pre-salt-like case. Furthermore, we demonstrate that the time-lapse seismic data provide useful information to improve production forecast in this situation.

Abbreviations:

BHP – Bottom-Hole Pressure

EOS – Equation of State

ESMDA – Ensemble Smoother with Multiple Data Assimilations

GOR – Gas-Oil Ratio

ICV – Interval Control Valve

IES – Iterative Ensemble Smoother

NQDS – Normalized Quadratic Deviation with Sign

PEM – Petroelastic Model

PVT – Pressure Volume Temperature

EnKF – Ensemble Kalman Filter

EnRML – Ensemble Randomized Maximum Likelihood

SEnRML – Subspace Ensemble Randomized Maximum Likelihood

TLS – Time-Lapse Seismic

TSVD – Truncated Singular Value Decomposition

WAG – Water-Alternating-Gas

Keywords

Model calibration; iterative ensemble smoothers; history matching; data assimilation; time-lapse seismic.

5.2 Introduction

Brazilian pre-salt is an important oil province, contributing to a significant portion of the country's reserves and daily production (ABELHA; PETERSOHN, 2018; DE MORAES CRUZ *et al.*, 2016; VASQUEZ; MORSCHBACHER; JUSTEN, 2019). Located in Santos and Campos sedimentary basins, the pre-salt reservoirs consist of microbial and coquina rocks at depths that surpass 5000 m (JOHANN; MONTEIRO, 2016). The task of building geological simulation models for these reservoirs is highly challenging, involving significant technical uncertainties. Some of the main uncertainties related to pre-salt simulation models are reservoir connectivity, facies and petrophysical properties distributions, response to the enhanced oil recovery strategy, distribution and behavior of faults and fractures, and fluid properties (MOCZYDLOWER, B.. *et al.*, 2012). Furthermore, the projects in this province involve huge investments (DE SANT´ANNA PIZARRO; BRANCO, 2012). All these characteristics corroborate the importance of mitigating the model uncertainties using all the information available. In this respect, well and time-lapse seismic (TLS) data are complementary sources of information to calibrate the reservoir simulation models. The former provides information abundant in time but scarce in space, especially in offshore projects, the pre-salt province situation. The latter provides information distributed in space, helping updating parameters far from the wells.

Iterative ensemble smoothers (IES) are a popular choice for assimilating well ant TLS data into reservoir models (EMERICK, Alexandre A., 2016; EMERICK, Alexandre A.; REYNOLDS, 2013b; FAHIMUDDIN; AANONSEN; SKJERVHEIM, 2010a; SKJERVHEIM, Jan-Arild *et al.*, 2007). This type of application is associated with relevant challenges related to pre-salt reservoirs. The reservoirs are big and highly heterogeneous in terms of permo-porous and facies distribution, diagenesis, faults, and fractures (JOHANN; MONTEIRO, 2016). These characteristics complicate the representation of the reservoir's main features using an ensemble with a limited number of models, each with a restricted number of active cells. Both the ensemble and the model sizes contribute to increasing computational costs. The production involves complex physics, with volatile fluids, high CO_2 contents, miscible gas injection alternated with water (JOHANN; MONTEIRO, 2016; MOCZYDLOWER, B.. *et al.*, 2012). Therefore, compositional fluid models are necessary to represent this process, increasing the complexity of both reservoir flow and time-lapse seismic forward models. The reservoir sizes and the application of seismic monitoring technologies (DEPLANTE *et al.*, 2019; JOHANN;

MONTEIRO, 2016) contribute to increasing data points, requiring efficient ensemble-based methods to handle big data sets.

Batzle and Wang (1992) proposed classical correlations, which have been the standard method for fluid characterization in quantitative TLS studies (AVSETH; MUKERJI; MAVKO, 2005). Nevertheless, the presence of volatile fluids containing significant amounts of CO_2 limits classical correlations' applicability to estimate the speed of sound in the oil phase (ALTUNDAS, Bilgin *et al.*, 2017; SILVA NETO, Gilson M. *et al.*, 2020; TAHANI, 2012). There are specific correlations for fluids with such characteristics (HAN; SUN; LIU, 2012, 2013). However, they require significant extra computations to integrate with compositional reservoir simulators, and one still needs to validate them for conditions that are different from the ranges applied during the experiments.

With this in mind, Silva Neto et al. (2020) proposed the application of a standard cubic equation of state (EOS) calibrated for the reservoir fluids to estimate the speed of sound in the reservoir hydrocarbons as part of the petroelastic model (PEM). This model has the advantage of efficiently coupling with the compositional reservoir flow simulator, which involves the same EOS. Furthermore, in their test with the Peng-Robinson EOS (PENG; ROBINSON, 1976; ROBINSON; PENG, 1978), they obtained simulations matching the experimental data in a similar level as Han et al. (2012, 2013) correlations, when they calibrated the EOS parameters with pressure-volume-temperature (PVT) data, a standard procedure in reservoir engineering. However, Silva Neto et al.'s study (2020) did not include a time-lapse seismic data assimilation experiment, which we perform using the same model in the present work. Therefore, this is the first application of this fluid model in a time-lapse seismic data assimilation case, including miscible gas injection and water-alternating-gas (WAG) injection.

The Ensemble Randomized Maximum Likelihood (EnRML) method initially represented an iterative scheme to improve the performance of the Ensemble Kalman Filter (EnKF) (EVENSEN, Geir, 1994) for highly nonlinear applications (GU; OLIVER, 2007). Later, Chen and Oliver (2012, 2013) adapted the method for batch data assimilation instead of the previous sequential approach, thus avoiding the need for time-consuming simulation restarts in reservoir applications. Recently, Raanes et al. (2019) improved the EnRML method conceptually and computationally using the property that the solution is in the ensemble subspace. Evensen et al. (2019) followed their work in the same year and proposed an efficient algorithm to calibrate reservoir models using big data sets, the Subspace EnRML (SEnRML). Silva Neto et al. (2021) applied the SEnRML method with local analysis to assimilate TLS data in synthetic applications. They concluded that this method could lead to equivalent results to the Ensemble Smoother with Multiple Data Assimilations (EMERICK, Alexandre A.; REYNOLDS, 2013a) with Kalman gain localization. They also reported that the SEnRML with local analysis has the advantage of requiring lower computational costs when there is a big data set.

In this work, we apply the SEnRML method with local analysis to assimilate well and TLS data. This method is promising for pre-salt-related applications due to the big reservoir models and seismic monitoring. SEnRML with local analysis provided reasonable results in the previous application in TLS data assimilation to update reservoir grid parameters (SILVA NETO, Gilson Moura *et al.*, 2021). In this work, we increase the problem complexity in terms of the model, the inclusion of well data, and the calibration of different parameter types, grid, scalar, and categorical. Therefore, this application mimics most of the challenges of a real field. The current case study uses the benchmark called UNISIM-III (CORREIA *et al.*, 2020), which follows a pre-salt reservoir's characteristics. To our knowledge, this work is the first application of SEnRML with local analysis in a data assimilation workflow to improve reservoir characterization using both well and seismic data in a pre-salt-like field.

The specific objectives of this study are:

- To apply the compositional fluid model in time-lapse seismic data assimilation in a realistic case.
- To validate the SEnRML method, with local analysis, in a complex, realistic case, including well and seismic data.
- To analyze the time-lapse seismic benefits in a pre-salt-like case.

5.3 Background Information and methods

This work integrates two methods proposed previously to enable the well and seismic data assimilation in a pre-salt-like synthetic application. We apply a compositional fluid model to compute the forward seismic simulation and use an iterative ensemble smoother that is suitable for big data sets and big reservoirs, called SEnRML, with local analysis, to calibrate the parameters. Figure 5.1 illustrates our workflow, including the reservoir flow model (1), which provides the inputs for the compositional model for speed of sound in the fluid (2), enabling the petroelastic model computations (3). We ran the reservoir flow models with the compositional reservoir simulator GEM version 2017.1 (CMG, 2017). The production and TLS data from the reservoir flow and the petroelastic models are compared to the observed data. The differences cause model parameter updates through the SEnRML with local analysis (4).

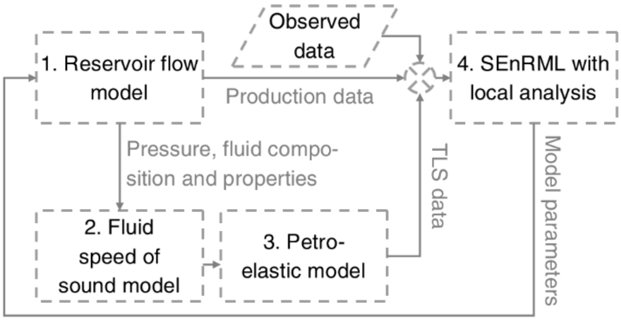


Figure 5.1: Well and time-lapse seismic data assimilation workflow.

We describe the fluid model and the data assimilation method in the following subsections. Besides the data assimilation workflow, we use a quadratic metric called Normalized Quadratic Deviation with a Sign (NQDS) to analyze the well data match. We define this metric in subsection 5.3.3. We address the petroelastic model, step three, in section 5.4.4.

5.3.1 Compositional fluid model for seismic simulation

The current work's application considers volatile oil with around 40% CO_2 content. The widely known Batzle and Wang (1992) correlation exhibit a relatively high deviation in representing the speed of sound in the oil phase with these characteristics (ALTUNDAS, Bilgin *et al.*, 2017; TAHANI, 2012). Silva Neto et al. (2020) proposed to use a calibrated cubic equation of state (EOS) to model the speed of sound in the hydrocarbon phases. We apply this model to compute the second step of the workflow depicted in Figure 5.1.

Considering the Peng-Robinson EOS (PENG; ROBINSON, 1976; ROBINSON; PENG, 1978) with volume translation (PÉNELOUX; RAUZY; FRÉZE, 1982), the pressure-volume-temperature relation is

$$P = \frac{RT}{(\bar{V} + c_{PR} - b_{PR})} - \frac{a_{PR}\left[1 + m_{PR}\left(1 - \frac{T^{0.5}}{T_c^{0.5}}\right)\right]^2}{(\bar{V} + c_{PR})(\bar{V} + c_{PR} + b_{PR}) + b_{PR}(\bar{V} + c_{PR} - b_{PR})}, \tag{5.1}$$

where P is the pressure, T the temperature, $\bar{V}$ the molar volume, R the gas constant, and the parameters a_{PR}, b_{PR}, c_{PR}, and m_{PR} undertake different values for different components. One can calculate them as a function of the acentric factor, critical pressure, P_c, and critical temperature, T_c. Furthermore, it is necessary to apply mixing rules to represent the oil and gas phases as mixtures of components and pseudo-components (PEDERSEN; CHRISTENSEN;

SHAIKH, 2015). From de relation defined in equation (5.1), one can compute the heat capacity at a constant volume

$$C_V = \left.\frac{\partial H^{id}}{\partial T}\right|_P - \frac{m_{PR} a_{PR}(1 + m_{PR})}{4\sqrt{2}T^{0.5}T_c^{0.5} b_{PR}} \ln\left\{\frac{\bar{V} - [(-1 + \sqrt{2})b_{PR} - c_{PR}]}{\bar{V} - [(-1 - \sqrt{2})b_{PR} - c_{PR}]}\right\} - R, \tag{5.2}$$

where the first term on the right side is a derivative of the ideal enthalpy at a constant pressure, which one can calculate from the relations that the Winprop (CMG, 2015) fluid simulator provides for the reservoir flow simulation. After calculating the heat capacity at a constant volume, it is possible to obtain the heat ratio using

$$\frac{C_P}{C_V} = 1 - \frac{T}{C_V} \frac{\left(\left.\frac{\partial P(T,\bar{V})}{\partial T}\right|_{\bar{V}}\right)^2}{\left.\frac{\partial P(T,\bar{V})}{\partial \bar{V}}\right|_T}, \tag{5.3}$$

in which the pressure partial derivative at a constant molar volume (numerator) and a constant temperature (denominator) are calculated analytically from equation (5.1). Finally, the speed of sound in the fluid is

$$v_P = \sqrt{\frac{C_P}{C_V} \times \frac{1}{\rho c_T}}, \tag{5.4}$$

where ρ is the fluid density and c_T is the isothermal compressibility, which are outputs from the reservoir flow simulator. We built the EOS model for the reservoir simulator using Winprop version 2015.10 (CMG, 2015).

Silva Neto et al. (2020) concluded that the present model could reasonably represent the speed of sound in the oil, as long as one calibrates the EOS parameters using PVT data, which is a standard procedure in reservoir engineering. Furthermore, if the speed of sound laboratory data is available, it is possible to calibrate the EOS with this information to improve the model without impairing the PVT data match.

5.3.2 The Subspace Ensemble Maximum Likelihood (SEnRML) method with local analysis

Focusing on the fourth step of the workflow depicted in Figure 5.1, the current SEnRML implementation with local analysis follows the revision presented by Raanes et al. (2019), the efficient algorithm for big data sets proposed by Evensen et al. (2019), and the local analysis scheme of Silva Neto et al. (2021). In this section, we present a method summary, highlighting the main features of this algorithm.

The SEnRML method aims at minimizing the objective function

$$\mathcal{J}(\boldsymbol{w}_j) = \frac{1}{2}\boldsymbol{w}_j^T\boldsymbol{w}_j + \frac{1}{2}\left[\boldsymbol{g}(\boldsymbol{x}^a) - \boldsymbol{d}_j\right]^T \boldsymbol{C}_{dd}^{-1}\left[\boldsymbol{g}(\boldsymbol{x}^a) - \boldsymbol{d}_j\right], \tag{5.5}$$

where $\boldsymbol{g}(\boldsymbol{x}^a)$ is the forward simulation model as a function of the updated parameters, $\boldsymbol{x}^a$, the variable $\boldsymbol{d}_j$ is the perturbed observed data and it follows the distribution $\mathcal{N}(\boldsymbol{d}^{obs}, \boldsymbol{C}_{dd})$, with a covariance matrix of measurement errors $\boldsymbol{C}_{dd}$, and $\boldsymbol{w}_j$ are column vectors that define the changes in the parameters during the calibration for each model. Therefore, the first term on the right side of equation (5.5) relates to the distance to the prior ensemble, and the second one refers to the data misfit. These two terms form the total cost function of the Bayesian methods (EVENSEN, Geir, 2009). The ensemble of updated parameters forms the matrix

$$\boldsymbol{X}^a = \boldsymbol{X}^f + \boldsymbol{AW}, \tag{5.6}$$

in which $\boldsymbol{X}^f$ is a matrix whose columns are prior parameters samples and $\boldsymbol{A}$ are ensemble anomalies defined as

$$\boldsymbol{A} = \boldsymbol{X}^f \frac{1}{\sqrt{N-1}}\left(\boldsymbol{I}_N - \frac{1}{N}\boldsymbol{1}_N\boldsymbol{1}_N^T\right) = \boldsymbol{X}^f\boldsymbol{\Pi}_N, \tag{5.7}$$

Where N is the ensemble size and the projector $\boldsymbol{\Pi}_N$ removes the mean and normalizes the matrix by $\sqrt{N-1}$. Note that the change of variables defined in equation (5.6) means that the model updates are a linear combination of the prior ensemble anomalies. The algorithm updates the matrix $\boldsymbol{W}$, that defines this linear combination.

One obtains the iterative procedure to update the matrix $\boldsymbol{W}$ and the parameters applying the Gauss-Newton method in the cost function equation (5.5). After some manipulations, it is possible to find

$$\boldsymbol{W}^{i+1} = \boldsymbol{W}^i - \boldsymbol{\gamma}\left[\boldsymbol{W}^i - \left(\boldsymbol{S}^i\right)^T\left(\boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^T + \boldsymbol{C}_{dd}\right)^{-\boldsymbol{1}}\boldsymbol{H}^i\right], \tag{5.8}$$

in which the step-length parameter, γ, controls the update speed, $\boldsymbol{S}^i$ is the matrix of predicted and deconditioned ensemble anomalies

$$\boldsymbol{S}^i = \left(\boldsymbol{D}^{sim}\right)^i\boldsymbol{\Pi}_N\left(\boldsymbol{I}_N + \boldsymbol{W}^i\boldsymbol{\Pi}_N\right)^{-1}, \tag{5.9}$$

where the simulated data form the matrix $\boldsymbol{D}^{sim}$, $\boldsymbol{I}_N$ is the identity matrix with size equals to the ensemble size, N. $\boldsymbol{H}^i$ is called the matrix of innovations, defined as

$$\boldsymbol{H}^i = \boldsymbol{S}^i\boldsymbol{W}^i + \boldsymbol{D} - \left(\boldsymbol{D}^{sim}\right)^i, \tag{5.10}$$

considering the matrix containing the perturbed observed data, $\boldsymbol{D}$. Note that we previously defined each column of $\boldsymbol{D}$ as $\boldsymbol{d}_j$.

We compute the matrix inverse $\left(\boldsymbol{S}^i\left(\boldsymbol{S}^i\right)^T + \boldsymbol{C}_{dd}\right)^{-1}$ by representing the covariance matrix $\boldsymbol{C}_{dd}$ with the approximation $\boldsymbol{C}_{dd} \approx \boldsymbol{E}\boldsymbol{E}^T$, where $\boldsymbol{E}$ columns are samples of the distribution $\mathcal{N}(0, \boldsymbol{C}_{dd})$ normalized by $\sqrt{N_E - 1}$. The number of samples that form $\boldsymbol{E}$, N_E, is a tradeoff between the approximation fidelity and the computational cost. After that, we project the approximated matrix onto the subspace defined by $\boldsymbol{S}$. Aiming at performing the matrix inverse, we compute a truncated singular value decomposition (TSVD) of $\boldsymbol{S}$ and an eigenvalue decomposition of the modified covariance matrix of measurement errors. It is relevant to note that the user needs to define the fraction of the $\boldsymbol{S}$ matrix singular values to keep in the TSVD. The final update equation scales linearly with the data size, improving the efficiency for big data sets compared to other ensemble-based methods that form the full matrix $\boldsymbol{C}_{dd}$. In these methods, the equations scale with the square of the number of data points. For instance, see the algorithm (EMERICK, Alexandre A., 2016). For more details regarding this method, we refer to (EVENSEN, Geir *et al.*, 2019; SILVA NETO, Gilson Moura *et al.*, 2021).

5.3.2.1 The local analysis scheme

The limited ensemble size makes it vital to apply a localization strategy in ensemble-based data assimilation. This technique mitigates exaggerated uncertainty reduction due to spurious correlations and limited degrees of freedom (EMERICK, Alexandre; REYNOLDS, 2011). Silva Neto et al. (2021) proposed a local analysis scheme to assimilate time-lapse seismic data using the efficient implementation of the SEnRML method. We apply this algorithm here to assimilate well and TLS data in the current case study.

In the local analysis scheme, we divide the data assimilation problem into independent analyses. In each one of them, we update a predefined subset of the parameters, called local group, using only the part of the data set that we assume correlated to the group. Each problem follows the same data assimilation procedure, described in equations (5.6) to (5.10).

One can segregate the parameters using their physical positions or considering the correlation between them and the data. A popular choice is to include in the same analysis all the parameters in vertical columns of grid cells from the reservoir model (CHEN, Yan; OLIVER, 2017; SILVA NETO, Gilson Moura *et al.*, 2021), which is the configuration that we adopted in the current study. Furthermore, we updated each scalar parameter in an individual group, enabling a refined analysis for these parameters that significantly impact the model response. It is worth mentioning that creating a local group for each scalar parameter causes a minor increase in the computational costs in practical applications because the number of parameters of this type is usually much lower than the number of grid parameters.

One approach of selecting the data that influence each local group is called distance-based localization, in which the algorithm computes the physical distance between each local group and the data point. Note that a well data is at the well position. The method includes any data located at the same position as the group with weight 1. Moreover, it tapers the influence of the remaining data using the Gaspari-Cohn function (GASPARI; COHN, 1999), defining the argument as the distance, normalized by the so-called localization lengths. One can consider these localization lengths as tunning parameters of the method. In our tests, the whole data set influences scalar parameters that do not have a specific physical position in the model, for instance, relative permeability tables. We call this procedure a global update. We assumed that one well data do not influence scalar parameters related to other wells' productivity or injectivity during the data assimilation.

Another method to select the data that influence the local groups is correlation-based localization (LUO; BHAKTA, 2020). In this case, the algorithm assumes the correlation threshold

$$\theta = \frac{1}{\sqrt{N}}\sqrt{2\ln(n_{ac})}, \quad \textbf{(5.11)}$$

where n_{ac} is the number of active cells in the reservoir model. This threshold relates to the statistical noise in the ensemble estimate of the correlation matrix between the data and parameters. For each parameter in a group and each data point, the influence tapering is the result of the Gaspari-Cohn function using the argument

$$z = \max\left(1.67 - 0.67\frac{|r|}{\theta}, 0\right), \quad \textbf{(5.12)}$$

in which z is called the pseudo-distance dummy variable, and r is the correlation between the parameter and the simulated data point, computed from the prior ensemble results (SILVA NETO, Gilson Moura *et al.*, 2021). Note that each group comprises a certain number of parameters. Therefore, it is necessary to define which pseudo-distance value will prevail for the group. If one chooses the minimum value, all data that influences at least one parameter in the group will influence the whole group. The maximum z will include only data points that relate to all parameters in each group. Finally, an intermediate option is to use a percentile of the z distribution in the groups. Note that the smaller the groups, the more insignificant this choice is. In this work, we tapered the data influence using the minimum value of z for each local group, which seems to be a conservative choice, avoiding neglecting correlated data at the cost of a more severe uncertainty reduction.

5.3.2.2 Configuration of the SEnRML method with local analysis

The SEnRML method with the local analysis scheme described in this section has some user-defined parameters. We list these parameters in Table 5.1. We comment on the parameters' most relevant influences in the results section 3.6.

Table 5.1: SEnRML with local analysis parameters.

Parameter	Configuration
E matrix size	1000 columns ($10 \times N$)
Fraction of the singular values in TSVD	0.99
Step-length control (γ)	Declining from 0.5 to 0.1
Parameter segregation	Grid: vertical columns of cells Scalar: one parameter per group
Localization distance (distance-based local analysis)	Well data: based on the influence area Seismic: 1400 m (7 grid cells)
Pseudo-distance (correlation-based local analysis)	Minimum value

5.3.3 NQDS metric

The Normalized Quadratic Deviation with Sign (NQDS) measures the distance between the observed and simulated data. The main characteristic of this quadratic norm is that it designates a negative sign to ensemble elements that underestimates the data and a positive one to the models that tend to overestimate them (AVANSI; MASCHIO; SCHIOZER, 2016). Therefore, it helps to identify bias and interpret the results' physical meaning. Commonly, one plots the NQDS statistical distribution of a subset of the data (CAVALCANTE *et al.*, 2020; FORMENTIN *et al.*, 2019), such as individual producers' oil rate and injectors' bottom-hole pressure. We define the NQDS related to a data subset l of an ensemble element j as

$$\left(NQDS_j\right)_l = \frac{\mathbf{1}_{m_l}^T\left[\left(\boldsymbol{d}_j^{sim}\right)_l - (\boldsymbol{d}^{obs})_l\right]}{\left|\mathbf{1}_{m_l}^T\left[\left(\boldsymbol{d}_j^{sim}\right)_l - (\boldsymbol{d}^{obs})_l\right]\right|} \times \frac{\left[\left(\boldsymbol{d}_j^{sim}\right)_l - (\boldsymbol{d}^{obs})_l\right]^T\left[\left(\boldsymbol{d}_j^{sim}\right)_l - (\boldsymbol{d}^{obs})_l\right]}{[(\boldsymbol{d}^{obs})_l \times \tau_l + \varphi_l]^T[(\boldsymbol{d}^{obs})_l \times \tau_l + \varphi_l]}, \quad \textbf{(5.13)}$$

where $\mathbf{1}_{m_l}^{\boldsymbol{T}}$ is a row vector whose elements are one and whose size is the number of points in the data subset, m_l, $\boldsymbol{d}_j^{sim}$ is the simulated data, $\boldsymbol{d}^{obs}$ is the observed data, τ is a tolerance value, and φ is a constant. The tolerance is analogous to the relative standard deviation of the measurement noise if one defines it as uncorrelated. The constant φ avoids division by numbers close to zero if the measured data is close to zero for a specific data type. We used 10% and 3% as the rates and pressure tolerances in this work, respectively. Furthermore, we set 40 as constant for the rates and 0.01 for the pressures. The latter is not meaningful because the pressure is never close to zero. Note that in equation (5.13), the first term on the right side relates to the positive or negative sign, while the second one is the Normalized Quadratic Deviation (NQD).

5.4 Application

5.4.1 General case description

The present case study uses the UNISIM-III benchmark model representing a fractured carbonate karst reservoir from the pre-salt province (CORREIA *et al.*, 2020). We only included the so-called Sector 1 as a hydraulic isolated reservoir model, which has a production strategy considering eight producers and nine injectors and corresponds to one platform's drainage area, as depicted in Figure 5.2a. The model has two zones. The upper consists mainly of stromatolites, while the lower corresponds to the coquinas. We present the model geometry and these zones in Figure 5.2b. There are three faults in this sector, whose locations are indicated in Figure 5.2a.

The observed data come from a fine-scale reference model, which we consider the true earth model or, in other words, the synthetic truth. The reservoir simulation model represents partial information from the synthetic truth at the wells' locations. Therefore, the reference model contains unknown characteristics during the simulation studies, mimicking a real field application. The geological uncertainty in the simulation model enables generating an ensemble of 100 geostatistical realizations, each one with different permeability and porosity fields. This ensemble size is a tradeoff between the representation of the uncertainty statistics and the computational costs to simulate the forward models. Table 5.2 lists the general models' characteristics. For more details regarding the geological models, we refer to (CORREIA *et al.*, 2020).

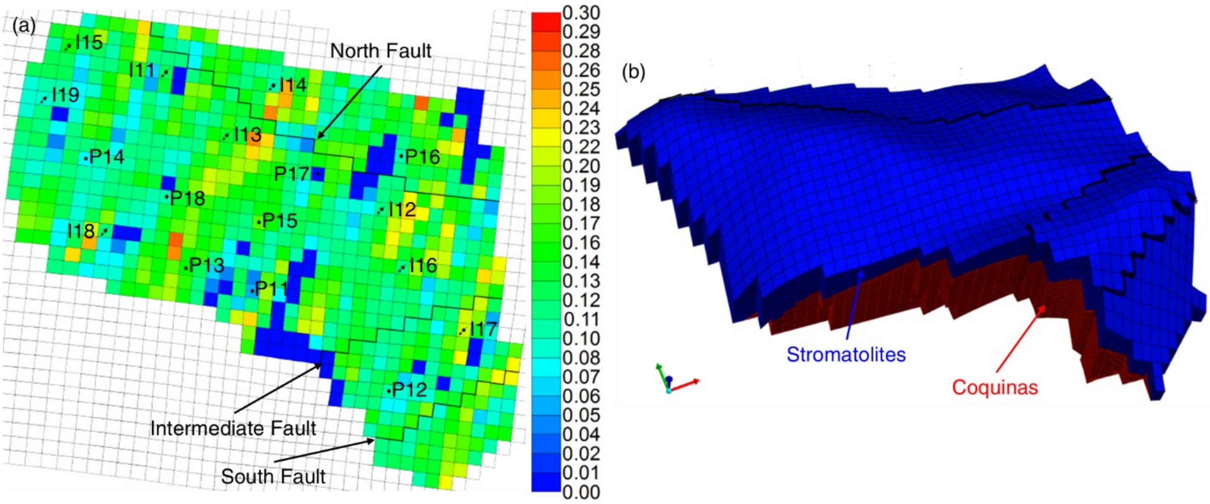


Figure 5.2: UNISIM-III Sector 1 model. The figure depicts an intermediate layer porosity map sample from the prior ensemble (a), including well and fault locations, and the 3D model (b), highlighting the two zones, stromatolites and coquinas. In the well's names, the letter "P" refers to a producer and "I" to an injector.

Table 5.2: General simulation model characteristics.

Characteristic	Value
Horizontal permeability (mD)	0 to 9000 (median ≈ 50)
Vertical permeability (mD)	0 to 900 (median ≈ 1)
Porosity	0 to 0.3 (median ≈ 0.11)
Average depth (m)	≈ 5500
Initial datum pressure (kPa)	63000
Average cell size (m)	≈ 200 × 200 × 5
Number of active cells	77071
Total number of cells	533403

The benchmark reservoir model includes hydrocarbon fluids whose characteristics reproduce the public report (PETROBRAS, 2015) regarding a pre-salt field. We present the reservoir fluids' main characteristics in Table 5.3. Among them, it is worth highlighting the high CO_2 content, around 40%. Furthermore, the fluid volatility is associated with the initial gas-oil ratio (GOR) of 415 and the high oil formation volume factor of 2. The oil phase at reservoir conditions presents low viscosity, roughly 0.4 cP, which is favorable to the recovery process sweep efficiency.

Table 5.3: General fluid properties.

Characteristic	Value
API (°)	27
Initial gas-oil ratio (L/L)	415
Oil formation volume factor (L/L)	2
Initial saturation pressure (kPa)	≈ 49000
Initial oil viscosity (cP)	≈ 0.4
CO_2 in the gas phase (%)	44
CO_2 in reservoir fluid	37%
Temperature at the reservoir (°C)	90

The benchmark case considers 17 vertical wells in Sector 1, eight producers and nine injectors, as presented in Figure 5.2a. All well completions include interval control valves (ICV), allowing the flow control in two zones for the injectors and three intervals for the producers. The injectors' ICV aim at uniformizing the injection of water and gas between the zones. Therefore, when it detects a predefined level of unbalance between the two intervals, the ICV closes the one that received a larger volume and opens the other. In the producers, the ICV aim at avoiding high GOR production from each interval. If it detects a GOR value above a predefined threshold, it closes the respective production zone. For more detail on the operation and optimization of the ICV for this benchmark, we refer to (BOTECHIA *et al.*, 2021).

During the history-matching process, each well operates with measured rates as boundary conditions. The total liquid rate represents this condition for each producer, while the injected water or gas rate plays the same role for each injector. It is worth mentioning that all the ICVs must reflect the same states as the actual operation during this period. Failing to report the valve restrictions in the forward simulation will act as a modeling error, which may impair the parameter calibration. Although not considered in this work, mechanical failure may be represented as an uncertain attribute and included in the data assimilation workflow (EVENSEN, Geir, 2019).

Besides those boundary conditions, it was vital to limit the pressure during the history-matching to avoid unphysically high or low values in any model. One of the benefits of preventing unphysical conditions in the simulation models is avoiding numerical problems related to exaggerated gas liberation in the porous media due to extremely low pressures. When one sets these pressure limits, it is crucial to avoid restricting the pressure at levels too close to the measured bottom-hole pressure (BHP), as this condition will tend to conceal productivity or injectivity mismatches. One way to do so is to check each well's BHP NQDS, using the respective pressure limit as the simulated data. All wells should exhibit a relatively high normalized quadratic deviation in this test, above 10, for instance.

It is essential to include the wells and platform operational constraints to simulate the forecast after the history-matching process. We present these constraints in Table 5.4. It is worth mentioning that the rate boundary conditions indirectly apply these constraints during the history matching. For instance, in this application, the platform operates at the maximum produced gas rate during most of the history period.

Table 5.4: Sector 1 operational constraints (CORREIA *et al.*, 2020).

Constraint	Value	Applies to
Maximum oil rate (m^3/d)	28617	Platform
Maximum liquid rate (m^3/d)	28617	Platform
Maximum water production rate (m^3/d)	23848	Platform
Maximum water injection rate (m^3/d)	35771	Platform
Maximum gas production rate (m^3/d)	12 million	Platform
Minimum BHP (kPa)	50000	Producer well
Maximum BHP (kPa)	75000	Injector well
Maximum liquid rate (m^3/d)	8000	Producer well
Maximum gas injection rate (m^3/d)	4 million	Injector well
Maximum water injection rate (m^3/d)	10000	Injector well

The recovery strategy in this benchmark case assumes that the injectors reinject all the produced gas in the reservoir. Furthermore, each injection well operates in WAG cycles of 6 months, except for well I16, which only injects gas. We control the total water injection rate to maintain the average reservoir pressure at a target value of 61000 kPa (BOTECHIA *et al.*, 2021).

5.4.2 Uncertain parameters

We apply the data assimilation workflow presented in Figure 5.1 to calibrate the reservoir simulation models of the UNISIM-III Sector 1 benchmark case. The simulation models represent partial information of the fine-scale reference model, and they carry geological uncertainties regarding the permeability fields at the three main directions and the porosity field. We represent all the four grid-related parameters using an ensemble of 100 models. Furthermore, we apply a logarithmic transformation to the permeability to compute the data assimilation, the fourth step of Figure 5.1. This transformation aims at approximating the problem to a linear Gaussian one since the permeability commonly has a log-normal distribution.

The benchmark case also considers 47 scalar parameters, $x \in \Re^{1\times1}$. We apply a logarithmic transformation to all these scalar parameters during the calibration process. Firstly, there are three transmissibility multipliers related to the three faults presented in Figure 5.2a. We represent the faults using a uniform distribution in the transformed domain, from closed to fully opened or equivalently from nearly zero transmissibility to 100% transmissibility. Secondly, each producer has uncertain productivity at each of the three intervals, represented by a multiplier with a gaussian distribution, centered at 1 in the transformed domain. The same occurs in each injector but with two intervals. These intervals relate to the ICV operation.

Finally, the simulation model assumes three categorical variables, the relative permeability curves of each zone (Figure 5.2b) and the equation of state. SEnRML, with local analysis, handles these variables by representing them with auxiliary continuous variables with a prior standard normal distribution. The algorithm selects thresholds to determine the discrete levels based on the prior probability of each category. During the calibration process, the analysis equation updates each model auxiliary variable to determine the calibrated categories. Note that one should order each variable's categories in a way that enables a monotonic tendency between the discrete levels and the impacts on the measured data. For instance, the relative permeability curves should cause an increasing or decreasing water and gas production among the categories. This procedure aims at reducing the nonlinearity of the relation between the parameters and the simulated data.

5.4.3 Observed data

We assimilated the production and TLS data during the first 2161 days of the UNISIM-III Sector 1 operation. We included the oil, gas, water rates, and BHP using a 30-day periodicity for each producer. Moreover, each injector provided the BHP and the individual rate, gas or water, depending on WAG cycles. Each measured data contains uncorrelated noise with a standard deviation of 10% for the rates and around 2% for the pressures. In a practical application, one must treat the correlated measurement errors properly when there are combinations of measurements to estimate individual well rates (EVENSEN, Geir; EIKREM, 2018).

Our seismic data set comprises one baseline acquisition, after 608 days of the extended well test (MOCZYDLOWER, Bruno; FIGUEIREDO JUNIOR; PIZARRO, 2019;

NAKANO *et al.*, 2009), and two monitors, in times 1583 and 1948 days. It is worth mentioning that the UNISIM-III benchmark considers permanent seismic monitoring, with one monitor every six months after this. The analysis regarding the seismic benefits with an increasing number of acquisitions will be the theme of a future study. Furthermore, it is essential to consider in the analysis that, depending on the WAG-cycles and TLS acquisition dates, physical effects with opposite influences, hardening and softening, may attenuate the time-lapse signal.

We generated the observed data applying a petroelastic model in the fine-scale reference model results. We represented the flow-related variations using each monitor acoustic impedance ratio, monitor 1 divided by the baseline, and monitor 2 divided by monitor 1 in the present work. After that, we extracted maps considering the vertical resolution of seismic data for this case, a permanent monitoring system, and performing a scale transference to the simulation scale. At this scale, each seismic map corresponds to five to ten model layers. In total, we worked with 30 time-lapse seismic maps, 15 for each pair of surveys, with a total of roughly 20000 active data points. One should note that assuming one new survey every six months, the number of data points will reach the order of 10^5 in 4 years from the second monitor's date only for Sector 1. Furthermore, this number will significantly increase when one includes other sectors. This analysis motivated the choice of a method that is suitable for bigger data sets. We consider the seismic data in this application as a quasi-ideal case since the complexities related to wave propagation are not present.

Aiming at describing the data set main characteristics, we present four TLS maps with both monitor times in Figure 5.3. Note that the bluish regions correspond to a hardening effect and the reddish areas to a softening. With around 4% impedance variation, the most intense anomalies relate to a porous pressure increase with gas substituting water or oil. We see the combination of these softening effects in the hot regions in the north of a2 and a3, maps 3 and 5. The bluish regions in these two maps indicate a pressure reduction around producers, causing around 2% variation on the impedance ratio. The water injection already started at four injectors located in the northern area, but the high pressure and gas saturation still conceal its effect. One can observe gas substituting water and oil anomalies with minor pressure effects in a4, map 12.

Focusing on the variation between the two monitors, Figure 5.3 row b, it is noticeable in maps 1 and 3, b1 and b2, pore pressure variations effects causing almost 2% variations in the acoustic impedance. In map 5, b3, besides the pressure-related variations, there is an apparent WAG anomaly, with water replacing gas, causing a nearly 3% impedance increase. There is a similar anomaly in map 12, b4, where a gas front advance also occurs in the north-east.

In the data assimilation process, we perturbed the seismic data using Gaussian noise with 0.7% standard deviation and 600 m exponential correlation length. We estimated the correlation length based on synthetic seismic modeling that considered the noise characteristics expected in permanent reservoir monitoring surveys. Furthermore, the standard deviation allowed balanced cost-function contributions between seismic and production data in the prior ensemble.

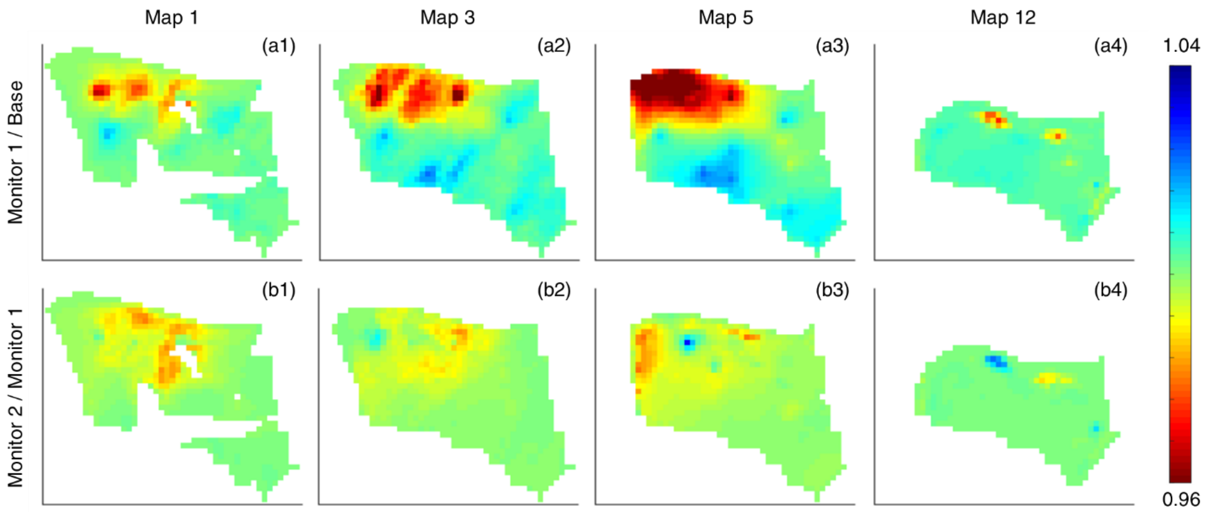


Figure 5.3: Time-lapse seismic observed data. The figure depicts four TLS maps, 1, 3, 5, and 12, in columns 1, 2, 3, 4, respectively. Furthermore, it presents the acoustic impedance ratio of monitor 1 divided by the baseline in row a and monitor 2 divided by monitor 1 in row b.

5.4.4 Petroelastic model

The petroelastic model applied in this work, the third step of Figure 5.1, considers a mixture of three minerals in the rock, calcite, dolomite, and quartz, with fractions of 85%, 11%, and 4%, respectively. The dry-rock properties as a function of effective pressure were logarithmic functions fitting laboratory data regarding a pre-salt field, as detailed in (COSTA *et al.*, 2016; SILVA *et al.*, 2020). Declining exponential functions matching data from a pre-salt reservoir represented the dry rock moduli as a porosity function. We considered different pressure and porosity laws for each reservoir zone, stromatolites and coquinas. Vasquez et al. (2019) and Silva et al. (2020) reported that the Gassmann equation (GASSMANN, 1951; apud MAVKO; MUKERJI; DVORKIN, 2009) provides a reasonable representation of the fluid substitution in reservoirs similar to ours. Therefore, we apply the Gassmann equation to the current PEM. The Batzle and Wang (1992) equations represent the water phase bulk modulus, while the fluid model described in section 5.3.1 (SILVA NETO, Gilson M. *et al.*, 2020) enables computing the gas and oil bulk moduli and densities.

5.5 Results and discussion

We divide the analysis of this research results into five sections. We start by verifying the well and seismic data match because the model ensemble must honor the past information before predicting the field behavior. After that, we analyze the SEnRML algorithm convergence when only well data and when well and seismic data are available. In the fourth subsection, we compare the well-rates forecast in the different data assimilation numerical experiments. Finally, we discuss the localization strategy in the last subsection.

For the results reported in this section, we ran SEnRML, with local analysis, using declining step-lengths, starting with 0.5 and ending with 0.1. Evensen (2019) applied a similar declining function when assimilating well data in a reservoir flow model with global analysis. The stop criterium was a maximum of 15 iterations. Nevertheless, we noticed that only minor changes occur after seven iterations. In future practical applications, it is worthwhile to include stop criteria based on parameter changes and cost function reduction throughout the iterations to avoid unnecessary flow simulations. We represented the covariance matrix of measurement errors with 1000 error samples in the matrix $\boldsymbol{E}$. The ESMDA, which serves as a quality reference for well data match, ran with eight data assimilations with constant inflation factors. We applied the ESMDA algorithm reported by Emerick (2016).

5.5.1 Well-data match

We use the NQDS metric defined in section 5.3.3 to evaluate the well data match, as depicted in Figure 5.4. The figure includes the prior ensemble, the ESMDA assimilating well data results, and the SEnRML results after assimilating well and well jointly with TLS data. For simplicity, we are going to call these three cases ESMDA, SEnRML well, and SEnRML TLS, respectively. Note that the ESMDA results represent a reference of the well-data match quality, as it is considered a standard option for production data history matching. In Figure 5.4, the boxes represent the models between the 25th and the 75th percentiles for the designed variable. The horizontal line inside the box is the median, while the dashed vertical lines show the distribution range. The points are outliers, corresponding to values whose distance to the 75th and 25th percentiles are greater than 1.5 times the difference of those percentiles.

In this analysis, we considered the tolerance as 10% of the rate measurements, the constant for the rates are 40, and the pressure tolerance is 3%.

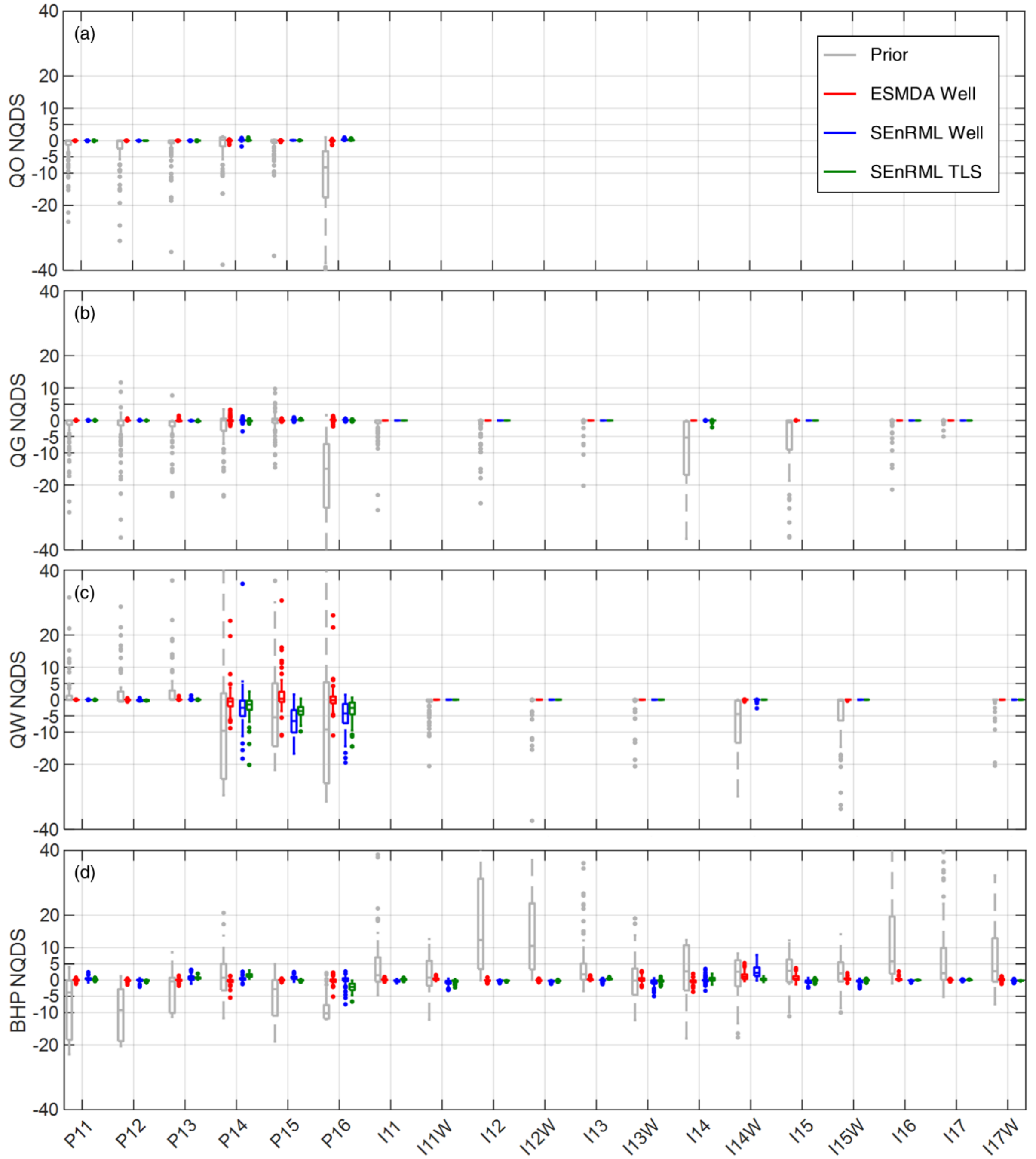


Figure 5.4: NQDS metric of all wells: (a) oil rate, (b) produced or injected gas rate, (c) produced or injected water rate, and (d) bottom-hole pressure. The WAG injectors appear twice. For instance, I11 is the injector injecting gas, and I11W is the same injector operating with water.

The oil rate, Figure 5.4a, and the gas rate, Figure 5.4b, are not critical parameters for the well data match. All methods were able to match these measurements, as all NQDS values are close to zero. These variables' prior results are already acceptable for most wells. A good prior match occurs because the liquid rate is a boundary condition for the producers, while the injected rate plays the same role for the injectors. Moreover, the produced water rate is much lower than the oil rate during the history period. One exception is P16, for which the initial ensemble underestimates the oil and gas production. P16 cannot honor the boundary condition in the prior ensemble due to the pressure constraint, which we included to avoid numerical problems related to exaggerated gas liberation, as discussed in section 5.4. Other exceptions are the injectors I14 and I15, whose rates also appear lower than expected initially. They also reach pressure limits, which we included to avoid unphysical high-pressure values. These difficulties of the prior ensemble in honoring the boundary conditions indicate a pessimistic bias in the transmissibility around these wells, causing flow compartmentalization. It is worth mentioning that uncertainty regarding reservoir connectivity is expected in this type of reservoir (DE SANT´ANNA PIZARRO; BRANCO, 2012; MOCZYDLOWER, B.. *et al.*, 2012).

The water rate, displayed in Figure 5.4c, was the most challenging variable to match for the producers based on the NQDS metric. One of the reasons for this difficulty is that the produced water rates are low compared to the oil rates. Therefore, the relative tolerance for this variable is lower. For instance, the SEnRML well case underestimated the water rate of P15, as shown in Figure 5.5. We mention P15 because it was the worst well data match that we obtained. As the liquid rate is a boundary condition, an error in the water rate leads to an equivalent error in the oil rate, with opposite direction. Nevertheless, the impact of this error is minor in the latter due to the variable magnitude (see Figure 5.5b).

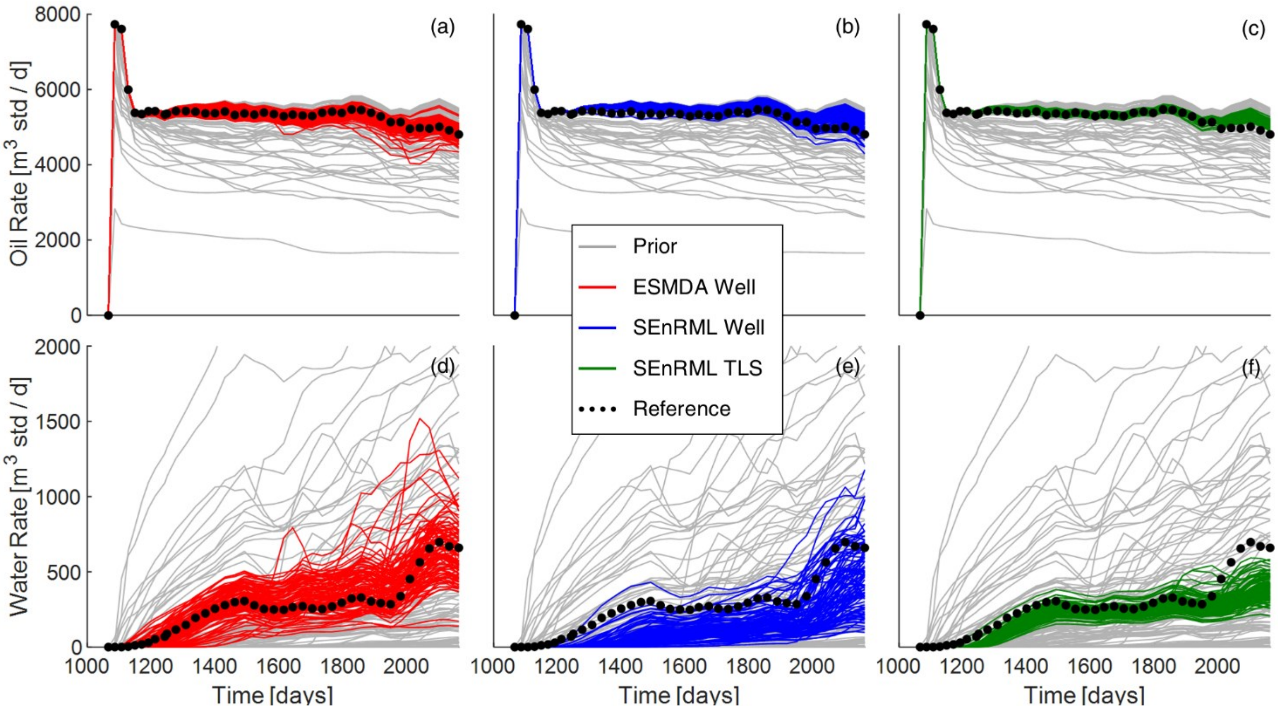


Figure 5.5: Oil rate (a, b, and c) and water rate (d, e, f) of P15. The figure shows the prior, ESMDA, SEnRML well, and SEnRML TLS results, respectively.

For the water production representation, the ESMDA results were better than the SEnRML when assimilating only well data. The latter method reduced the water-production-related cost function, but it did not reach values as low as the former. We believe that this result is related to a similar issue reported by Chen and Oliver (2013) related to the approximation of the sensitivity matrix to compute the model updates in more nonlinear problems. Evensen et al. (2019) suggested reducing the step size (γ) to overcome instabilities and improve the cost function reduction throughout the iterations. However, we could not improve the results beyond the ones reported here. The limit, in this case, is a tradeoff between the step size and the number of required iterations.

The water rate match quality improved for the SEnRML when we added the time-lapse seismic data to the data assimilation problem. As the seismic provides spatially rich information related to the reservoir fluid distribution, it helped the method reach lower water rate NQDS values, as depicted in Figure 5.4c and Figure 5.5f for P15. This result was closer to the ESMDA assimilating only well data, our reference for history-match quality.

One must note that the NQDS metric used for history matching quality assessment is mathematically different from the data assimilation cost function defined in equation (5.5). In the IES cost function, the inverse of the measurement noise covariance matrix normalizes the quadratic distance from the simulated to the observed data. If one assumes that this matrix is diagonal, which is a common choice for production data, this is equivalent to dividing each measurement by the respective standard deviation before computing the distance. Since we defined the standard deviation as 10% of the observed data, each denominator is different. In NQDS, equation (5.13), all data of the same type have the same denominator. This difference explains the result of Figure 5.6 in comparison with Figure 5.4c. We computed the cost function in Figure 5.6 using

$$\mathcal{L}_{l,j} = \sum_{l} \frac{\left(\boldsymbol{d}_{l,j}^{sim} - \boldsymbol{d}_{l,j}^{obs}\right)^2}{\sigma_l^2}, \tag{5.14}$$

where $\mathcal{L}$ is the data cost function, l indexes a specific subset of the data, the produced water rate in Figure 5.6, and σ^2 is the variance of each measured data point. Note that the SEnRML method assimilating well and TLS data provided water rate cost function slightly lower than ESMDA (Figure 5.6). Nevertheless, the absolute value of the NQDS of P14, P15, and P16 in Figure 5.4c are marginally higher. This combination of results indicates that the SEnRML results are closer to the observed data for the lower rates and farther for the higher rates, as it is possible to confirm for P15 in Figure 5.5d and Figure 5.5f, ESMDA and SEnRML TLS, respectively. One should be cautious in comparing the results depicted in Figure 5.4 and Figure 5.6 since the former highlight the data match metric on a well-by-well basis, while the latter represents a global assessment, in which a better matched well can compensate another in the comparison.

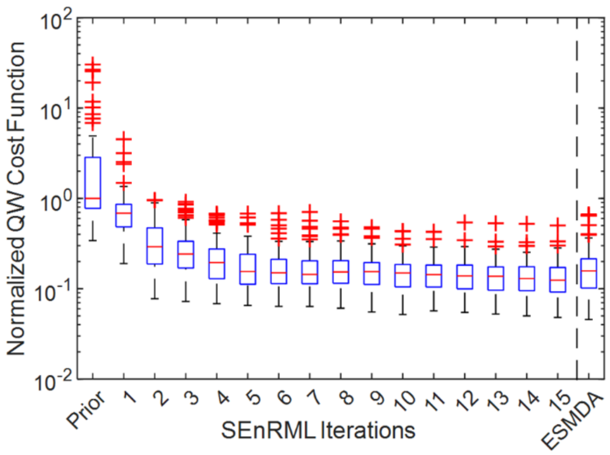


Figure 5.6: Normalized water rate cost function throughout the SEnRML TLS iterations compared to the final ESMDA result. The prior ensemble's median cost value is the normalization factor.

We evaluate the bottom-hole pressure match using the results in Figure 5.4d. Most producers exhibit a significant pressure deviation with an underestimation tendency in the prior ensemble, while most injectors show expressive deviation with the opposite sign. This information indicates that the initial models are somewhat compartmentalized. In other words, effective transmissibility between injectors and producers is lower than in the synthetic truth (fine-scale reference model). The relatively poor communication between the wells is related to two characteristics of the prior ensemble. Firstly, the transmissibility multiplier across the three faults in the model is low for most of the models, as the median value is 10^{-4}. Secondly, the permeability field is anisotropic and heterogeneous, with very low permeability values for most cells. Despite the prior ensemble's behavior, the three data assimilation examples provided a good match for all wells' bottom-hole pressure, as depicted in Figure 5.4d. The well data managed to provide enough information to correct the connection degree between producers and injectors.

In summary, the SEnRML method managed to provide well data match comparable to ESMDA for most of the measurements. One exception is the water rate match of three wells, which was poorer with SEnRML when assimilating only well data. The inclusion of TLS data in the data assimilation with SEnRML did not jeopardize the well-data match. On the contrary, it improved the overall water rate, providing a cost function for the well data slightly lower than the ESMDA case, which is the reference for the well-data match in this work.

5.5.2 Seismic-data match

We evaluate the seismic data match using the sum of the cost function terms related to this information source, equation (2.3), in Figure 5.7. We notice a significant deviation reduction in the first four iterations and minor reductions afterward. The SEnRML method reduced the cost function to less than 20% of the initial median value. We also compare this result with the SEnRML well case. To do so, we took the final ensemble from this case, ran a forward seismic simulation, and computed the deviation from the seismic measurements. It is possible to note that the well data provide indirect information regarding the pressure and fluid spatial distribution, improving the TSL cost function compared to the prior. Nevertheless, this information is limited, as the well data is scarce in space. Therefore, the well normalized TLS cost function is roughly 35% of the initial cost function, but it is about twice as higher as the TLS cost function when the SEnRML method integrated TLS data.

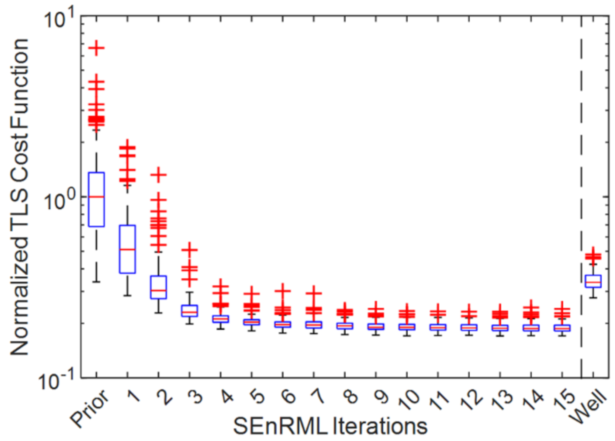


Figure 5.7: Normalized TLS data cost function throughout the SEnRML TLS iterations, compared to the final SEnRML well result. The prior ensemble's median value is the normalization factor.

It is also possible to evaluate the seismic data assimilation by comparing the maps of Figure 5.8, where we show for three maps and two monitors: the observed data, the results of the prior ensemble, the calibrated ensemble of the SEnRML well data assimilation, and the SEnRML TLS data assimilation. Note that in the latter, the well data is also present. The first general characteristic we note is the compartmentalization of the prior ensemble, which is noticeable from the discontinuities of the maps of Figure 5.8 column 2, especially b2. Furthermore, this compartmentalization is not present at the same level in the observed data of column 1 nor the calibrated ensembles of columns 3 and 4. Therefore, it is possible to conclude that both well and TLS data provide information to improve the models' connectivity compared to the synthetic truth. For instance, the north fault (Figure 5.2a) that causes the apparent discontinuity in maps d2 and e2 behaves differently in the calibrated ensembles d3, d4, e3, and e4. The same occurs for the intermediate fault in e2. Nevertheless, the available information from wells and TLS was not enough to change the south fault behavior, which impairs transmissibility in maps a3 and, to a lesser extent, a4.

In a practical application, we recommend that the differences in compartmentalization between the observed data, Figure 5.8 column 1, and the prior results, Figure 5.8 column 2, should lead to a new prior geological modeling. One of the likely changes is the fault transmissibility multiplier distribution, which should admit open faults with a more pronounced probability. The new prior should also exhibit an improved communication through the permeability field than the former. This process would bring the flow dynamics closer to the observed seismic, even before the data assimilation process. Furthermore, starting with an ensemble that fully represents the available data interpretation, the Bayesian data assimilation process would lead to an ensemble that better describes the reservoir flow dynamics.

Nevertheless, the prior ensemble redefinition in a new geological modeling process is beyond the present work scope.

Another relevant piece of information from Figure 5.8 is the difference between the SEnRML well and SEnRML TLS cases in columns 3 and 4. As expected, the TLS provides richer data regarding the spatial distribution of pore pressure and fluids in the porous media. Therefore, we expected that the ensemble of models calibrated only with well data could result in system behavior far from the wells that are different from the synthetic truth. It is possible to notice this difference in the west and south regions of map a3, compared to the observed data in a1. However, one could also note similarities between the two calibrated ensembles that justify the cost function reduction of the SEnRML well case compared to the prior in Figure 5.7. For instance, the main characteristics of maps f3 and f4 are similar to each other and the observed map f1.

Figure 5.8 also allows us to verify which kind of information the TLS can provide when we have a carbonate reservoir with volatile fluid at high pressure. Based on our petroelastic model results described in section 5.4.4, it is possible to observe a WAG anomaly, the bluish changes surrounded by a hot area in the maps of row e. These maps represent injected water after a period of injecting gas in the porous media. Furthermore, pressure-related anomalies, combined with localized increased gas saturation, are also present, as one can observe in the softening areas of the b maps. The water-related hardening effect increased the impedance by 3%, while the pore-pressure and gas softening effect decreased the acoustic impedance by 4%.

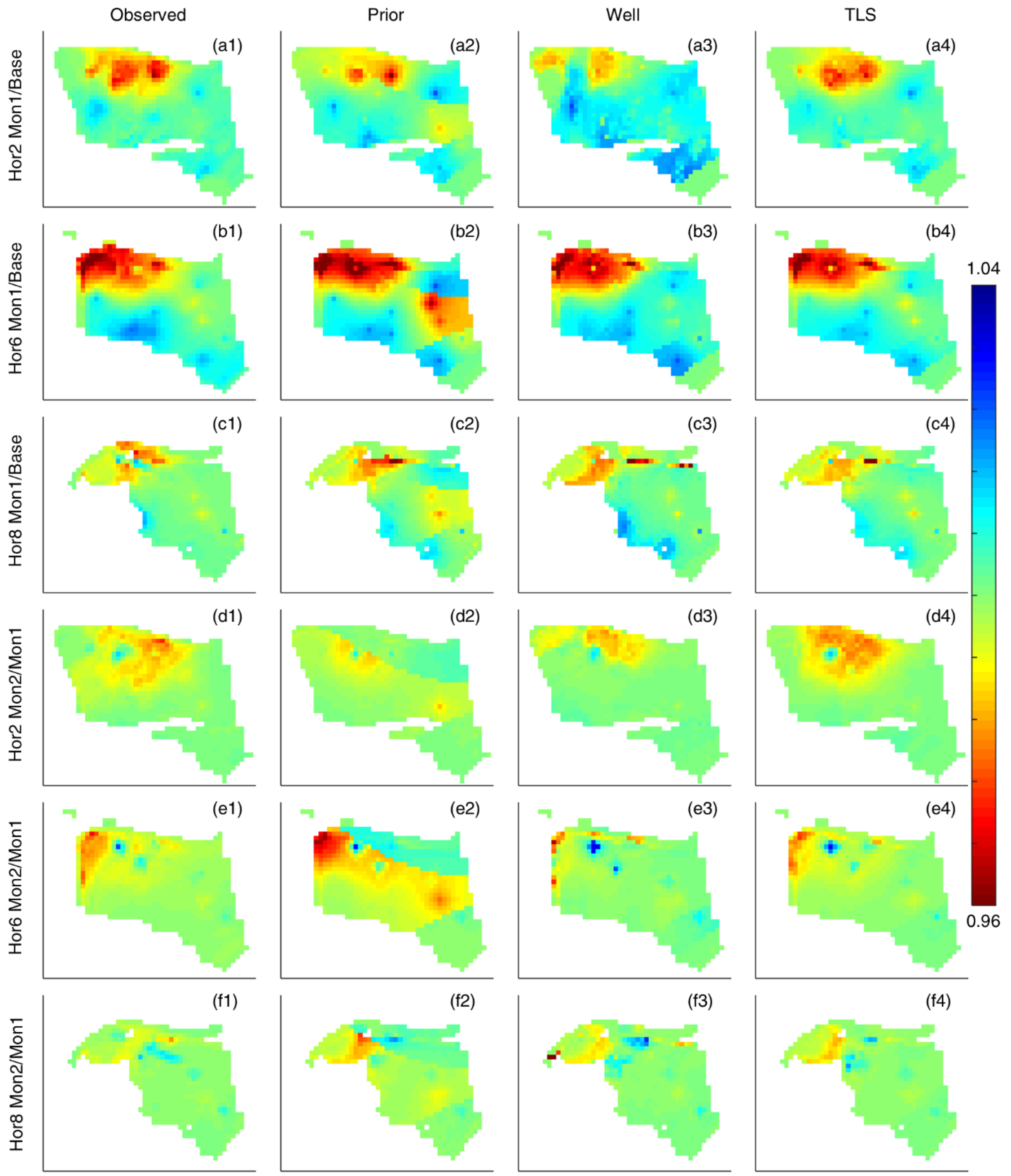


Figure 5.8: Time-lapse seismic maps, acoustic impedance ratio, of horizons 2 (a and d), 6 (b and e), 8 (c and f), monitor one divided by baseline (a, b, and c) and monitor two divided by monitor one (d, e, and f). The figure compares the observed data (1), the prior ensemble mean (2), the calibrated ensemble mean of SEnRML wells (3), and SEnRML TLS (4).

In summary, the SEnRML method managed to use the TLS information to improve the models' capabilities to describe the pore-pressure and fluids distribution in the porous media. Using the well data alone, one could improve the models' general behavior, but some discrepancies remain when comparing their results to the TLS data from the synthetic truth, especially in regions far from the wells. It is worth mentioning that the most appealing advantage of SEnRML in this application is the capability of handling bigger data sets originated from the TLS acquisitions, which we assimilate simultaneously with production data.

5.5.3 Convergence speed and cost-function behavior

The SEnRML method is an iterative ensemble smoother. As such, it differs from the classical ensemble smoother (EVENSEN, Geir, 2009; EVENSEN, Geir; VAN LEEUWEN, 1996) because it makes parameters updates in steps, which are usually smaller than the ensemble smoother's single time update. In this context, it is necessary to check in which iteration the method stopped to perform a significant change in the models' behavior, an indication of convergence. Obviously, the sooner we stop data assimilation, the lower the data assimilation process computational cost. Nevertheless, it is vital to ensure that we do not impose a premature stop, jeopardizing the data match.

We show the data cost function, equation (2.3), of each ensemble member throughout the iterations in Figure 5.9. We compare the results of two data assimilation cases, the first assimilating only well data and the second well and TLS. If we consider a stop criterium based on the reduction level of the average cost function (CHEN, Yan; OLIVER, 2013), it is possible to note that both cases would stop around iteration 8. Moreover, in the two cases, the calibration process reduced the cost function by roughly one order of magnitude.

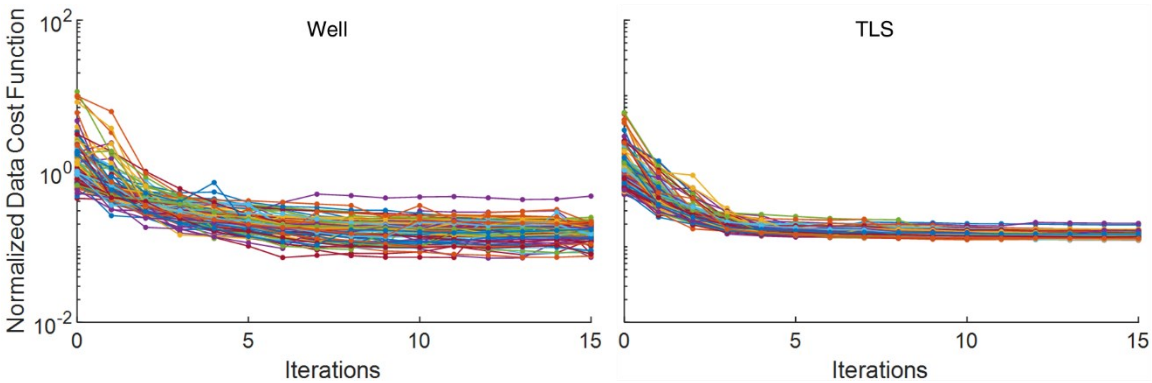


Figure 5.9: All the ensemble members' data cost function throughout the iterations of SEnRML, assimilating well and well jointly with TLS data. In each graph, the prior median value is the normalization factor.

Based on the analysis of Figure 5.9, it is possible to note that the presence of TLS data changed the problem in a way that allowed a more monotonic reduction of each ensemble member's data cost function throughout the iterations. We believe that there are at least two possible reasons for this behavior. Firstly, the TLS data seem to have a relation with the parameters that are closer to linear. Furthermore, it provides complementary information to the well data, reducing the tendency of occurring multiple minima close to each other in the parameter space. In other words, it reduces the ill-posedness of the problem.

When only well data was available, some models presented oscillating cost functions at some iterations. One could reduce this tendency by reducing the SEnRML step-length parameter (EVENSEN, Geir *et al.*, 2019). However, it is vital to keep in mind that more iterations will be necessary if the step length is too low. In this context, we believe that the results presented in Figure 5.9 are a tradeoff between the importance of fast convergence and allowing stable results for most of the models. The ESMDA algorithm also exhibited oscillating data cost functions of models during the multiple data assimilations of the well data, as depicted in Figure 5.10. These results corroborate the complexity and nonlinearity of this problem.

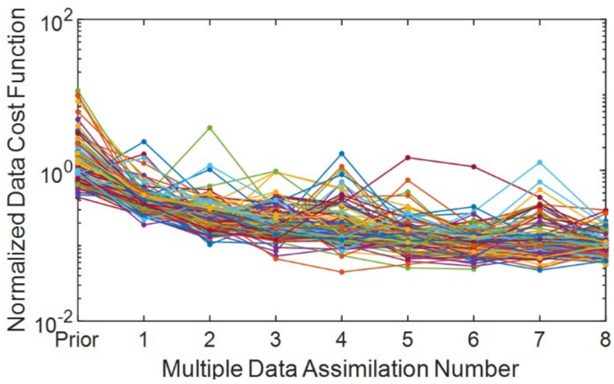


Figure 5.10: All the ensemble members' data cost function throughout the multiple data assimilations of ESMDA, assimilating well data. The prior median value is the normalization factor.

5.5.4 Well rates forecast

Most of the importance of having calibrated reservoir simulation models comes from their application in production forecasts to support the decision-making process. Therefore, it is interesting to evaluate the data assimilation results in predicting field behavior. Since this is a synthetic application, we can compare the calibrated model responses with the synthetic truth, which we commonly call a reference solution. It is worth mentioning that we apply a bottom-hole pressure lower limit for all wells during the forecast. The platform capacities impose an overall limit for each produced and injected rate. Furthermore, we employ the same control rules to the intelligent control valves to uniformize the injection profile in the reservoir zones and avoid excessive gas or water production. For more detail regarding these strategies, we refer to (BOTECHIA *et al.*, 2021).

We show the forecasted overall field oil, gas, and water rates in Figure 5.11. The two data assimilation cases are present: SEnRML well and SEnRML TLS. In terms of overall field behavior, the results are relatively similar. As expected, SEnRML TLS results, column 2, exhibit less variability due to the higher amount of information available during the calibration process.

In this application, the gas rate limits the oil rate due to the gas-oil ratio and the platform capacity to process the produced gas. Therefore, we observe the constant gas production rates in Figure 5.11, row b. High gas production is a common bottleneck in pre-salt reservoirs (DE MORAES CRUZ *et al.*, 2016; DEPLANTE *et al.*, 2019). Both calibrated ensembles seem to capture this behavior until around day 8000. Nonetheless, some models exhibit declining gas rates at the end of the forecast period in both cases, but more severely in the SEnRML well example. This discrepancy in some models occurs after around day 8000, which seems plausible since we only used data from the first 2161 days of production. The TLS data assimilation seemed to alleviate this discrepancy, but it could not avoid it for some models.

Both calibrated ensembles exhibit a declining oil rate whose tendency is similar to the reference in Figure 5.11, row a. However, there is an optimistic bias that the available TLS data did not mitigate. The GOR defines the oil production rate, and we associate this bias with the lack of data regarding the gas front advance in the porous media until the end of the history period. All wells exhibit relatively low gas rates during this period, with a GOR of less than 510 $m^3\ std/m^3\ std$. Furthermore, the gas-related anomalies in the TLS data are close to the injectors, providing limited information about how this fluid will advance in the reservoir.

Although the most challenging data to match during the history matching, the water rate does not play a crucial role during the forecast. Despite the increasing values during the history period, the forecasted water production remains lower than the oil production. In general, both calibrated ensembles capture this behavior (see Figure 5.11c). The SEnRML TLS case provides results closer to the reference, with a significant uncertainty reduction compared to the SEnRML-well case. The relatively worse results for the water rate match of SEnRML well did not jeopardize this ensemble's capability to provide forecast results comparable to the SEnRML-TLS case. This similarity corroborates that the wells' water rates during the history matching are not critical to describing future field behavior in this application.

One relevant aspect of the results presented in Figure 5.11 is how we define reservoir management during the forecast. Firstly, we impose the platform limits of Table 5.4, of which the maximum gas production rate is the most important. This limit will cause well rate restrictions throughout the forecast. Secondly, each injector ICV will balance the fluid volumes injected in each interval. Thirdly, each producer ICV will shut in any interval that produces gas above a predefined GOR limit (BOTECHIA *et al.*, 2021). It is important to highlight that the management rules are identical in the reference and the simulation models. Nevertheless, the well restrictions and the ICV operation will differ in each model due to differences in reservoir properties among the models, which complicates the model comparison using forecast results alone. For instance, the opened producer wells and intervals at the end of the history period differ in each model, explaining part of the gas rate discrepancies in Figure 5.11b1 and Figure 5.11b2 after day 8000.

There are cases where a minor difference in a model can lead to a substantial impact due to differences in reservoir management operations, for instance, ICV changes and well shut-in. It is possible to reduce these effects on the model evaluation. For instance, one can run the reservoir simulation beyond the history period applying the same operation sequence on all well valves. In a real field application, reserving part of the history for validation enables this analysis. Furthermore, in a benchmark synthetic study such as the current work, one could run the synthetic truth and the reservoir models without reservoir management rules, a human-intervention free simulation. Nevertheless, it is worth highlighting that both tests are appropriate for validation purposes, but they are not realistic production forecasts. While the former uses information regarding the forecast's actual reservoir management operations, the latter applies a naïve strategy, neglecting any well intervention to reduce undesired fluid production. Comparing these methods to validate and compare the models' predictive capabilities will be the focus of a future study.

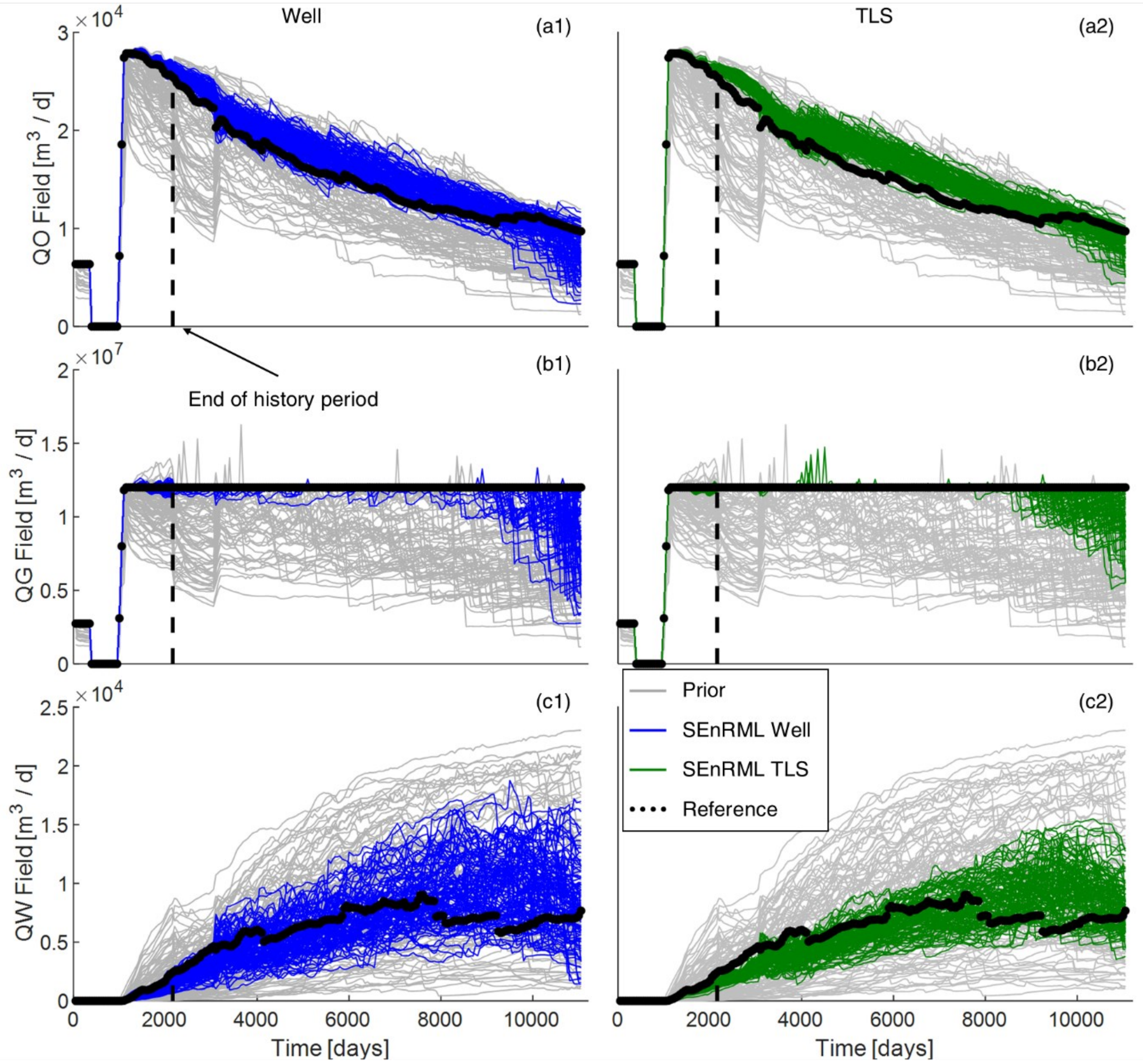


Figure 5.11: Forecast field oil, gas, and water rates (a, b, c, respectively), of SEnRML Well (1) and SEnRML TLS (2) cases.

Although it is relevant to check the calibrated ensembles' performance at the field level, the analysis of field rates presented in Figure 5.11 may be misleading. These rates are the sum of all wells' contributions, and errors of different directions attenuate each other. Therefore, it is also essential to evaluate the forecast on a well-by-well basis. Aiming at comparing the ensembles' results with the synthetic truth in detail, we computed the sum of the squared differences between the well's rates and the reference solution for the forecast period. We normalized this value by the prior ensemble's median and called it the normalized forecast cost function. Figure 5.12 depicts this variable for the oil production rates (a), gas production rates (b), and water production rates (c).

For both the oil and gas production rates in Figure 5.12a and Figure 5.12b, the well data assimilation significantly reduced the normalized forecast cost function compared to the prior. The addition of TLS data to the calibration process further reduced this metric for all models, as expected. For the water rates presented in Figure 5.12c, the data assimilation improved the responses of the worst models for this metric from the prior. Nonetheless, the best solutions in the prior ensemble are close to the calibrated ensembles' best solutions. Furthermore, incorporating TLS data does not seem to improve the water rate forecast on a well-by-well basis. This fact may be related to the fewer water-related anomalies in the TLS maps, as depicted in Figure 5.8. Moreover, three out of the six opened wells do not produce water during the history period, while the other three produce low water rates compared to the oil.

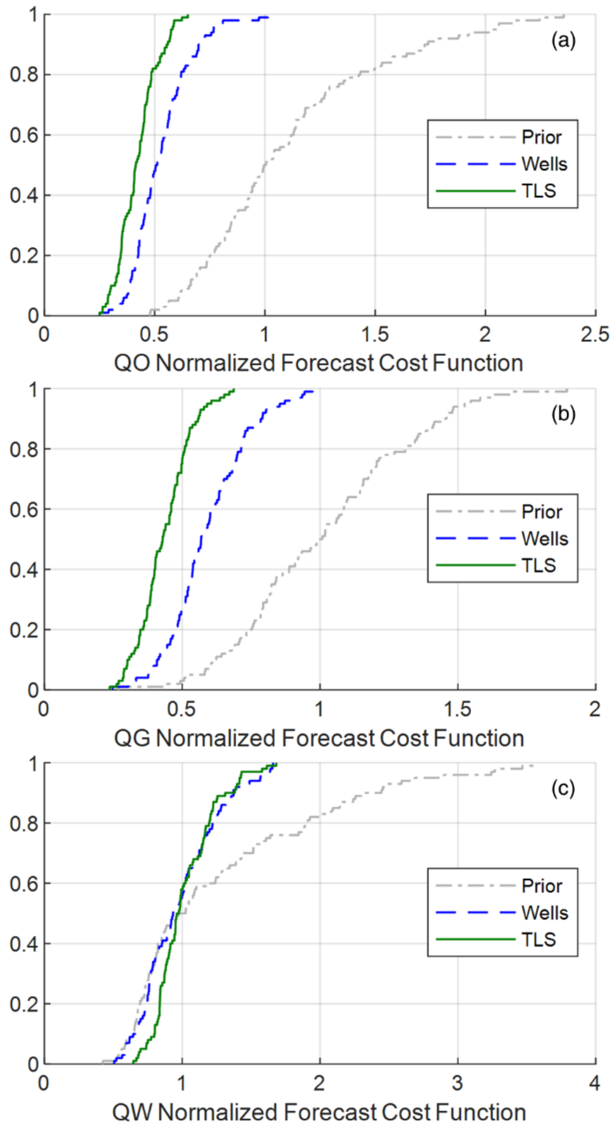


Figure 5.12: Forecast period cost function. Each graph is an experimental cumulative distribution curve, including each model's results. The figure shows the oil rate (a), gas rate (b), and the water rate (c) cost functions.

The producers P17 and P18 open after the end of the history period. Therefore, they are examples of how the models can predict a new well performance. These wells are in a central portion of the reservoir, as shown in Figure 5.2. Figure 5.13 depicts P17 predicted rates in rows a and b. The TLS case provided oil, gas, and water production forecasts closer to the reference solution with reduced uncertainty compared to the case where only well data is available. P18 oil rates in Figure 5.13c1 and Figure 5.13d1 are close to the reference, with an early pessimistic tendency in the former and optimistic in the latter. P18 gas and water production evidence the TLS benefits for this well, as the calibrated ensemble results are closer to the reference in the SEnRML TLS case. The SEnRML well results of P17 and P18 reflect the lack of spatial information regarding the reservoir properties far from the wells that open during the history. In this sense, the TLS data provides complementary information to the former.

Note that, in the reference solution of Figure 5.13, rows a and b, P17 shut in around time 4100 days due to the gas production limit. This early shut-in also occurs in the calibrated models but at different dates for each model, which results from how the reservoir management rules function in models with different reservoir properties. For instance, a model may keep the well open for a more extended period because the peak GOR was slightly lower than the limit at around time 4100 days. A similar effect in the opposite direction occurs for P18, in Figure 5.13, rows c and d. In the reference, P18 shuts in at around day 10000. Nevertheless, some simulation models cause earlier shut-in, before day 9000.

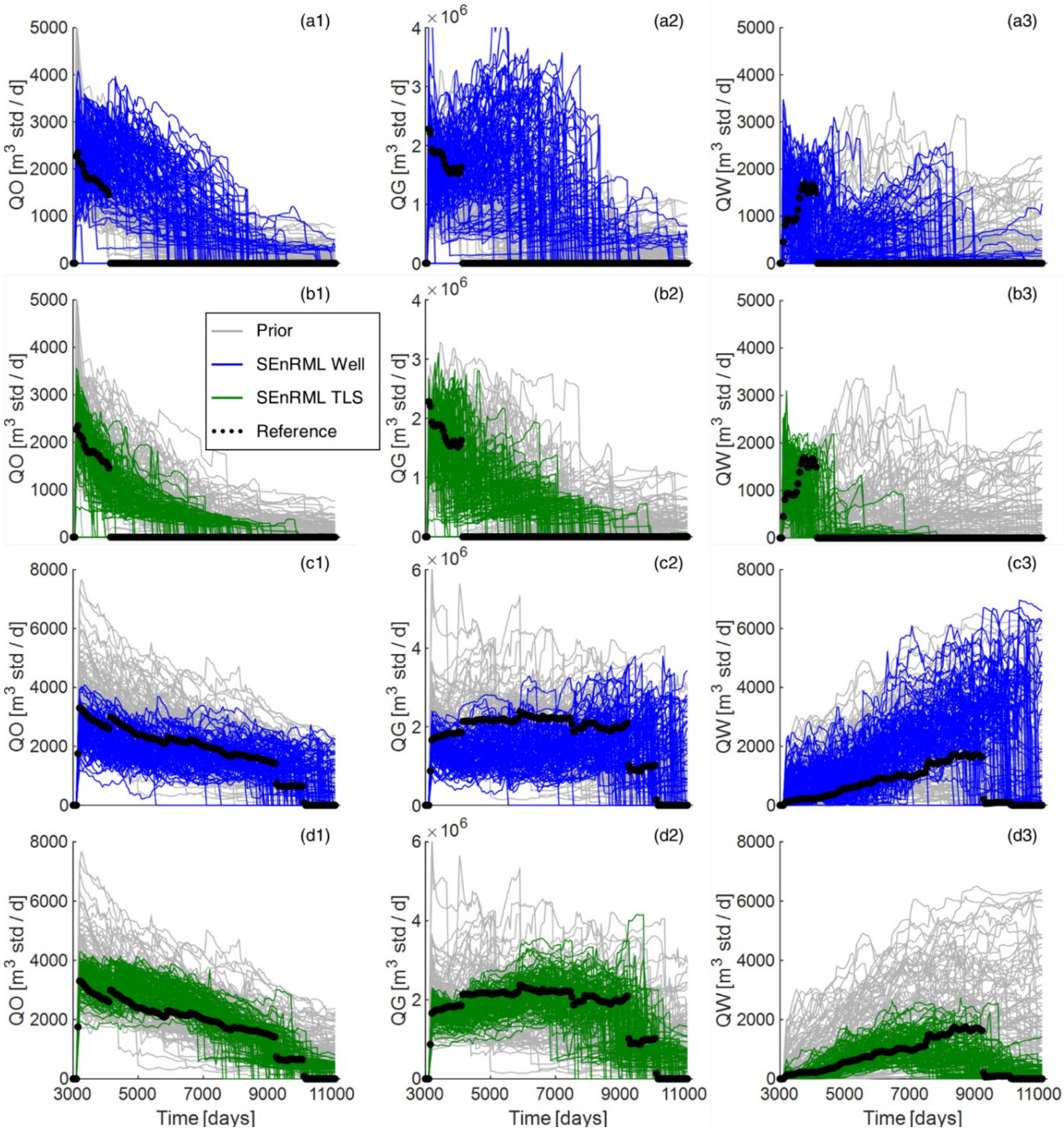


Figure 5.13: Oil (column 1), gas (column 2), and water (column 3) rate forecasts of wells P17 (rows a and b) and P18 (rows c and d). The figure compares the SEnRML well case (rows a and c) to the SEnRML TLS case (rows b and d).

The forecast results indicate that the well and TLS data assimilation using SEnRML with local analysis improved the models' capability to represent the field's future behavior. The TLS data contributed to improving the oil and gas rate forecasts on a well-by-well basis.

5.5.5 Additional analysis: localization strategy

It is crucial to employ a localization strategy to avoid exaggerated uncertainty reduction when applying an ensemble-based data assimilation algorithm with a limited ensemble size (EMERICK, Alexandre; REYNOLDS, 2011). Silva Neto et al. (2021) proposed two local analysis schemes for the SEnRML method, one based on the distance between the parameters and the data and another based on their correlation.

All sections 5.5.1 to 5.5.4 results used distance-based localization. The well-data localization lengths related to the drainage area and each well's influence region, estimated from streamlines (EMERICK, Alexandre; REYNOLDS, 2011; SOARES; MASCHIO; SCHIOZER, 2018). It is worth mentioning that ESMDA with Kalman gain localization, a reference for well match in section 5.5.1, employed the same localization metrics. Moreover, we set an arbitrary value for the TLS-data localization length of 1400 m, or seven reservoir cells.

In addition to the distance-based, we tested the automatic or correlation-based localization strategy in the present study. Figure 5.14 displays a comparison of this strategy to the distance-based scheme. It is apparent that the correlation-based localization led to worse data matches than the distance-based method. This difference was enough to impair the pressure and rates well match quality, based on the NQDS metric.

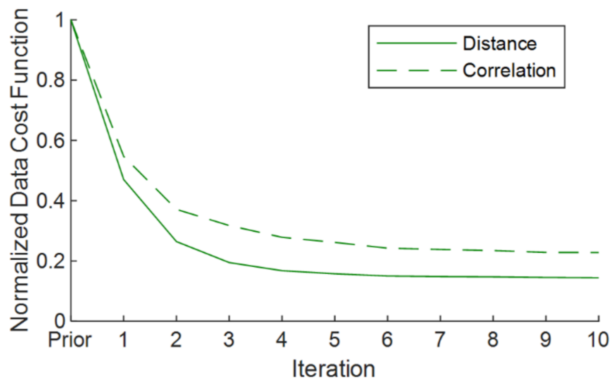


Figure 5.14: Comparison of the data cost function reduction between the distance-based and the correlation-based localization methods. Each line represents the median of the ensembles throughout the iterations.

We identified three characteristics of the present case study that led to the correlation-based localization method's worse performance. Firstly, the correlation level between the local groups and the TLS data is relatively low, resulting in data tapering below 0.5 for most of the localization matrix, even in positions close to the local group of parameters, as depicted in Figure 5.15 row a. The dot in the maps shows the local group location, while the maps relate the group to the TLS data. There are four different combinations of local groups and TLS horizons in the figure.

Secondly, the correlation between data and parameters seems to change as the data assimilation scheme updates the models. To illustrate that, we compare the localization value computed from the correlations using the prior ensemble and the final ensemble from the SEnRML Well case in Figure 5.15 rows a and b, respectively. For instance, one can note that the localization matrix includes data far from the local group to the southeast in map a1, and it is limited to an area close to the group in map b1. On the other hand, maps a2, a3, and a4 do not assign significant weight to any data around the local groups, while the maps b2, b3, and b4 do. It is worth mentioning that we expected lower correlation levels in b due to the calibrated ensemble's lower variability.

The low correlation levels and the variation during the calibration seem to be related to the highly heterogeneous reservoir, with complex flow paths in the prior ensemble, as illustrated by the permeability cross-section and maps of one arbitrary model Figure 5.16a. The figure represents the permeability in three levels, low permeability in blue, intermediate in green, and high permeability in red. The horizontal and vertical prior permeability distributions are discontinuous, which means that there are multiple regions with intermediate to high permeability, separated by low permeability cells. This pattern occurs in the vertical communication in (a1) and horizontal communication (a2).

The prior permeability distribution, jointly with the three faults transmissibility multipliers, create poorly communicated regions in the reservoir, corroborating the pressure discrepancies discussed in section 5.5.1. Furthermore, the calibration process alleviates the compartmentalization, as in Figure 5.16b1 and Figure 5.16b2, equalizing the pressure close to injectors and producers (Figure 5.4d). This process seems to change the flow dynamics and the correlations between the parameters and the simulated data. In summary, the prior ensemble is biased towards poor connectivity, which changes after the calibration procedure.

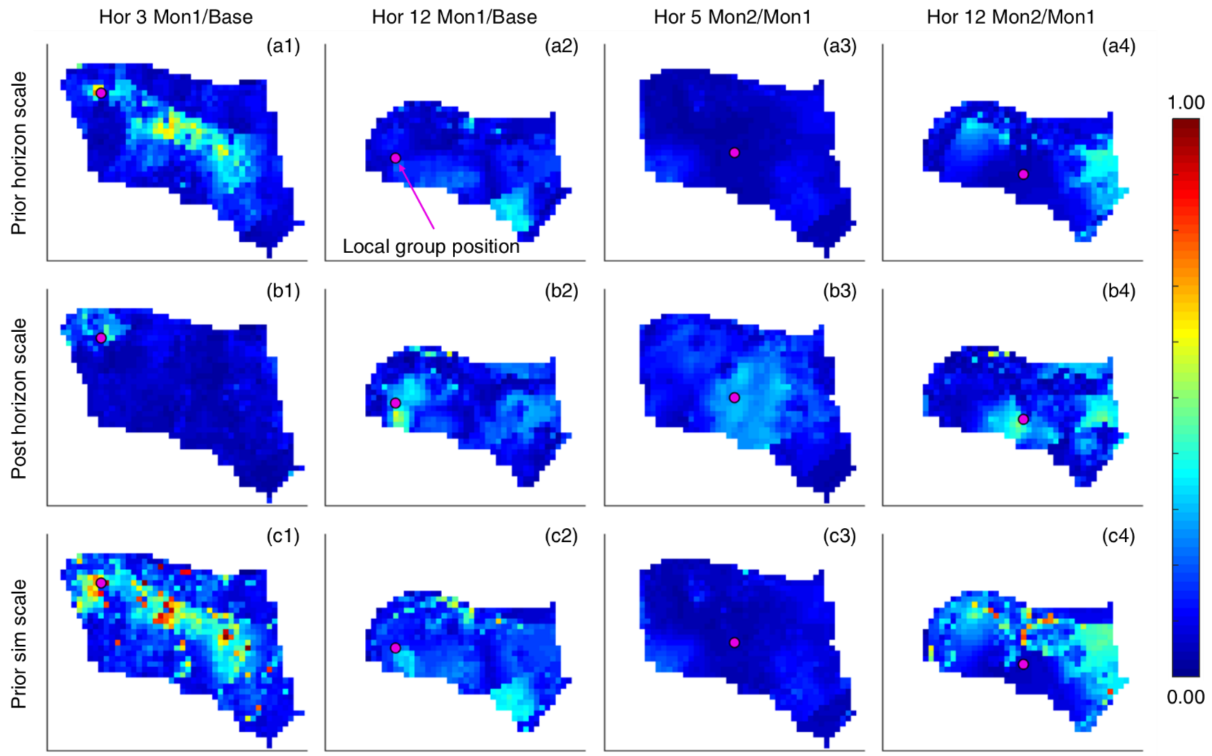


Figure 5.15: Correlation-based localization maps. The figure depicts the tapering results using the prior ensemble in the TLS horizon scale (a), the calibrated ensemble with only well data in the TLS horizon scale (b), and the prior ensemble in the reservoir simulation scale (c). It also shows four combinations of local groups and TLS horizons in columns 1 to 4.

The third reason for the low correlations in Figure 5.15 is the fact that the seismic horizons are vertically upscaled, representing the weighted mean of five to ten simulation layers for most regions. If we compare this correlation to the one computed with the simulated data in the reservoir model scale, row c, it is apparent that the upscaled data attenuated some correlations. It is worth mentioning that, for each TLS map, Figure 5.15 row c displays the highest correlation in the layers that form the map. This procedure highlighted correlated data influence on the parameters, as indicated by the higher tapering values around the group in Figure 5.15c1 and Figure 5.15c2 compared to a1 and a2. Nevertheless, it can also amplify the spurious correlation effects.

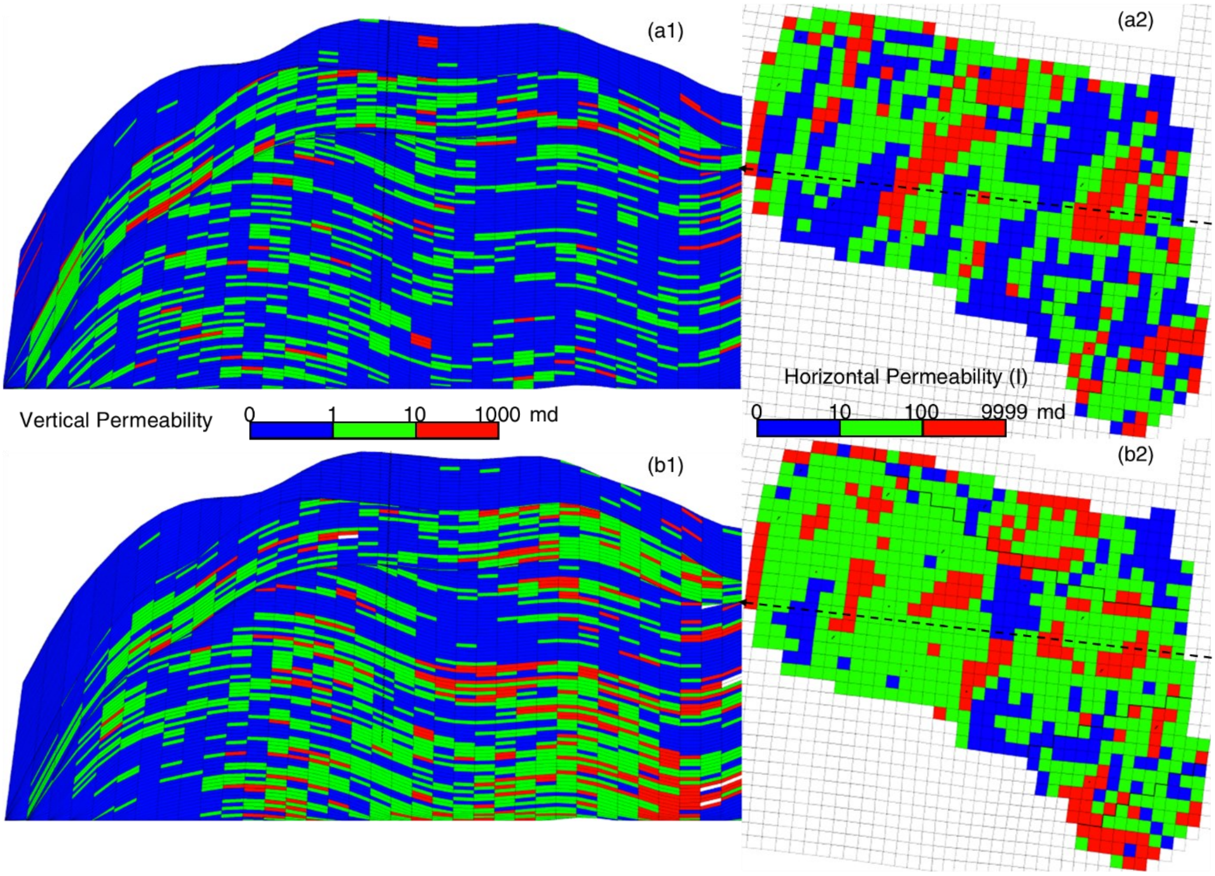


Figure 5.16: Model 80 permeability. The figure shows the prior (a) and calibrated (b) fields using well and TLS data. The vertical permeability cross-sections (J = 20) are in column 1, and the maps (K = 96) of horizontal permeability in one direction are in column 2. The dashed lines indicate the position of the horizontal and vertical cuts.

The current results indicate that the distance-based localization performance was better than the correlation-based scheme in this application. We identified some characteristics of the present study that contributed to this difference. Some of these characteristics are related to our highly heterogeneous prior ensemble, which has a general flow behavior different from the calibrated ensembles in reservoir connectivity. Furthermore, a bigger ensemble would also improve the correlation representation by lowering the statistical noise at the cost of increasing simulation time. We indicate that it is relevant to investigate different ensembles' influence on the correlation-based data tapering during iterative calibration in future studies.

5.6 Summary and conclusions

We performed well and time-lapse seismic data assimilation in a realistic synthetic case that represents challenges similar to a Brazilian pre-salt reservoir. We considered a compositional fluid model for both reservoir flow and petroelastic models, water-alternating-gas injection for enhanced oil recovery, interval control valves in the wells, and permanent seismic monitoring. We employed the SEnRML method with local analysis to assimilate the data and compared the well data match quality with the ESMDA.

The specific conclusions of this work are:

- The SEnRML method, with local analysis, managed to assimilate well and time-lapse seismic data from two monitor surveys simultaneously, leading to models that represent the production history information similarly to ESMDA assimilating well data. The former has the advantage of assimilating bigger data sets, as the permanent seismic monitoring provides multiple monitor acquisitions, with lower computational costs when compared to the latter (SILVA NETO, Gilson Moura *et al.*, 2021).
- The assimilation of time-lapse seismic data jointly with well data improved the reservoir forecast compared to the prior ensemble and the one calibrated only with production data. This result corroborates the importance of time-lapse seismic data in a pre-salt reservoir application. Moreover, it demonstrates that integrating the compositional fluid model and SEnRML with local analysis is a viable solution to take advantage of the time-lapse seismic information.
- SEnRML with local analysis and ESMDA with Kalman gain localization led data cost function oscillating throughout the iterations or multiple data assimilations for some models in well data assimilation only. Moreover, the SEnRML data match was slightly worse than ESMDA in terms of the final data cost function. However, incorporating time-lapse seismic data improved the SEnRML method's performance in terms of stable data cost function reduction throughout the iterations. Note that this is the type of problem for which SEnRML shows more advantages due to more giant data sets.
- In the present case study, the well data reveal part of the information associated with the time-lapse seismic data. The assimilation of the former significantly improves the representation of the latter in terms of the cost function.
- The assimilation of time-lapse seismic data improved the reservoir characterization in regions far from the wells opened during the history period, improving the time-lapse seismic data match and new wells production forecasts.
- Seismic simulation using a compositional fluid model provided useful information for reservoir parameters calibration in a miscible gas injection alternating water synthetic case.

- The distance-based localization performed better than the correlation-based localization in terms of data match. This result is related to the highly heterogeneous and biased prior models, whose flow dynamics differ from the calibrated ones.

We indicate four future studies derived from this work's observations. Firstly, it is relevant to test further the SEnRML method in assimilating well data in highly nonlinear problems. Secondly, it is worth investigating how to apply correlation-based localization when the correlation pattern varies during the calibration process. Thirdly, future research will evaluate methodologies to validate the predictive capabilities of reservoir models. Lastly, the seismic monitoring benefits to the reservoir characterization and production forecast, with an increasing number of monitors, will be a theme of future work.

5.7 Nomenclature

Variables

$\mathbf{1}$	Vector whose elements are equal to one
a_{PR}	Equation of state parameter
b_{PR}	Equation of state parameter
c_{PR}	Equation of state volume-shift parameter
c_T	Isothermal compressibility
$\boldsymbol{d}$	Data vector
$\boldsymbol{g}$	Forward model
m	Number of datapoints
m_{PR}	Equation of state parameter
n	Number of parameters
n_{ac}	Number of active cells
r	Correlation coefficient
v_P	P-wave velocity
$\boldsymbol{w}$	Vector of coefficients of parameters updates
$\boldsymbol{x}$	Vector of parameters
z	Element of pseudo-distance dummy variable
$\boldsymbol{A}$	Matrix of ensemble anomalies
$\boldsymbol{C}_{dd}$	Covariance matrix of measurement errors
C_P	Heat capacity at constant pressure
C_V	Heat capacity at constant volume
$\boldsymbol{D}$	Matrix of data realizations
$\boldsymbol{E}$	Matrix of measurement perturbations
$\boldsymbol{H}$	Matrix of innovations
H^{id}	Ideal enthalpy
$\boldsymbol{I}$	Identity matrix
$\mathcal{J}$	Cost function
$\mathcal{L}$	Data cost function
N	Ensemble size

N_E	Number of samples that form the $\boldsymbol{E}$ matrix
P	Pressure
R	Gas constant
$\boldsymbol{S}$	Matrix of predicted and deconditioned ensemble anomalies
$\bar{V}$	Molar volume
$\boldsymbol{W}$	Matrix of coefficients of parameters updates
$\boldsymbol{X}$	Matrix of the ensemble of parameters
γ	Step-length parameter
θ	Correlation threshold
ρ	Density
σ	Standard deviation
τ	NQDS data tolerance
φ	NQDS data constant
$\boldsymbol{\Pi}$	Projector that subtracts the mean and normalizes parameters or data matrices

Subscripts

c	Critical point
j	Ensemble member
l	Subset of the data
m	Indicates that the size is equal to the number of data points
N	Indicates that the size is equal to the ensemble size

Superscripts

a	Updated, also known as posterior or analysis result
f	Prior, also known as background
i	Iterations
obs	Observed data
sim	Simulated data
T	Matrix transpose

5.8 Acknowledgments

We conducted this work with the support of Libra Consortium (Petrobras, Shell Brasil, Total, CNOOC, CNPC, and PPSA) within the ANP R&D levy as “commitment to research and development investments” (5850.0105307.17.9) and Energi Simulation. The authors are grateful for the support of the Center of Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil), and the Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil). Also, a special thanks to CMG and Schlumberger for software licenses and Alexandre A. Emerick (from Petrobras) for providing the EHM tool to UNISIM.

6 CONCLUSIONS

This thesis addressed two critical challenges related to seismic data assimilation in reservoir simulation models: (1) to mitigate model errors and their effects and (2) incorporate big data sets using ensemble-based methods. The first scientific study presented a methodology to handle significant spatially-correlated model errors in time-lapse seismic data assimilation. The second proposed a local analysis method with the subspace ensemble randomized maximum likelihood method to assimilate seismic big data sets in reservoir models. The third work addressed a compositional fluid modeling method to simulate seismic responses. The fourth and last study integrated the two previous studies, aiming to assimilate well and time-lapse seismic data in a synthetic study resembling a pre-salt reservoir.

The first study methodology consists of a weak constraint formulation to the data assimilation problem. We include an additive error-related grid parameter in the data assimilation process, whose prior distribution relates to data assimilation residuals. The results show that model errors with long-range spatial correlation can jeopardize the data assimilation results. However, the proposed methodology avoids physically inconsistent updates and improves the reservoir characterization and production forecast using time-lapse seismic data in the presence of spatially-correlated model errors.

The second study addresses a local analysis scheme to the subspace ensemble randomized maximum likelihood method to assimilate time-lapse seismic data. The proposed algorithm is suitable for big reservoirs with big data sets. Furthermore, it can provide data assimilation results comparable to a popular ensemble-based method in the calibration of grid parameters with time-lapse seismic data. The distance-based and correlation-based schemes mitigated limitations related to the small ensemble, such as exaggerated uncertainty reduction and reduced degrees of freedom.

The third work focused on improving the seismic forward modeling, adopting a compositional fluid model to compute the bulk moduli of the oil and gas phases. The proposed method can represent volatile reservoir fluid with significant amounts of CO_2. Furthermore, since it uses the same equation of state applied in the reservoir flow simulators, one can readily integrate it with the latter in a data assimilation workflow. The results show that the presented model provides a match to experimental data similar to a correlation built for this type of fluid when only pressure-volume-temperature data is available. Moreover, it is possible to improve the model by incorporating speed of sound experimental data.

This thesis's last work represents a synthetic case study that mimics some Brazilian pre-salt challenges. It considers volatile fluid containing CO_2, miscible gas alternating water injection, interval control valves, highly heterogeneous carbonate rock, and seismic monitoring with multiple monitor times. This work integrates the proposed data assimilation method and fluid model to assimilate production and time-lapse seismic data in this situation. Compared to the application presented in the second study, this case increased the complexity by including some relevant features for a practical application, such as well data, scalar and categorical parameters, and a more physically complex problem. The results showed that it was possible to improve the reservoir characterization and production forecast by assimilating well and time-lapse seismic data using the proposed methods even in a complex nonlinear problem. Moreover, seismic monitoring can play an important role in improving reservoir characterization in Brazilian pre-salt projects.

Besides these four most relevant studies to this thesis's objective, it presents, as an appendix, the fifth work with a methodology to mitigate the influence of correlated model errors in the calibration of reservoir models using 3D seismic data. This study helps applications where the prior ensemble of reservoir models does not represent the 3D baseline seismic data due to geological modeling limitations. Furthermore, one could potentially adopt a similar methodology to time-lapse seismic data assimilation.

The main contributions and conclusions are:

- The weak-constraint formulation to the TLS data assimilation problem enabled the mitigation of the spatially correlated model error effects.
- It was possible to reduce the limitations regarding the data set size using the proposed iterative ensemble smoother method with local analysis for TLS data assimilation.
- The fluid modeling methodology contributed to reducing model errors by providing a physically consistent representation of volatile fluid production with significant CO_2 content.
- The two latter methods provided a viable solution to the data assimilation problem considering challenges similar to a Brazilian pre-salt field.

7 FUTURE STUDIES

Based on the observations of the five academic studies that comprise this thesis, it is possible to list some topics for future research:

- In chapter 2, a weak constraint formulation was the basis of a methodology to assimilate time-lapse seismic data considering spatially-correlated model error. Even though the error was additive, it is possible to build error models with different relations between the error parameter and the simulation output, inserting physical relations to the problem, which can be a topic for future developments.
- In the field application presented in chapter 2, the unmodeled resolution loss caused model errors that affected the data assimilation results. Although this effect was mitigated using a weak constraint formulation, it is also possible to improve the TLS forward model by including the seismic modeling process after the PEM. Future studies may address this approach to TLS data assimilation in complex realistic cases.
- In chapter 3, the SEnRML method assimilated TLS data in a local analysis scheme, leading to an efficient algorithm for big data sets. Nevertheless, it is possible to formulate the ESMDA method with a similar subspace inversion method, enabling an efficient formulation for problems with a vast number of measured data points. Comparing these two algorithms for production and time-lapse seismic data assimilation is a relevant topic for future research.
- The local analysis algorithm applied in chapter 3 grouped the parameters based on their physical positions in the model grid. Nevertheless, this process's idea is to consider the parameters related to the same subset of the data in a single analysis step. Therefore, future work can propose a correlation-based method to select local groups in the local analysis scheme.
- In chapter 4, the compositional fluid model for time-lapse seismic simulation provided speed of sound estimates in agreement with experimental data. However, measurements of the speed of sound with varying compositions were not available. The variation of light components in the porous media will occur when a miscible gas injection is part of the recovery strategy. Therefore, it is relevant to validate and possibly calibrate the current fluid model in this situation, which can be part of future research.
- In chapter 5, the SEnRML method with local analysis exhibited significant cost function oscillations when assimilating only well data. Moreover, the final data cost function

value was slightly worse than ESMDA considering this type of data. Therefore, more efforts to test the SEnRML method in highly nonlinear problems, like production data assimilation, are worthwhile. Developments in the sensitivity matrix estimation and the step-length control strategies could probably improve the data assimilation results.

- The application of the correlation-based local analysis method in the pre-salt-like case in chapter 5 led to results worse than the distance-based scheme in terms of data cost function value. One reason for this behavior was a biased prior ensemble that led to low correlation estimates different from the updated ensembles. Thus, it seems necessary to investigate how the prior ensemble definition can influence correlation-based localization and propose a methodology to mitigate such influences.
- The results presented in chapter 5 highlighted the influence of reservoir management operations in the production forecast evaluation. In some cases, a minor difference in the simulation model can cause a drastic change in the forecasts due to well valve changes and shut-in. Therefore, future research will evaluate different methods of validating and comparing models regarding their predictive capabilities.
- The UNISIM-III benchmark brings opportunity for other researches related to time-lapse seismic data assimilation. Firstly, it is possible to investigate the impact of multiple monitors in the reservoir calibration process as a total of eight seismic monitor acquisitions will be available. Secondly, one can study the influence of resolution loss, other seismic modeling processes, noise levels, and noise correlated with 3D seismic amplitude in the TLS data assimilation results. If there is a significant influence of spatially correlated errors, it is possible to apply the methodology presented in chapter 2 to improve the results.
- Another relevant topic for future studies is to assess the use of different seismic attributes in the TLS data assimilation workflow, such as elastic impedance, data with varying offsets, and time-shift.

8 REFERENCES

ABADPOUR, A.; BERGEY, P.; PIASECKI, R. 4D Seismic History Matching With Ensemble Kalman Filter-Assimilation on Hausdorff Distance to Saturation Front. *In*: SPE RESERVOIR SIMULATION SYMPOSIUM, 2013. **SPE Reservoir Simulation Symposium** [...]. The Woodlands, Texas, USA: Society of Petroleum Engineers, 2013. DOI 10.2118/163635-MS. Available at: http://www.onepetro.org/doi/10.2118/163635-MS. Accessed on: 4 Sep. 2018.

ABELHA, M.; PETERSOHN, E. The State of the Art of the Brazilian Pre-Salt Exploration. **AAPG ACE (2018). Salt Lake City Utah: AAPG**, 2018. .

AKTER, F.; IMTIAZ, S.; ZENDEHBOUDI, S.; HOSSAIN, K. Modified Ensemble Kalman filter for reservoir parameter and state estimation in the presence of model uncertainty. **Journal of Petroleum Science and Engineering**, vol. 199, p. 108323, Apr. 2021. https://doi.org/10.1016/j.petrol.2020.108323.

ALFONZO, M.; OLIVER, D. S. Seismic data assimilation with an imperfect model. **Computational Geosciences**, vol. 24, no. 2, p. 889–905, 1 Apr. 2020. https://doi.org/10.1007/s10596-019-09849-0.

ALMEIDA, F. L. R.; DAVOLIO, A.; SCHIOZER, D. J. Systematic Approach To Reduce Uncertainties When Quantitatively Assimilating 4D Seismic and Well Data. **SPE Reservoir Evaluation & Engineering**, vol. 23, no. 01, p. 013–030, 17 Feb. 2020. https://doi.org/10.2118/187081-PA.

ALTUNDAS, B.; CHUGUNOV, N.; RAMAKRISHNAN, T. S.; WILL, R. Quantifying the effect of CO2 dissolution on seismic monitoring of CO2 in CO2-EOR. *In*: SEG TECHNICAL PROGRAM EXPANDED ABSTRACTS 2017, 17 Aug. 2017. **SEG Technical Program Expanded Abstracts 2017** [...]. Houston, Texas: Society of Exploration Geophysicists, 17 Aug. 2017. p. 3771–3775. DOI 10.1190/segam2017-17792996.1. Available at: https://library.seg.org/doi/10.1190/segam2017-17792996.1. Accessed on: 1 Jul. 2021.

ALTUNDAS, Y. B.; CHUGUNOV, N.; RAMAKRISHNAN, T. S. On the importance of accurate CO_2 fluid and fluid substitution models for the seismic monitoring of CO_2. *In*: SEG TECHNICAL PROGRAM EXPANDED ABSTRACTS 2013, Sep. 2013. **SEG Technical Program Expanded Abstracts 2013** [...]. [*S. l.*]: Society of Exploration Geophysicists, Sep. 2013. p. 2716–2721. DOI 10.1190/segam2013-1451.1. Available at: http://library.seg.org/doi/abs/10.1190/segam2013-1451.1. Accessed on: 13 Sep. 2019.

AVANSI, G. D.; MASCHIO, C.; SCHIOZER, D. J. Simultaneous History-Matching Approach by Use of Reservoir-Characterization and Reservoir-Simulation Studies. **SPE Reservoir Evaluation & Engineering**, vol. 19, no. 04, p. 694–712, 1 Oct. 2016. https://doi.org/10.2118/179740-PA.

AVANSI, G. D.; SCHIOZER, D. J. UNISIM-I: synthetic model for reservoir development and management applications. **International Journal of Modeling and Simulation for the Petroleum Industry**, vol. 9, no. 1, 2015. .

AVSETH, P.; MUKERJI, T.; MAVKO, G. **Quantitative seismic interpretation: applying rock physics tools to reduce interpretation risk**. Cambridge, UK ; New York: Cambridge University Press, 2005.

BA, J.; XU, W.; FU, L.-Y.; CARCIONE, J. M.; ZHANG, L. Rock anelasticity due to patchy saturation and fabric heterogeneity: A double double-porosity model of wave propagation: Double Double-Porosity Wave Modeling. **Journal of Geophysical Research: Solid Earth**, 2017. DOI 10.1002/2016JB013882. Available at: http://doi.wiley.com/10.1002/2016JB013882. Accessed on: 14 May 2020.

BA, J.; ZHAO, J.; CARCIONE, J. M.; HUANG, X. Compressional wave dispersion due to rock matrix stiffening by clay squirt flow: Clay Squirt Flow in Tight Siltstone. **Geophysical Research Letters**, vol. 43, no. 12, p. 6186–6195, 28 Jun. 2016. https://doi.org/10.1002/2016GL069312.

BARREAU, A.; GAILLARD, K.; BÉHAR, E.; DARIDON, J. L.; LAGOURETTE, B.; XANS, P. Volumetric properties, isobaric heat capacity and sound velocity of condensate gas measurements and modelling. **Fluid Phase Equilibria**, vol. 127, no. 1–2, p. 155–171, Jan. 1997. https://doi.org/10.1016/S0378-3812(96)03140-8.

BATZLE, M.; WANG, Z. Seismic properties of pore fluids. **GEOPHYSICS**, vol. 57, no. 11, p. 1396–1408, Nov. 1992. https://doi.org/10.1190/1.1443207.

BERTOLINI, A. C.; MASCHIO, C.; SCHIOZER, D. J. A methodology to evaluate and reduce reservoir uncertainties using multivariate distribution. **Journal of Petroleum Science and Engineering**, vol. 128, p. 1–14, Apr. 2015. https://doi.org/10.1016/j.petrol.2015.02.003.

BISHOP, C. H.; HODYSS, D. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. **Quarterly Journal of the Royal Meteorological Society**, vol. 133, no. 629, p. 2029–2044, Oct. 2007. https://doi.org/10.1002/qj.169.

BOTECHIA, V. E.; ARAÚJO DE LEMOS, R.; HOHENDORFF FILHO, J. C. von; SCHIOZER, D. J. Well and ICV management in a carbonate reservoir with high gas content. **Journal of Petroleum Science and Engineering**, vol. 200, p. 108345, May 2021. https://doi.org/10.1016/j.petrol.2021.108345.

CARRASSI, A.; VANNITSEM, S. Deterministic Treatment of Model Error in Geophysical Data Assimilation. *In*: ANCONA, F.; CANNARSA, P.; JONES, C.; PORTALURI, A. (eds.). **Mathematical Paradigms of Climate Science**. Cham: Springer International Publishing, 2016. vol. 15, p. 175–213. DOI 10.1007/978-3-319-39092-5_9. Available at: http://link.springer.com/10.1007/978-3-319-39092-5_9. Accessed on: 7 Mar. 2019.

CAVALCANTE, C. C. B.; MASCHIO, C.; SANTOS, A. A.; SCHIOZER, D.; ROCHA, A. History matching through dynamic decision-making. **PLOS ONE**, vol. 12, no. 6, p. e0178507, 5 Jun. 2017. https://doi.org/10.1371/journal.pone.0178507.

CAVALCANTE, C. C. B.; MASCHIO, C.; SCHIOZER, D.; ROCHA, A. A stochastic learning-from-data approach to the history-matching problem. **Engineering Applications of Artificial Intelligence**, vol. 94, p. 103767, Sep. 2020. https://doi.org/10.1016/j.engappai.2020.103767.

CHASSAGNE, R.; OBIDEGWU, D.; DAMBRINE, J.; MACBETH, C. Binary 4D seismic history matching, a metric study. **Computers & Geosciences**, vol. 96, p. 159–172, Nov. 2016. https://doi.org/10.1016/j.cageo.2016.08.013.

CHEN, Yan; OLIVER, D. S. Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother. **Mathematical Geosciences**, vol. 44, no. 1, p. 1–26, Jan. 2012. https://doi.org/10.1007/s11004-011-9376-z.

CHEN, Yan; OLIVER, D. S. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. **Computational Geosciences**, vol. 17, no. 4, p. 689–703, Aug. 2013. https://doi.org/10.1007/s10596-013-9351-5.

CHEN, Yan; OLIVER, D. S. Localization and regularization for iterative ensemble smoothers. **Computational Geosciences**, vol. 21, no. 1, p. 13–30, Feb. 2017. https://doi.org/10.1007/s10596-016-9599-7.

CHEN, Yilun; WIESEL, A.; ELDAR, Y. C.; HERO, A. O. Shrinkage Algorithms for MMSE Covariance Estimation. **IEEE Transactions on Signal Processing**, vol. 58, no. 10, p. 5016–5029, Oct. 2010. https://doi.org/10.1109/TSP.2010.2053029.

CHENG, W.; BA, J.; FU, L.-Y.; LEBEDEV, M. Wave-velocity dispersion and rock microstructure. **Journal of Petroleum Science and Engineering**, vol. 183, p. 106466, Dec. 2019. https://doi.org/10.1016/j.petrol.2019.106466.

CHRISTIE, M. A. Upscaling for Reservoir Simulation. **Journal of Petroleum Technology**, vol. 48, no. 11, p. 1004–1010, 1 Nov. 1996. https://doi.org/10.2118/37324-JPT.

CHRISTIE, M. A.; BLUNT, M. J. Tenth SPE Comparative Solution Project: A Comparison of Upscaling Techniques. **SPE Reservoir Evaluation & Engineering**, vol. 4, no. 04, p. 308–317, 1 Aug. 2001. https://doi.org/10.2118/72469-PA.

CMG, C. M. G. LTD. **GEM Compositional & Unconventional Simulator**. https://www.cmgl.ca/gem: Computer Modelling Group LTD., 2017(version 2017.10). Available at: https://www.cmgl.ca/gem. Accessed on: 30 Mar. 2020.

CMG, C. M. G. LTD. **WinProp Fluid Property Characterization Tool**. https://www.cmgl.ca/winprop: Computer Modelling Group LTD., 2015(version 2015.10). Available at: https://www.cmgl.ca/winprop. Accessed on: 30 Mar. 2020.

CMG, COMPUTER MODELLING GROUP LTD. **IMEX Black Oil & Unconventional Simulator**. [*S. l.*: *s. n.*], 2017. Available at: https://www.cmgl.ca/imex.

CONNOLLY, P. Calibration and inversion of non-zero Offset seismic. *In*: SEG TECHNICAL PROGRAM EXPANDED ABSTRACTS 1998, Jan. 1998. **SEG Technical Program Expanded Abstracts 1998** [...]. [*S. l.*]: Society of Exploration Geophysicists, Jan. 1998. p. 182–184. DOI 10.1190/1.1820288. Available at: http://library.seg.org/doi/abs/10.1190/1.1820288. Accessed on: 7 Apr. 2020.

CORREIA, M.; BOTECHIA, V.; PIRES, L.; RIOS, V.; SANTOS, S.; RIOS, V.; HOHENDORFF, J.; CHAVES, M.; SCHIOZER, D. UNISIM-III: Benchmark Case Proposal Based on a Fractured Karst Reservoir. *In*: ECMOR XVII, 2020. **ECMOR XVII** [...]. Online Event,: European Association of Geoscientists & Engineers, 2020. p. 1–14. DOI 10.3997/2214-

4609.202035018. Available at: https://www.earthdoc.org/content/papers/10.3997/2214-4609.202035018. Accessed on: 3 Mar. 2021.

COSTA, M. M.; SILVA, E. P.; SANTOS, M. S.; VASQUEZ, G. F. Quantification of 4D seismic response in pre-salt carbonates via time shift analysis. 2016. **3rd EAGE/SBGf Workshop 2016: Quantitative Seismic Interpretation of Lacustrine Carbonates** [...]. [*S. l.*]: European Association of Geoscientists and Engineers, EAGE, 2016. p. 85–89. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84973468804&partnerID=40&md5=5bf7be4f51a134cf8126b2da6ac04dcc %[6 April 2016 through 7 April 2016.

DA NÓBREGA, D. V.; DE MORAES, F. S.; EMERICK, A. A. Data assimilation of a legacy 4D seismic in a brown field. **Journal of Geophysics and Engineering**, vol. 15, no. 6, p. 2585–2601, 1 Dec. 2018. https://doi.org/10.1088/1742-2140/aadd68.

DANAEI, S.; SILVA NETO, G. M.; SCHIOZER, D. J.; DAVOLIO, A. Using petro-elastic proxy model to integrate 4D seismic in ensemble based data assimilation. **Journal of Petroleum Science and Engineering**, vol. 194, p. 107457, Nov. 2020. https://doi.org/10.1016/j.petrol.2020.107457.

DARIDON, J. L.; LAGOURETTE, B.; XANS, P.; MONTEL, F. Petroleum characterization from ultrasonic measurement. **Journal of Petroleum Science and Engineering**, vol. 19, no. 3–4, p. 281–293, Mar. 1998. https://doi.org/10.1016/S0920-4105(97)00050-8.

DASHTI, H. H.; RIAZI, M. R. Acoustic velocities in petroleum fluids: Measurement and prediction. **Journal of Petroleum Science and Engineering**, vol. 124, p. 94–104, Dec. 2014. https://doi.org/10.1016/j.petrol.2014.10.013.

DAVOLIO, A.; SCHIOZER, D. J. A Proper Data Comparison for Seismic History Matching Processes. *In*: SPE EUROPEC FEATURED AT 81ST EAGE CONFERENCE AND EXHIBITION, 3 Jun. 2019. **SPE Europec featured at 81st EAGE Conference and Exhibition** [...]. London, England, UK: SPE, 3 Jun. 2019. p. D031S005R004. DOI 10.2118/195549-MS. Available at: https://onepetro.org/SPEEURO/proceedings/19EURO/3-19EURO/London,%20England,%20UK/217899. Accessed on: 1 Jul. 2021.

DAVOLIO, A.; SCHIOZER, D. J. Probabilistic seismic history matching using binary images. **Journal of Geophysics and Engineering**, vol. 15, no. 1, p. 261–274, 1 Feb. 2018. https://doi.org/10.1088/1742-2140/aa99f4.

DE MORAES CRUZ, R. O.; ROSA, M. B.; BRANCO, C. C. M.; DE SANT'ANNA PIZARRO, J. O.; DE SOUZA SILVA, C. T. Lula NE Pilot Project - An Ultra-Deep Success in the Brazilian Pre-Salt. *In*: OFFSHORE TECHNOLOGY CONFERENCE, 2 May 2016. **Offshore Technology Conference** [...]. Houston, Texas, USA: OTC, 2 May 2016. p. D011S007R001. DOI 10.4043/27297-MS. Available at: https://onepetro.org/OTCONF/proceedings/16OTC/1-16OTC/Houston,%20Texas,%20USA/84739. Accessed on: 10 Mar. 2021.

DE SANT´ANNA PIZARRO, J. O.; BRANCO, C. C. Challenges in Implementing an EOR Project in the Pre-Salt Province in Deep Offshore Brasil. *In*: SPE EOR CONFERENCE AT OIL AND GAS WEST ASIA, 16 Apr. 2012. **SPE EOR Conference at Oil and Gas West Asia** [...]. Muscat, Oman: SPE, 16 Apr. 2012. p. SPE-155665-MS. DOI 10.2118/155665-MS.

Available at: https://onepetro.org/SPEOGWA/proceedings/12OGWA/All-12OGWA/Muscat,%20Oman/158339. Accessed on: 10 Mar. 2021.

DE SOUZA, R. M. **Quantitative Analysis of 4D Seismic and Production Data for Saturation Estimation and Fluid-flow Model Assessment**. 2018. PhD Thesis – The University of Western Australia, Perth, Australia, 2018.

DEPLANTE, C.; COSTA, M.; DOS SANTOS, M. S.; DE MELO DIAS, R.; MELLO, V. L.; MEIRELLES, B. R.; DA SILVA, S.; ARAUJO, D.; SANSONOWSKI, R. Using full wave seismic modeling to test 4D repeatability for Libra pre-salt field. *In*: 16TH INTERNATIONAL CONGRESS OF THE BRAZILIAN GEOPHYSICAL SOCIETY, 2019. **16th International Congress of the Brazilian Geophysical Society** [...]. Rio de Janeiro: Sociedade Brasileira de Geofísica, 2019. Available at: https://sbgf.org.br/mysbgf/eventos/expanded_abstracts/16th_CISBGf/Using%20full%20wave%20seismic%20modeling%20to%20test%204D%20repeatability%20for%20Libra%20pre-salt%20field.pdf. Accessed on: 10 Mar. 2021.

DOHERTY, J.; WELTER, D. A short exploration of structural noise: A SHORT EXPLORATION OF STRUCTURAL NOISE. **Water Resources Research**, vol. 46, no. 5, May 2010. DOI 10.1029/2009WR008377. Available at: http://doi.wiley.com/10.1029/2009WR008377. Accessed on: 5 May 2020.

DONG, Y.; GU, Y.; OLIVER, D. S. Sequential assimilation of 4D seismic data for reservoir description using the ensemble Kalman filter. **Journal of Petroleum Science and Engineering**, vol. 53, no. 1–2, p. 83–99, Aug. 2006. https://doi.org/10.1016/j.petrol.2006.03.028.

DOS SANTOS, J. M. C.; DAVOLIO, A.; SCHIOZER, D. J.; MACBETH, C. Semiquantitative 4D seismic interpretation integrated with reservoir simulation: Application to the Norne field. **Interpretation**, vol. 6, no. 3, p. T601–T611, 1 Aug. 2018. https://doi.org/10.1190/INT-2017-0122.1.

DOYEN, P. M. **Seismic reservoir characterization: an earth modelling perspective**. Houten, The Netherlands: EAGE publications, 2007.

EMERICK, A. **History Matching and Uncertainty Characterization Using Ensemble-based Methods**. 2012. 339 pp. f. Doctoral Thesis – The University of Tulsa, Tulsa, OK, 2012. https://doi.org/10.13140/2.1.2117.4723.

EMERICK, A.; REYNOLDS, A. Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. **Computational Geosciences**, vol. 15, no. 2, p. 251–269, 2011. https://doi.org/10.1007/s10596-010-9198-y.

EMERICK, Alexandre A. Analysis of geometric selection of the data-error covariance inflation for ES-MDA. **Journal of Petroleum Science and Engineering**, vol. 182, p. 106168, Nov. 2019. https://doi.org/10.1016/j.petrol.2019.06.032.

EMERICK, Alexandre A. Analysis of the performance of ensemble-based assimilation of production and seismic data. **Journal of Petroleum Science and Engineering**, vol. 139, p. 219–239, 2016. .

EMERICK, Alexandre A. Deterministic ensemble smoother with multiple data assimilation as an alternative for history-matching seismic data. **Computational Geosciences**, vol. 22, no. 5, p. 1175–1186, Oct. 2018. https://doi.org/10.1007/s10596-018-9745-5.

EMERICK, Alexandre A.; REYNOLDS, A. C. Ensemble smoother with multiple data assimilation. **Computers & Geosciences**, vol. 55, p. 3–15, Jun. 2013a. https://doi.org/10.1016/j.cageo.2012.03.011.

EMERICK, Alexandre A.; REYNOLDS, A. C. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. **Computational Geosciences**, vol. 16, no. 3, p. 639–659, Jun. 2012. https://doi.org/10.1007/s10596-012-9275-5.

EMERICK, Alexandre A.; REYNOLDS, A. C. History-matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation. 2013b. **SPE Reservoir Simulation Symposium** [...]. [*S. l.*]: Society of Petroleum Engineers, 2013. https://doi.org/10.2118/163675-MS.

EMERICK, Alexandre Anozé; MORAES, R.; RODRIGUES, J.; OTHERS. Calculating seismic attributes within a reservoir flow simulator. 2007. **Latin American & Caribbean Petroleum Engineering Conference** [...]. [*S. l.*]: Society of Petroleum Engineers, 2007. Available at: https://www.onepetro.org/conference-paper/SPE-107001-MS. Accessed on: 10 Oct. 2017.

EVENSEN, G. Consistent Formulation and Error Statistics for Reservoir History Matching. *In*: ECMOR XVII, 2020., 14 Sep. 2020. **ECMOR XVII** [...]. [*S. l.*]: European Association of Geoscientists & Engineers, 14 Sep. 2020. vol. 2020, p. 1–22. DOI 10.3997/2214-4609.202035022. Available at: https://www.earthdoc.org/content/papers/10.3997/2214-4609.202035022. Accessed on: 23 Oct. 2020.

EVENSEN, G. Introducing Stochastic Model Errors In Ensemble-Based History Matching. *In*: ECMOR XVI - 16TH EUROPEAN CONFERENCE ON THE MATHEMATICS OF OIL RECOVERY, 3 Sep. 2018. **ECMOR XVI - 16th European Conference on the Mathematics of Oil Recovery** [...]. Barcelona, Spain: [*s. n.*], 3 Sep. 2018. DOI 10.3997/2214-4609.201802280. Available at: http://www.earthdoc.org/publication/publicationdetails/?publication=94015. Accessed on: 19 Feb. 2019.

EVENSEN, Geir. Accounting for model errors in iterative ensemble smoothers. **Computational Geosciences**, vol. 23, no. 4, p. 761–775, Aug. 2019. https://doi.org/10.1007/s10596-019-9819-z.

EVENSEN, Geir. Analysis of iterative ensemble smoothers for solving inverse problems. **Computational Geosciences**, vol. 22, no. 3, p. 885–908, Jun. 2018. https://doi.org/10.1007/s10596-018-9731-y.

EVENSEN, Geir. **Data Assimilation**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. DOI 10.1007/978-3-642-03711-5. Available at: http://link.springer.com/10.1007/978-3-642-03711-5. Accessed on: 17 Aug. 2018.

EVENSEN, Geir. Sampling strategies and square root analysis schemes for the EnKF. **Ocean Dynamics**, vol. 54, no. 6, p. 539–560, Dec. 2004. https://doi.org/10.1007/s10236-004-0099-2.

EVENSEN, Geir. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. **Journal of Geophysical Research**, vol. 99, no. C5, p. 10143, 1994. https://doi.org/10.1029/94JC00572.

EVENSEN, Geir. The Ensemble Kalman Filter: theoretical formulation and practical implementation. **Ocean Dynamics**, vol. 53, no. 4, p. 343–367, 1 Nov. 2003. https://doi.org/10.1007/s10236-003-0036-9.

EVENSEN, Geir; EIKREM, K. S. Conditioning reservoir models on rate data using ensemble smoothers. **Computational Geosciences**, vol. 22, no. 5, p. 1251–1270, Oct. 2018. https://doi.org/10.1007/s10596-018-9750-8.

EVENSEN, Geir; RAANES, P. N.; STORDAL, A. S.; HOVE, J. Efficient Implementation of an Iterative Ensemble Smoother for Data Assimilation and Reservoir History Matching. **Frontiers in Applied Mathematics and Statistics**, vol. 5, p. 47, 4 Oct. 2019. https://doi.org/10.3389/fams.2019.00047.

EVENSEN, Geir; VAN LEEUWEN, P. J. Assimilation of Geosat Altimeter Data for the Agulhas Current Using the Ensemble Kalman Filter with a Quasigeostrophic Model. **Monthly Weather Review**, vol. 124, no. 1, p. 85–96, Jan. 1996. https://doi.org/10.1175/1520-0493(1996)124<0085:AOGADF>2.0.CO;2.

FAHIMUDDIN, A.; AANONSEN, S. I.; SKJERVHEIM, J.-A. 4D Seismic History Matching of a Real Field Case With EnKF: Use of Local Analysis for Model Updating. *In*: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, 2010a. **SPE Annual Technical Conference and Exhibition** [...]. Florence, Italy: Society of Petroleum Engineers, 2010. DOI 10.2118/134894-MS. Available at: http://www.onepetro.org/doi/10.2118/134894-MS. Accessed on: 18 Jun. 2020.

FAHIMUDDIN, A.; AANONSEN, S. I.; SKJERVHEIM, J.-A. Ensemble based 4D seismic history matching: Integration of different levels and types of seismic data. 2010b. **SPE EUROPEC/EAGE Annual Conference and Exhibition** [...]. [*S. l.*]: Society of Petroleum Engineers, 2010. https://doi.org/10.2118/131453-MS.

FIROOZABADI, A. **Thermodynamics of hydrocarbon reservoirs**. New York: McGraw-Hill, 1999.

FORMENTIN, H. N.; ALMEIDA, F. la R.; AVANSI, G. D.; MASCHIO, C.; SCHIOZER, D. J.; CAIADO, C.; VERNON, I.; GOLDSTEIN, M. Gaining more understanding about reservoir behavior through assimilation of breakthrough time and productivity deviation in the history matching process. **Journal of Petroleum Science and Engineering**, vol. 173, p. 1080–1096, Feb. 2019. https://doi.org/10.1016/j.petrol.2018.10.045.

FOSSUM, K.; LORENTZEN, R. J. Assisted History Matching of 4D Seismic Data - A Comparative Study. *In*: PETROLEUM GEOSTATISTICS 2019, 2019. **Petroleum Geostatistics 2019** [...]. Florence, Italy,: European Association of Geoscientists & Engineers, 2019. p. 1–5. DOI 10.3997/2214-4609.201902180. Available at: https://www.earthdoc.org/content/papers/10.3997/2214-4609.201902180. Accessed on: 8 Apr. 2021.

GASPARI, G.; COHN, S. E. Construction of correlation functions in two and three dimensions. **Quarterly Journal of the Royal Meteorological Society**, vol. 125, no. 554, p. 723–757, Jan. 1999. https://doi.org/10.1002/qj.49712555417.

GASSMANN, F. Über die elastizität poröser medien: Vierteljahrss-chrift der Naturforschenden Gesellschaft in Zurich 96, 1-23. **Paper translation at http://sepwww. stanford. edu/sep/berryman/PS/gassmann. pdf**, 1951. .

GOSSELIN, O.; AANONSEN, S. I.; AAVATSMARK, I.; COMINELLI, A.; GONARD, R.; KOLASINSKI, M.; FERDINANDI, F.; KOVACIC, L.; NEYLON, K. History matching using time-lapse seismic (HUTS). 2003. **SPE Annual Technical Conference and Exhibition** [...]. [*S. l.*]: Society of Petroleum Engineers, 2003.

GU, Y.; OLIVER, D. S. An Iterative Ensemble Kalman Filter for Multiphase Fluid Flow Data Assimilation. **SPE Journal**, vol. 12, no. 04, p. 438–446, 1 Nov. 2007. https://doi.org/10.2118/108438-PA.

HAN, D.; SUN, M.; LIU, J. Velocity and density of CO2-oil miscible mixtures. *In*: 2012 SEG ANNUAL MEETING, Sep. 2012. **2012 SEG Annual Meeting** [...]. Las Vegas: Society of Exploration Geophysicists, Sep. 2012. p. 1–5. DOI 10.1190/segam2012-1242.1. Available at: http://library.seg.org/doi/abs/10.1190/segam2012-1242.1. Accessed on: 24 May 2018.

HAN, D.; SUN, M.; LIU, J. Velocity and density of oil-HC-CO2 miscible mixtures. *In*: 2013 SEG ANNUAL MEETING, Sep. 2013. **2013 SEG Annual Meeting** [...]. Houston: Society of Exploration Geophysicists, Sep. 2013. p. 2831–2835. DOI 10.1190/segam2013-1104.1. Available at: http://library.seg.org/doi/abs/10.1190/segam2013-1104.1. Accessed on: 24 May 2018.

HARLIM, J. Model Error in Data Assimilation. **arXiv:1311.3579 [physics]**, 14 Nov. 2013. Available at: http://arxiv.org/abs/1311.3579. Accessed on: 7 Mar. 2019.

HASHIN, Z.; SHTRIKMAN, S. A variational approach to the elastic behavior of multiphase materials: Journal Mechanics Physical Solids, 11, 127–140. **CrossRef Google Scholar**, 1963. .

HUNT, B. R.; KOSTELICH, E. J.; SZUNYOGH, I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. **Physica D: Nonlinear Phenomena**, vol. 230, no. 1–2, p. 112–126, Jun. 2007. https://doi.org/10.1016/j.physd.2006.11.008.

JOHANN, P. R. S.; MONTEIRO, R. C. Geophysical Reservoir Characterization and Monitoring at Brazilian Pre-Salt Oil Fields. *In*: OFFSHORE TECHNOLOGY CONFERENCE, 2 May 2016. **Offshore Technology Conference** [...]. Houston, Texas, USA: OTC, 2 May 2016. p. D021S020R003. DOI 10.4043/27246-MS. Available at: https://onepetro.org/OTCONF/proceedings/16OTC/2-16OTC/Houston,%20Texas,%20USA/84680. Accessed on: 10 Mar. 2021.

JUNG, S.; LEE, K.; PARK, C.; CHOE, J. Ensemble-Based Data Assimilation in Reservoir Characterization: A Review. **Energies**, vol. 11, no. 2, p. 445, 17 Feb. 2018. https://doi.org/10.3390/en11020445.

KETINENI, S. P.; KALLA, S.; OPPERT, S.; BILLITER, T. Quantitative Integration of 4D Seismic with Reservoir Simulation. **SPE Journal**, vol. 25, no. 04, p. 2055–2066, 1 Aug. 2020. https://doi.org/10.2118/191521-PA.

KUMAR, A.; FARMER, C. L.; JERAULD, G. R.; LI, D. Efficient Upscaling from Cores to Simulation Models. *In*: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, 1997. **SPE Annual Technical Conference and Exhibition** [...]. San Antonio, Texas: Society of Petroleum Engineers, 1997. DOI 10.2118/38744-MS. Available at: http://www.onepetro.org/doi/10.2118/38744-MS. Accessed on: 20 Sep. 2019.

LACERDA, J. M.; EMERICK, A. A.; PIRES, A. P. Methods to mitigate loss of variance due to sampling errors in ensemble data assimilation with non-local model parameters. **Journal of Petroleum Science and Engineering**, vol. 172, p. 690–706, Jan. 2019. https://doi.org/10.1016/j.petrol.2018.08.056.

LANCASTER, S.; WHITCOMBE, D. Fast-track 'coloured' inversion. Jan. 2000. **SEG Technical Program Expanded Abstracts 2000** [...]. [*S. l.*]: Society of Exploration Geophysicists, Jan. 2000. p. 1572–1575. DOI 10.1190/1.1815711. Available at: http://library.seg.org/doi/abs/10.1190/1.1815711. Accessed on: 3 Feb. 2018.

LE RAVALEC, M.; TILLIER, E.; DA VEIGA, S.; ENCHERY, G.; GERVAIS, V. Advanced Integrated Workflows for Incorporating Both Production and 4D Seismic-Related Data into Reservoir Models. **Oil & Gas Science and Technology – Revue d'IFP Energies nouvelles**, vol. 67, no. 2, p. 207–220, Mar. 2012. https://doi.org/10.2516/ogst/2011159.

LEDOIT, O.; WOLF, M. A well-conditioned estimator for large-dimensional covariance matrices. **Journal of Multivariate Analysis**, vol. 88, no. 2, p. 365–411, Feb. 2004. https://doi.org/10.1016/S0047-259X(03)00096-4.

LEEUWENBURGH, O.; ARTS, R. Distance parameterization for efficient seismic history matching with the ensemble Kalman Filter. **Computational Geosciences**, vol. 18, no. 3–4, p. 535–548, Aug. 2014. https://doi.org/10.1007/s10596-014-9434-y.

LEEUWENBURGH, O.; BROUWER, J.; TRANI, M. Ensemble-based conditioning of reservoir models to seismic data. **Computational Geosciences**, vol. 15, no. 2, p. 359–378, Mar. 2011. https://doi.org/10.1007/s10596-010-9209-z.

LIE, K.-A. **An introduction to reservoir simulation using MATLAB/GNU Octave: user guide for the MATLAB Reservoir Simulation Toolbox (MRST)**. Cambridge, United Kingdom ; New York, NY, USA: Cambridge University Press, 2019.

LIMA, M. M.; EMERICK, A. A.; ORTIZ, C. E. P. Data-space inversion with ensemble smoother. **Computational Geosciences**, vol. 24, no. 3, p. 1179–1200, Jun. 2020. https://doi.org/10.1007/s10596-020-09933-w.

LIU, M.; GRANA, D. Time-lapse seismic history matching with an iterative ensemble smoother and deep convolutional autoencoder. **GEOPHYSICS**, vol. 85, no. 1, p. M15–M31, 1 Jan. 2020. https://doi.org/10.1190/geo2019-0019.1.

LORENTZEN, R. J.; LUO, X.; BHAKTA, T.; VALESTRAND, R. History Matching the Full Norne Field Model Using Seismic and Production Data. **SPE Journal**, vol. 24, no. 04, p. 1452–1467, 1 Aug. 2019. https://doi.org/10.2118/194205-PA.

LU, M.; CHEN, Y. Improved Estimation and Forecasting Through Residual-Based Model Error Quantification. **SPE Journal**, 1 Jan. 2020. DOI 10.2118/199358-PA. Available at: http://www.onepetro.org/doi/10.2118/199358-PA. Accessed on: 15 Jan. 2020.

LUO, X. Ensemble-based kernel learning for a class of data assimilation problems with imperfect forward simulators. **PLOS ONE**, vol. 14, no. 7, p. e0219247, 11 Jul. 2019. https://doi.org/10.1371/journal.pone.0219247.

LUO, X.; BHAKTA, T. Automatic and adaptive localization for ensemble-based history matching. **Journal of Petroleum Science and Engineering**, vol. 184, p. 106559, Jan. 2020. https://doi.org/10.1016/j.petrol.2019.106559.

LUO, X.; BHAKTA, T.; JAKOBSEN, M.; NAEVDAL, G. An Ensemble 4D-Seismic History-Matching Framework With Sparse Representation Based On Wavelet Multiresolution Analysis. **SPE Journal**, vol. 22, no. 03, p. 0985–1010, 1 Jun. 2017. https://doi.org/10.2118/180025-PA.

LUO, X.; BHAKTA, T.; JAKOBSEN, M.; NÆVDAL, G. Efficient big data assimilation through sparse representation: A 3D benchmark case study in petroleum engineering. **PLOS ONE**, vol. 13, no. 7, p. e0198586, 27 Jul. 2018. https://doi.org/10.1371/journal.pone.0198586.

LUO, X.; BHAKTA, T.; NÆVDAL, G. Correlation-Based Adaptive Localization With Applications to Ensemble-Based 4D-Seismic History Matching. **SPE Journal**, vol. 23, no. 02, p. 396–427, 1 Apr. 2018. https://doi.org/10.2118/185936-PA.

LUO, X.; LORENTZEN, R. J.; BHAKTA, T. Accounting for model errors of rock physics models in 4D seismic history matching problems: A perspective of machine learning. **Journal of Petroleum Science and Engineering**, vol. 196, p. 107961, Jan. 2021. https://doi.org/10.1016/j.petrol.2020.107961.

LUO, X.; LORENTZEN, R. J.; VALESTRAND, R.; EVENSEN, G. Correlation-Based Adaptive Localization for Ensemble-Based History Matching: Applied To the Norne Field Case Study. **SPE Reservoir Evaluation & Engineering**, vol. 22, no. 03, p. 1084–1109, 1 Aug. 2019. https://doi.org/10.2118/191305-PA.

LUO, X.; STORDAL, A. S.; LORENTZEN, R. J.; NAEVDAL, G. Iterative Ensemble Smoother as an Approximate Solution to a Regularized Minimum-Average-Cost Problem: Theory and Applications. **SPE Journal**, vol. 20, no. 05, p. 0962–0982, 1 Oct. 2015. https://doi.org/10.2118/176023-PA.

LYGREN, M.; FAGERVIK, K.; VALEN, T. S.; HETLELID, A.; BERGE, G.; DAHL, G. V.; SØNNELAND, L.; LIE, H. E.; MAGNUS, I. A method for performing history matching of reservoir flow models using 4D seismic data. **Petroleum Geoscience**, vol. 9, no. 1, p. 83–90, 2003. .

MA, W.; JAFARPOUR, B.; QIN, J. Dynamic characterization of geologic CO2 storage aquifers from monitoring data with ensemble Kalman filter. **International Journal of Greenhouse Gas Control**, vol. 81, p. 199–215, Feb. 2019. https://doi.org/10.1016/j.ijggc.2018.10.009.

MACBETH, C.; FLORICICH, M.; SOLDO, J. Going quantitative with 4D seismic analysis. **Geophysical Prospecting**, vol. 54, no. 3, p. 303–317, May 2006. https://doi.org/10.1111/j.1365-2478.2006.00536.x.

MALEKI, M.; DAVOLIO, A.; SCHIOZER, D. J. Quantitative integration of 3D and 4D seismic impedance into reservoir simulation model updating in the Norne Field: Reservoir simulation model updating in the Norne Field. **Geophysical Prospecting**, vol. 67, no. 1, p. 167–187, Jan. 2019. https://doi.org/10.1111/1365-2478.12717.

MASCHIO, C.; AVANSI, G. D.; SANTOS, A. A.; SCHIOZER, D. J. UNISIM-IH: case study for history matching. 2013. **UNISIM website**. Available at: https://www.unisim.cepetro.unicamp.br/benchmarks/br/unisim-i/unisim-i-h. Accessed on: 4 Sep. 2019.

MASCHIO, Célio; SCHIOZER, D. J. A new methodology for history matching combining iterative discrete Latin Hypercube with multi-start simulated annealing. **Journal of Petroleum Science and Engineering**, vol. 169, p. 560–577, Oct. 2018. https://doi.org/10.1016/j.petrol.2018.06.004.

MASCHIO, Célio; SCHIOZER, D. J. Integration of geostatistical realizations in data assimilation and reduction of uncertainty process using genetic algorithm combined with multi-start simulated annealing. **Oil & Gas Science and Technology – Revue d'IFP Energies nouvelles**, vol. 74, p. 73, 2019. https://doi.org/10.2516/ogst/2019045.

MASCHIO, Célio; SCHIOZER, D. J. Probabilistic history matching using discrete Latin Hypercube sampling and nonparametric density estimation. **Journal of Petroleum Science and Engineering**, vol. 147, p. 98–115, Nov. 2016. https://doi.org/10.1016/j.petrol.2016.05.011.

MAVKO, G.; MUKERJI, T.; DVORKIN, J. **The rock physics handbook: Tools for seismic analysis of porous media**. 2nd ed. Cambridge, UK ; New York: Cambridge University Press, 2009.

MOCZYDLOWER, B..; SALOMÃO, M. C.; BRANCO, C. C.; ROMEU, R. K.; HOMEM, T. R.; FREITAS, L. C.; LIMA, H. A. Development of the Brazilian Pre-Salt Fields – When to Pay for Information and When to Pay for Flexibility. *In*: SPE LATIN AMERICA AND CARIBBEAN PETROLEUM ENGINEERING CONFERENCE, 16 Apr. 2012. **SPE Latin America and Caribbean Petroleum Engineering Conference** [...]. Mexico City, Mexico: SPE, 16 Apr. 2012. p. SPE-152860-MS. DOI 10.2118/152860-MS. Available at: https://onepetro.org/SPELACP/proceedings/12LACP/All-12LACP/Mexico%20City,%20Mexico/157656. Accessed on: 10 Mar. 2021.

MOCZYDLOWER, Bruno; FIGUEIREDO JUNIOR, F. P.; PIZARRO, J. O. S. Libra Extended Well Test - An Innovative Approach to De-Risk a Complex Field Development. *In*: OFFSHORE TECHNOLOGY CONFERENCE, 26 Apr. 2019. **Offshore Technology Conference** [...]. Houston, Texas: OTC, 26 Apr. 2019. p. D011S009R003. DOI 10.4043/29653-MS. Available at: https://onepetro.org/OTCONF/proceedings/19OTC/1-19OTC/Houston,%20Texas/181538. Accessed on: 22 Mar. 2021.

MUKERJI, T.; JØRSTAD, A.; MAVKO, G.; GRANLI, J. R. Near and far offset impedances: Seismic attributes for identifying lithofacies and pore fluids. **Geophysical Research Letters**, vol. 25, no. 24, p. 4557–4560, 15 Dec. 1998. https://doi.org/10.1029/1998GL900187.

NAKANO, C. M. F.; CAPELEIRO PINTO, A. C.; MARCUSSO, J. L.; MINAMI, K. Pre-Salt Santos Basin -Extended Well Test and Production Pilot in the Tupi Area - The Planning Phase.

In: OFFSHORE TECHNOLOGY CONFERENCE, 4 May 2009. **Offshore Technology Conference** [...]. Houston, Texas: OTC, 4 May 2009. p. OTC-19886-MS. DOI 10.4043/19886-MS. Available at: https://onepetro.org/OTCONF/proceedings/09OTC/All-09OTC/Houston,%20Texas/35850. Accessed on: 22 Mar. 2021.

OBIDEGWU, D.; CHASSAGNE, R.; MACBETH, C. Seismic assisted history matching using binary maps. **Journal of Natural Gas Science and Engineering**, vol. 42, p. 69–84, Jun. 2017. https://doi.org/10.1016/j.jngse.2017.03.001.

OLIVEIRA, G. S.; SCHIOZER, D. J.; MASCHIO, C. History matching by integrating regional multi-property image perturbation methods with a multivariate sensitivity analysis. **Journal of Petroleum Science and Engineering**, vol. 153, p. 111–122, May 2017. https://doi.org/10.1016/j.petrol.2017.03.031.

OLIVER, D. S.; REYNOLDS, A. C.; LIU, N. **Inverse Theory for Petroleum Reservoir Characterization and History Matching**. [*S. l.*]: Cambridge University Press, 2008. Available at: https://books.google.com.br/books?id=oPuy3OfbtfIC.

OLIVER, Dean S.; CHEN, Y. Recent progress on reservoir history matching: a review. **Computational Geosciences**, vol. 15, no. 1, p. 185–221, Jan. 2011. https://doi.org/10.1007/s10596-010-9194-2.

OLIVER, D.S.; ALFONZO, M. A. Calibration of imperfect models to biased observations. **Computational Geosciences**, vol. 22, no. 1, p. 145–161, Feb. 2018a. https://doi.org/10.1007/s10596-017-9678-4.

OLIVER, D.S.; ALFONZO, M. A. Seismic Data Assimilation With An Imperfect Model. *In*: ECMOR XVI - 16TH EUROPEAN CONFERENCE ON THE MATHEMATICS OF OIL RECOVERY, 3 Sep. 2018b. **ECMOR XVI - 16th European Conference on the Mathematics of Oil Recovery** [...]. Barcelona, Spain: [*s. n.*], 3 Sep. 2018. DOI 10.3997/2214-4609.201802283. Available at: http://www.earthdoc.org/publication/publicationdetails/?publication=94018. Accessed on: 25 Sep. 2018.

O'SULLIVAN, A. E. **Modelling simulation error for improved reservoir prediction**. 2004. PhD Thesis – Heriot-Watt University, Edinburgh, Scotland, UK, 2004.

PEDERSEN, K. S.; CHRISTENSEN, P. L.; SHAIKH, J. A. **Phase behavior of petroleum reservoir fluids**. Second edition. Boca Raton: CRC Press, Taylor & Francis Group, 2015.

PÉNELOUX, A.; RAUZY, E.; FRÉZE, R. A consistent correction for Redlich-Kwong-Soave volumes. **Fluid Phase Equilibria**, vol. 8, no. 1, p. 7–23, Jan. 1982. https://doi.org/10.1016/0378-3812(82)80002-2.

PENG, D.-Y.; ROBINSON, D. B. A New Two-Constant Equation of State. **Industrial & Engineering Chemistry Fundamentals**, vol. 15, no. 1, p. 59–64, Feb. 1976. https://doi.org/10.1021/i160057a011.

PETROBRAS. Teste de Longa Duração e Sistemas de Produção Antecipada de Libra – Bacia de Santos – II. Caracterização da Atividade. 2015. Available at: http://licenciamento.ibama.gov.br/Petroleo/Producao/Producao%20-

%20Bacia%20de%20Santos%20-%20TLD%20e%20SPAs%20de%20Libra%20-%20Petrobras/EIA/II_2_CaracAtividade/II_2-CaracAtividade.pdf. Accessed on: 19 Sep. 2019.

PICARD, D. J.; BISHNOI, P. R. Calculation of the thermodynamic sound velocity in two-phase multicomponent fluids. **International Journal of Multiphase Flow**, vol. 13, no. 3, p. 295–308, May 1987. https://doi.org/10.1016/0301-9322(87)90050-4.

RAANES, P. N.; STORDAL, A. S.; EVENSEN, G. Revising the stochastic iterative ensemble smoother. **Nonlinear Processes in Geophysics**, vol. 26, no. 3, p. 325–338, 17 Sep. 2019. https://doi.org/10.5194/npg-26-325-2019.

RAMMAY, Muzammil H.; ELSHEIKH, A. H.; CHEN, Y. Identifiability of Model Discrepancy Parameters in History Matching. *In*: SPE RESERVOIR SIMULATION CONFERENCE, 2019. **SPE Reservoir Simulation Conference** [...]. Galveston, Texas, USA: Society of Petroleum Engineers, 2019. DOI 10.2118/193838-MS. Available at: http://www.onepetro.org/doi/10.2118/193838-MS. Accessed on: 3 Apr. 2019.

RAMMAY, Muzammil Hussain; ELSHEIKH, A. H.; CHEN, Y. Flexible iterative ensemble smoother for calibration of perfect and imperfect models. **Computational Geosciences**, vol. 25, no. 1, p. 373–394, Feb. 2021. https://doi.org/10.1007/s10596-020-10008-z.

RAMMAY, Muzammil Hussain; ELSHEIKH, A. H.; CHEN, Y. Quantification of prediction uncertainty using imperfect subsurface models with model error estimation. **Journal of Hydrology**, 2019. .

RANAZZI, P. H.; SAMPAIO, M. A. Influence of the Kalman gain localization in adaptive ensemble smoother history matching. **Journal of Petroleum Science and Engineering**, vol. 179, p. 244–256, Aug. 2019. https://doi.org/10.1016/j.petrol.2019.04.079.

REDLICH, Otto.; KWONG, J. N. S. On the Thermodynamics of Solutions. V. An Equation of State. Fugacities of Gaseous Solutions. **Chemical Reviews**, vol. 44, no. 1, p. 233–244, Feb. 1949. https://doi.org/10.1021/cr60137a013.

RIOS, V. S.; SANTOS, L. O. S.; QUADROS, F. B.; SCHIOZER, D. J. New upscaling technique for compositional reservoir simulations of miscible gas injection. **Journal of Petroleum Science and Engineering**, vol. 175, p. 389–406, Apr. 2019. https://doi.org/10.1016/j.petrol.2018.12.061.

RIOS, V.; SANTOS, L.; ESPÓSITO, R. Methodology to incorporate the miscibility on the adjust of the equations of state. *In*: RIO OIL & GAS EXPO AND CONFERENCE 2016, 2016. **Rio Oil & Gas Expo and Conference 2016** [...]. [*S. l.*: *s. n.*], 2016.

ROBINSON, D. B.; PENG, D.-Y. **The characterization of the heptanes and heavier fractions for the GPA Peng-Robinson programs**, n. RR-028. Edmonton, Alberta: Gas processors association, 1978. Available at: https://gpamidstream.org/publications/item/?id=113. Accessed on: 22 May 2020.

SAGITOV, I.; STEPHEN, K. D. Optimizing the Integration of 4D Seismic Data in History Matching: Which Data Should We Compare? *In*: EAGE ANNUAL CONFERENCE & EXHIBITION INCORPORATING SPE EUROPEC, 2013. **EAGE Annual Conference & Exhibition incorporating SPE Europec** [...]. London, UK: Society of Petroleum Engineers,

2013. DOI 10.2118/164852-MS. Available at: http://www.onepetro.org/doi/10.2118/164852-MS. Accessed on: 11 Mar. 2019.

SAKOV, P.; BERTINO, L. Relation between two common localisation methods for the EnKF. **Computational Geosciences**, vol. 15, no. 2, p. 225–237, Mar. 2011. https://doi.org/10.1007/s10596-010-9202-6.

SALIMI, M.; BAHRAMIAN, A. The Prediction of the Speed of Sound in Hydrocarbon Liquids and Gases: The Peng-Robinson Equation of State Versus SAFT-BACK. **Petroleum Science and Technology**, vol. 32, no. 4, p. 409–417, 16 Feb. 2014. https://doi.org/10.1080/10916466.2011.580301.

SHABANI, M. R.; RIAZI, M. R.; SHABAN, H. I. Use of velocity of sound in predicting thermodynamic properties of dense fluids from cubic equations of state. **The Canadian Journal of Chemical Engineering**, vol. 76, no. 2, p. 281–289, Apr. 1998. https://doi.org/10.1002/cjce.5450760217.

SILVA, E. P. A.; DAVÓLIO, A.; DOS SANTOS, M. S.; SCHIOZER, D. J. 4D petro-elastic modeling based on a pre-salt well. **Interpretation**, vol. in press, 2020. .

SILVA NETO, Gilson M.; DAVOLIO, A.; SCHIOZER, D. J. 3D seismic data assimilation to reduce uncertainties in reservoir simulation considering model errors. **Journal of Petroleum Science and Engineering**, vol. 189, p. 106967, 1 Jun. 2020. https://doi.org/10.1016/j.petrol.2020.106967.

SILVA NETO, Gilson M.; RIOS, V. de S.; DAVOLIO, A.; SCHIOZER, D. J. Improving fluid modeling representation for seismic data assimilation in compositional reservoir simulation. **Journal of Petroleum Science and Engineering**, vol. 194, p. 107446, Nov. 2020. https://doi.org/10.1016/j.petrol.2020.107446.

SILVA NETO, Gilson Moura; SOARES, R. V.; EVENSEN, G.; DAVOLIO, A.; SCHIOZER, D. J. Subspace Ensemble Randomized Maximum Likelihood with Local Analysis for Time-Lapse-Seismic-Data Assimilation. **SPE Journal**, , p. 1–21, 1 Feb. 2021. https://doi.org/10.2118/205029-PA.

SKJERVHEIM, J.; EVENSEN, G. An Ensemble Smoother for Assisted History Matching. *In*: SPE RESERVOIR SIMULATION SYMPOSIUM, 2011. **SPE Reservoir Simulation Symposium** [...]. The Woodlands, Texas, USA: Society of Petroleum Engineers, 2011. DOI 10.2118/141929-MS. Available at: http://www.onepetro.org/doi/10.2118/141929-MS. Accessed on: 12 Mar. 2019.

SKJERVHEIM, J.-A.; EVENSEN, G.; AANONSEN, S. I.; RUUD, B. O.; JOHANSEN, T.-A. Incorporating 4D Seismic Data in Reservoir Simulation Models Using Ensemble Kalman Filter. *In*: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, 2005. **SPE Annual Technical Conference and Exhibition** [...]. Dallas, Texas: Society of Petroleum Engineers, 2005. DOI 10.2118/95789-MS. Available at: http://www.onepetro.org/doi/10.2118/95789-MS. Accessed on: 20 Feb. 2019.

SKJERVHEIM, J.-A.; EVENSEN, G.; AANONSEN, S. I.; RUUD, B. O.; JOHANSEN, T.-A. Incorporating 4D Seismic Data in Reservoir Simulation Models Using Ensemble Kalman Filter. **SPE Journal**, vol. 12, no. 03, p. 282–292, 1 Sep. 2007. https://doi.org/10.2118/95789-PA.

SOARES, R. V.; LUO, X.; EVENSEN, G. Sparse Representation of 4D Seismic Signal Based on Dictionary Learning. *In*: SPE NORWAY ONE DAY SEMINAR, 2019. **SPE Norway One Day Seminar** [...]. Bergen, Norway: Society of Petroleum Engineers, 2019. DOI 10.2118/195599-MS. Available at: http://www.onepetro.org/doi/10.2118/195599-MS. Accessed on: 9 Jan. 2020.

SOARES, R. V.; MASCHIO, C.; SCHIOZER, D. J. A novel localization scheme for scalar uncertainties in ensemble-based data assimilation methods. **Journal of Petroleum Exploration and Production Technology**, vol. 9, no. 4, p. 2497–2510, Dec. 2019. https://doi.org/10.1007/s13202-019-0727-5.

SOARES, R. V.; MASCHIO, C.; SCHIOZER, D. J. Applying a localization technique to Kalman Gain and assessing the influence on the variability of models in history matching. **Journal of Petroleum Science and Engineering**, vol. 169, p. 110–125, 2018. .

SOAVE, G. Equilibrium constants from a modified Redlich-Kwong equation of state. **Chemical Engineering Science**, vol. 27, no. 6, p. 1197–1203, Jun. 1972. https://doi.org/10.1016/0009-2509(72)80096-4.

STANDING, M. B. Oil-system correlations. **Petroleum production handbook**, vol. 2, no. part 19, 1962. .

STEPHEN, K. D. Scale and Process Dependent Model Errors in Seismic History Matching. **Oil & Gas Science and Technology - Revue de l'IFP**, vol. 62, no. 2, p. 123–135, Mar. 2007. https://doi.org/10.2516/ogst:2007011.

STEPHEN, Karl D.; KAZEMI, A. Improved normalization of time-lapse seismic data using normalized root mean square repeatability data to improve automatic production and seismic history matching in the Nelson field: Seismic history matching by NRMS calibration. **Geophysical Prospecting**, vol. 62, no. 5, p. 1009–1027, Sep. 2014. https://doi.org/10.1111/1365-2478.12109.

STEPHEN, Karl D.; MACBETH, C. Reducing Reservoir Prediction Uncertainty by Updating a Stochastic Model Using Seismic History Matching. **SPE Reservoir Evaluation & Engineering**, vol. 11, no. 06, p. 991–999, 1 Dec. 2008. https://doi.org/10.2118/100295-PA.

STEPHEN, Karl D.; SHAMS, A.; MACBETH, C. Faster Seismic History Matching in a United Kingdom Continental Shelf Reservoir. **SPE Reservoir Evaluation & Engineering**, vol. 12, no. 04, p. 586–594, 1 Aug. 2009. https://doi.org/10.2118/107147-PA.

STEPHEN, Karl D.; SOLDO, J.; MACBETH, C.; CHRISTIE, M. A. Multiple Model Seismic and Production History Matching: A Case Study. **SPE Journal**, vol. 11, no. 04, p. 418–430, 1 Dec. 2006. https://doi.org/10.2118/94173-PA.

SUN, W.; VINK, J. C.; GAO, G. A Practical Method to Mitigate Spurious Uncertainty Reduction in History Matching Workflows with Imperfect Reservoir Models. *In*: SPE RESERVOIR SIMULATION CONFERENCE, 2017. **SPE Reservoir Simulation Conference** [...]. Montgomery, Texas, USA: Society of Petroleum Engineers, 2017. DOI 10.2118/182599-MS. Available at: http://www.onepetro.org/doi/10.2118/182599-MS. Accessed on: 19 Feb. 2019.

TAHA, T.; WARD, P.; PEACOCK, G.; HERITAGE, J.; BORDAS, R.; ASLAM, U.; WALSH, S.; HAMMERSLEY, R.; GRINGARTEN, E. History Matching Using 4D Seismic in an Integrated Multi-Disciplinary Automated Workflow. *In*: SPE RESERVOIR CHARACTERISATION AND SIMULATION CONFERENCE AND EXHIBITION, 17 Sep. 2019. **SPE Reservoir Characterisation and Simulation Conference and Exhibition** [...]. Abu Dhabi, UAE: SPE, 17 Sep. 2019. p. D031S022R001. DOI 10.2118/196680-MS. Available at: https://onepetro.org/SPERCSC/proceedings/19RCSC/3-19RCSC/Abu%20Dhabi,%20UAE/219080. Accessed on: 9 Apr. 2021.

TAHANI, H. **Determination of the Velocity of Sound in Reservoir Fluids Using an Equation of State**. 2012. 1–217 f. Doctoral Thesis – Heriot-Watt University, Edinburgh, United Kingdom, 2012. Available at: http://www.ros.hw.ac.uk/handle/10399/2537. Accessed on: 18 May 2018.

TARANTOLA, A. **Inverse problem theory and methods for model parameter estimation**. Philadelphia, PA: SIAM, Society for Industrial and Applied Mathematics, 2005.

TOLSTUKHIN, E.; LYNGNES, B.; SUDAN, H. H. Ekofisk 4D Seismic - Seismic History Matching Workflow. *In*: SPE EUROPEC/EAGE ANNUAL CONFERENCE, 2012. **SPE Europec/EAGE Annual Conference** [...]. Copenhagen, Denmark: Society of Petroleum Engineers, 2012. DOI 10.2118/154347-MS. Available at: http://www.onepetro.org/doi/10.2118/154347-MS. Accessed on: 26 Mar. 2019.

TRANI, M.; ARTS, R.; LEEUWENBURGH, O. Seismic History Matching of Fluid Fronts Using the Ensemble Kalman Filter. **SPE Journal**, vol. 18, no. 01, p. 159–171, 1 Feb. 2013. https://doi.org/10.2118/163043-PA.

ULLMANN DE BRITO, D.; CALETTI, L.; MORAES, R. Incorporation of 4D Seismic in the Re-construction and History Matching of Marlim Sul Deep Water Field Flow Simulation Model. *In*: SPE EUROPEC/EAGE ANNUAL CONFERENCE AND EXHIBITION, 2011. **SPE EUROPEC/EAGE Annual Conference and Exhibition** [...]. Vienna, Austria: Society of Petroleum Engineers, 2011. DOI 10.2118/143048-MS. Available at: http://www.onepetro.org/doi/10.2118/143048-MS. Accessed on: 26 Mar. 2019.

VAN DER WAALS, J. D. **Over de Continuiteit van den Gas-en Vloeistoftoestand**. [*S. l.*]: Sijthoff, 1873. vol. 1, .

VASQUEZ, G. F.; MORSCHBACHER, M. J.; JUSTEN, J. C. R. Experimental efforts to access 4D feasibility and interpretation issues of Brazilian presalt carbonate reservoirs. **Interpretation**, vol. 7, no. 4, p. SH1–SH18, 1 Nov. 2019. https://doi.org/10.1190/INT-2018-0203.1.

VINK, J. C.; GAO, G.; CHEN, C. Bayesian Style History Matching: Another Way to Under-Estimate Forecast Uncertainty? *In*: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, 2015. **SPE Annual Technical Conference and Exhibition** [...]. Houston, Texas, USA: Society of Petroleum Engineers, 2015. DOI 10.2118/175121-MS. Available at: http://www.onepetro.org/doi/10.2118/175121-MS. Accessed on: 19 Feb. 2019.

WANG, Z.; NUR, A. M.; BATZLE, M. L. Acoustic Velocities in Petroleum Oils. *In*: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, 1988. **SPE Annual Technical Conference and Exhibition** [...]. Houston, Texas: Society of Petroleum Engineers, 1988. DOI

10.2118/18163-MS. Available at: http://www.onepetro.org/doi/10.2118/18163-MS. Accessed on: 26 Mar. 2020.

WHITCOMBE, D. N. Elastic impedance normalization. **GEOPHYSICS**, vol. 67, no. 1, p. 60–62, Jan. 2002. https://doi.org/10.1190/1.1451331.

WOJNAR, K.; S?TROM, J.; MUNCK, T. F.; STUNELL, M.; SVILAND-ØSTRE, S.; HETTERVIK, K. O.; KRISTOFFERSEN, I.; ANDORSEN, K. Quantitative 4D Seismic Assisted History Matching Using Ensemble-Based Methods on the Vilje Field. *In*: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, 19 Oct. 2020. **SPE Annual Technical Conference and Exhibition** [...]. Virtual: SPE, 19 Oct. 2020. p. D041S060R002. DOI 10.2118/201750-MS. Available at: https://onepetro.org/SPEATCE/proceedings/20ATCE/4-20ATCE/Virtual/451256. Accessed on: 9 Apr. 2021.

XU, S.; PAYNE, M. A. Modeling elastic properties in carbonate rocks. **The Leading Edge**, vol. 28, no. 1, p. 66–74, Jan. 2009. https://doi.org/10.1190/1.3064148.

YE, S.; LAGOURETTE, B.; ALLIEZ, J.; SAINT-GUIRONS, H.; XANS, P.; MONTEL, F. Reservoir Fluid Characterization Using Ultrasonic Velocity. *In*: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, 1991., SPE-22862-MS. **SPE Annual Technical Conference and Exhibition** [...]. Dallas: Society of Petroleum Engineers, 1991. DOI 10.2118/22862-MS. Available at: http://www.onepetro.org/doi/10.2118/22862-MS. Accessed on: 18 May 2018.

YIN, Z.; FENG, T.; MACBETH, C. Fast assimilation of frequently acquired 4D seismic data for reservoir history matching. **Computers & Geosciences**, vol. 128, p. 30–40, Jul. 2019. https://doi.org/10.1016/j.cageo.2019.04.001.

ZHANG, Yanfen; OLIVER, D. S. Evaluation and error analysis: Kalman gain regularization versus covariance regularization. **Computational Geosciences**, vol. 15, no. 3, p. 489–508, Jun. 2011. https://doi.org/10.1007/s10596-010-9218-y.

ZHANG, Yanhui; LEEUWENBURGH, O. Image-oriented distance parameterization for ensemble-based seismic history matching. **Computational Geosciences**, vol. 21, no. 4, p. 713–731, Aug. 2017. https://doi.org/10.1007/s10596-017-9652-1.

ZHAO, Y.; REYNOLDS, A. C.; LI, G. Generating Facies Maps by Assimilating Production Data and Seismic Data With the Ensemble Kalman Filter. *In*: SPE SYMPOSIUM ON IMPROVED OIL RECOVERY, 2008. **SPE Symposium on Improved Oil Recovery** [...]. Tulsa, Oklahoma, USA: Society of Petroleum Engineers, 2008. DOI 10.2118/113990-MS. Available at: http://www.onepetro.org/doi/10.2118/113990-MS. Accessed on: 22 Jun. 2020.

APPENDIX A - 3D SEISMIC DATA ASSIMILATION TO REDUCE UNCERTAINTIES IN RESERVOIR SIMULATION CONSIDERING MODEL ERRORS

Authors:

Gilson M. Silva Neto

Alessandra Davolio

Denis J. Schiozer

A.1 Abstract

Decisions in petroleum reservoir management usually involve a high level of uncertainties. Therefore, information of different types is used to calibrate reservoir models for the production forecasts and decision analysis. One source of information is 3D seismic, which is highly correlated to petrophysical properties. These properties are a major source of uncertainty. The incorporation of 3D seismic data in flow models is affected by errors caused by discretization, scale differences, seismic modeling uncertainties, seismic propagation related distortions, among others. Nevertheless, these errors are commonly neglected in conventional model calibration workflows. This work treats seismic resolution loss as a form of model error that needs to be considered in the data assimilation process. In our tests, we used synthetic data from a realistic benchmark case. First, we extended the methodology proposed by Oliver and Alfonzo (2018a) to 3D seismic data assimilation. We focus on the model improvement by estimating a "total" observation error covariance matrix. Furthermore, we reduced the influence of systematic errors by including a simple analytical function in our forward model. The function is defined based on physical premises and the parameter is calibrated in the data assimilation workflow. This procedure increases the dimension of the problem. The error covariance update improved the reservoir volume characterization in all of our tests. Moreover, we show that the update provides a way to improve the determination of the residual weights in the data assimilation problem. These weights are difficult to define in practice and the results were relatively insensitive to the initial values. By using the proposed methodology, we were

able to improve the reservoir volume calibration using relatively low-resolution data. If the correlated errors were neglected, the data assimilation would lead to implausible parameter distributions.

Abbreviations:

ESMDA – Ensemble Smoother with Multiple data Assimilations

OF – Objective Function

PEM – Petro-elastic Model

VOIP - Volume of Oil in Place

Keywords:

Model calibration; ensemble data assimilation; history matching; model error; model improvement; ESMDA

A.2 Introduction

Oil and gas production forecasts based on numerical flow simulation models are used in the decision-making process involving important investments and complex challenges. Ultimately aiming to improve the outcome of the decisions made based on the predicted production, it is important to use all the relevant information available to reduce the uncertainty in these forecasts. Therefore, data assimilation methods that apply field measured data have become increasingly popular in reservoir engineering (OLIVER, D. S.; REYNOLDS; LIU, 2008; OLIVER, Dean S.; CHEN, 2011). Here, we apply the probabilistic ensemble-based method ESMDA (EMERICK, Alexandre A.; REYNOLDS, 2013a) to reduce the uncertainty of our reservoir flow models through seismic data assimilation.

3D seismic is a well-known source of information in reservoir modeling, providing crucial data to structural modeling and petrophysical properties distribution (DOYEN, 2007). In this work, the 3D seismic data assimilation is performed directly in the simulation models using ESMDA. It is important to assure that these models represent the 3D baseline data before the 4D (or time-lapse) data is incorporated. Similar calibration of flow simulation models using

3D seismic data have been reported (EMERICK, Alexandre A., 2016; EMERICK, Alexandre A.; REYNOLDS, 2013b; LEEUWENBURGH; BROUWER; TRANI, 2011; STEPHEN, Karl D. *et al.*, 2006).

We recognize that all models are simplifications of the real systems. In other words, all data assimilation workflows will be affected by model errors. Some common sources of model errors are incorrect parametrization, numerical discretization, lack of important physical phenomena description, etc. (CARRASSI; VANNITSEM, 2016; HARLIM, 2013). Focusing specifically on seismic data assimilation, there will be deviations caused by scale differences, discretization, petro-elastic model (PEM) uncertainties, seismic propagation-related distortions, among others. Nevertheless, these sources of errors are commonly neglected in conventional uncertainty reduction workflows (DONG; GU; OLIVER, 2006; EMERICK, Alexandre A.; REYNOLDS, 2013b; LEEUWENBURGH; BROUWER; TRANI, 2011; SKJERVHEIM, Jan-Arild *et al.*, 2005). This may cause incorrect adjustments and uncertainty underestimation, due to the assimilation of large datasets without considering relevant errors (EVENSEN, G., 2018; OLIVER, D.S.; ALFONZO, 2018a; SUN; VINK; GAO, 2017).

There are several studies from other research areas that address data assimilation considering model errors (HARLIM, 2013). Although this topic is less addressed in the petroleum literature, there is previous work which considered model errors in the data assimilation workflows. Some of them require the knowledge of a reference model, which has reduced model errors. Therefore, it can be used to quantify bias or to model the covariance matrix of observation errors (O'SULLIVAN, 2004; RAMMAY, Muzammil H.; ELSHEIKH; CHEN, 2019; STEPHEN, K. D., 2007; STEPHEN, Karl D.; SHAMS; MACBETH, 2009). This requirement limits the application of such methodologies in some practical cases. Sun et al. (2017) and Vink et al. (2015) proposed ways to inflate the covariance matrix of errors to mitigate the influence of model imperfections. Furthermore, Evensen (2018) has shown that one way to treat model errors is to increase the dimension of the problem, e.g. by estimating another set of parameters related to model errors. This leads to weaker updates of the uncertain parameters but improves the forecasts.

Oliver and Alfonzo (2018a) proposed a workflow that comprises model construction, model calibration, model criticism, and model improvement to perform data assimilation considering model imperfections. Their model diagnostics method was based on the analysis of the magnitude and the spatial or temporal distribution of the mismatch of the last updated ensemble. According to the author's methodology, if the diagnostics indicate that there are important model errors, it is possible to repeat the data assimilation. In the subsequent calibrations, a "total" observation error covariance matrix should be estimated from the residuals of the last conditioned ensemble. Later, Rammay et al. (2019) investigated different approaches to deal with model errors in production data history matching, including Oliver and Alfonzo's methodology. Oliver and Alfonzo (2018b) applied this methodology to time-lapse seismic data assimilation, but some of their premises do not hold in our case.

An adaptation of Oliver and Alfonzo's (2018a) methodology to 3D data assimilation is proposed here. The main source of model errors is the simplified forward model, which comprises only the reservoir flow simulator and a PEM to compute acoustic impedance, neglecting the effects of resolution loss and seismic inversion. We apply our workflow to a realistic benchmark case called UNISIM-I-H, which is based on the data from the Namorado Field, an offshore turbidite sandstone reservoir in the Campus Basin, Brazil (AVANSI; SCHIOZER, 2015).

The specific objectives of our work are: (1) show the possible impacts of resolution loss in 3D seismic data assimilation; (2) extend the application of the methodology proposed by Oliver and Alfonzo (2018a) to 3D seismic data assimilation; (3) propose a way to reduce systematic model errors; (4) present a method to update the covariance matrix of "total" observation errors in 3D seismic data assimilation, allowing for heterogeneous variance; and (5) show that this methodology provides a way to estimate the error covariance that is not sensitive to the initial covariance values.

A.3 Theoretical background

This section provides a brief description of three important elements that are applied in the proposed methodology. First, we address the data assimilation algorithm employed here, namely Ensemble Smoother with Multiple Data Assimilations. Then, the model errors treatment and how it affects the definition of the "total" observation error covariance matrix is addressed.

A.3.1 Ensemble Smoother with Multiple data Assimilations (ESMDA)

The ensemble-based methods have been investigated and applied extensively for conditioning reservoir models to production and seismic data (JUNG *et al.*, 2018). Among them, the Ensemble Smoother with Multiple Data Assimilations (EMERICK, Alexandre, 2012; EMERICK, Alexandre A.; REYNOLDS, 2013a) has been applied in reservoir engineering problems, due to the simple formulation and the easy implementation using popular commercial tools. For reservoir engineering applications, the ESMDA has two advantages over the Ensemble Kalman Filter (EVENSEN, Geir, 1994). First, it does not perform sequential data assimilation in time. Secondly, it does not update the vector of states. Furthermore, the method's iterative scheme leads to better matches than the Ensemble Smoother (EVENSEN, Geir, 2009; EVENSEN, Geir; VAN LEEUWEN, 1996) in reservoir data assimilation. In the following paragraphs, we will show the basic formulation of the method. For a full description of ESMDA, refer to (EMERICK, Alexandre, 2012; EMERICK, Alexandre A.; REYNOLDS, 2013a).

Here we represent a model with N_m unknown parameters with an ensemble of N_e members. Moreover, we have a set of N_d observed data to reduce the uncertainty of those parameters. The ESMDA analysis equation with Kalman gain localization (ZHANG, Yanfen; OLIVER, 2011) is:

$$\boldsymbol{m}_j^{l+1} = \boldsymbol{m}_j^l + (\boldsymbol{R}_{md} \circ \boldsymbol{K}^l)\left(\boldsymbol{d}_{obs} + \boldsymbol{e}_j^l - \boldsymbol{g}(\boldsymbol{m}_j^l)\right) \tag{A.1}$$

where $\boldsymbol{m}_j^l$ is the vector containing the parameters of the jth model realization at the lth data assimilation of the ESMDA method, $\boldsymbol{R}_{md}$ is the localization matrix, $\circ$ denotes the Schur product, $\boldsymbol{d}_{obs}$ is the vector of observed data, $\boldsymbol{e}_j^l \in \mathbb{R}^{N_d}$ is a vector randomly drawn from the distribution $\mathcal{N}(0, \alpha^{l+1} C_e)$, $\boldsymbol{C}_e$ is the error covariance matrix and $\boldsymbol{g}(\boldsymbol{m}_j^l)$ is the simulated data using the vector of parameters $\boldsymbol{m}_j^l$. The Kalman gain is defined in equation (A.2),

$$\boldsymbol{K}^l = \widetilde{\boldsymbol{C}}_{md}^l\left(\widetilde{\boldsymbol{C}}_{dd}^l + \alpha^{l+1}\boldsymbol{C}_e\right)^{-1}, \tag{A.2}$$

where α^l is known as the error covariance inflation factor of the lth step, $\widetilde{\boldsymbol{C}}_{md}^l$ and $\widetilde{\boldsymbol{C}}_{dd}^l$ are covariance matrices computed at each step using equations (A.3) and (A.4).

$$\widetilde{\boldsymbol{C}}_{md}^l = \Delta\boldsymbol{M}^l(\Delta\boldsymbol{D}^l)^T \tag{A.3}$$

$$\widetilde{\boldsymbol{C}}_{dd}^l = \Delta\boldsymbol{D}^l(\Delta\boldsymbol{D}^l)^T \tag{A.4}$$

Furthermore, $\Delta\boldsymbol{M}^l$ and $\Delta\boldsymbol{D}^l$ are computed at each step using equations (A.5) and (A.6).

$$\Delta\boldsymbol{M}^l = \frac{1}{\sqrt{N_e - 1}}\left[\boldsymbol{m}_1^l - \overline{\boldsymbol{m}^l} \ldots \boldsymbol{m}_{N_e}^l - \overline{\boldsymbol{m}^l}\right] \tag{A.5}$$

$$\Delta \boldsymbol{D}^l = \frac{1}{\sqrt{N_e - 1}} \left[\boldsymbol{g}(\boldsymbol{m}_1^l) - \overline{g(\boldsymbol{m}^l)} \ldots \boldsymbol{g}(\boldsymbol{m}_{N_e}^l) - \overline{g(\boldsymbol{m}^l)} \right] \tag{A.6}$$

To perform multiple data assimilations, equation (A.1) is used N_a number of times, which are previously selected. Apart from this, the inflation factor of each step, α^l, must satisfy the equation (A.7).

$$\sum_{l=1}^{N_a} (\alpha^l)^{-1} = 1 \tag{A.7}$$

Among the variables in equation (A.1), the error covariance matrix, $\boldsymbol{C}_e$, is especially important to the data assimilation results because it influences the weights of each data deviation on the parameter modifications. However, its computation considering the data-error correlations is not well defined in petroleum literature. In the following subsection, we address the model errors and how they influence the definition of $\boldsymbol{C}_e$.

A.3.2 Model errors and the error covariance matrix

The measurement errors are not the only source of deviations when the output of the models is compared to the observed data. The model's limited capabilities of accurately predicting the true responses cause what we call model errors. Therefore, even if it was possible to use the "true" parameters, the output of the model would not be exactly the true response of the reservoir. For instance, focusing on seismic data assimilation, common sources of model errors include discretization, scale differences, PEM uncertainties, seismic propagation related distortions, among others.

Oliver and Alfonzo (2018a) proposed a methodology to treat the model errors in a data assimilation process by replacing $\boldsymbol{C}_e$ by a total observation error covariance matrix, $\boldsymbol{C}_D = \boldsymbol{C}_e + \boldsymbol{C}_g$, where $\boldsymbol{C}_g$ is the covariance matrix of model error, that is assumed to be Gaussian with zero mean. In their methodology, $\boldsymbol{C}_D$ is iteratively estimated from the behavior of the observation residuals of the previously calibrated ensemble. Here, we apply this idea in the 3D seismic data assimilation workflow.

Neglecting observation bias, we can assume that our observed data comprise the so-called "true" response of the reservoir and a measurement error, $\boldsymbol{e}_d$. Furthermore, neglecting model bias for now, we assume that our model output is the "true" response summed with a model error $\boldsymbol{e}_g$. Assuming that both errors are Gaussian with zero mean, they are characterized by the covariance matrices $\boldsymbol{C}_e$ and $\boldsymbol{C}_g$. Therefore, the mismatch between the model output and the observed data will be the total observation error:

$$\boldsymbol{g}(\boldsymbol{m}_j^k) - \boldsymbol{d}_{obs} = (\boldsymbol{e}_d)_j + (\boldsymbol{e}_g)_j. \tag{A.8}$$

Oliver and Alfonzo (2018a) proposed estimating the total observation error covariance matrix by the iterative procedure represented by the equation (A.9). The matrix $\boldsymbol{M}^k$ contains the vectors $\boldsymbol{m}_j^k$ in each of its columns, which in turn represents the calibrated parameters from the last ESMDA run, using the total observation error covariance matrix $\widetilde{\boldsymbol{C}}_D^k$. The matrix $\boldsymbol{g}(\boldsymbol{M}^k)$ contains the reservoir simulation results using the last calibrated ensemble. In the first step, $k = 0$, the initial covariance matrix can be defined based on the estimated measurement error covariance matrix, usually assumed to be diagonal. The authors have shown in a simple nonlinear application that this procedure improves the data assimilation results in the presence of model errors and is also able to converge to the true measurement error variance when there is only uncorrelated observation error in the data and no model errors.

$$\widetilde{\boldsymbol{C}}_D^{k+1} = \mathrm{cov}(\boldsymbol{g}(\boldsymbol{M}^k) - \boldsymbol{d}_{\boldsymbol{obs}}). \tag{A.9}$$

In this section, we revisited some key topics that were important to the development of the proposed workflow to assimilate 3D seismic data considering model imperfections. In the next section, we detail each step of the methodology.

A.4 Methodology

The methodology shown here follows the general steps for model improvement and calibration proposed by Oliver and Alfonzo (2018a) and applied to 4D seismic data assimilation by Oliver and Alfonzo (2018b). However, there are some key differences in the assumptions to develop each step of the 3D seismic data assimilation workflow, based on the application in the realistic case that we have investigated. First, in our application, the addition of a systematic error reduction term was necessary to reduce the biased discrepancies originated by the simplification in the forward seismic model. Furthermore, we realized that it was not possible to assume stationarity to simplify the computations involving the covariance matrix of the total observation errors.

Our model improvement and calibration workflow using 3D seismic data are presented in Figure A.1. It comprises four steps, namely data assimilation, model diagnosis, systematic error reduction, and “total” observation error covariance estimation. Each of these steps is detailed in the following subsections.

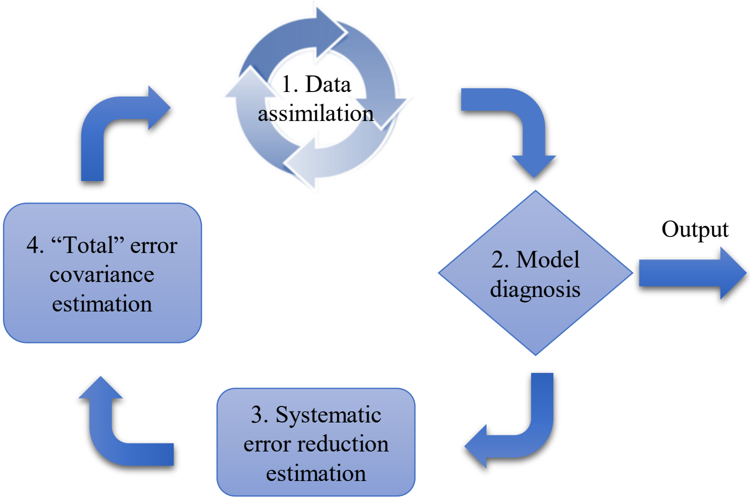


Figure A.1: Model improvement workflow.

A.4.1 Data assimilation

The data assimilation workflow is detailed in Figure A.2. The flow simulator coupled with the PEM and the systematic error reduction represents the improved forward model, whose output is compared to the observed data. The reservoir flow simulator and the petro-elastic model are detailed in sections A.5.1 and A.5.2, respectively. For each ensemble member, the ESMDA method computes the parameter modifications from the deviations between the perturbed vector of observations, $\boldsymbol{d}_{obs} + \boldsymbol{e}$, and $\boldsymbol{d}_{sim} = \boldsymbol{g}(\boldsymbol{m})$. The new ensemble, with altered parameters, is simulated again and this process is repeated until the previously specified number of data assimilations is reached. This workflow follows the same basic steps of previous applications, e.g. (EMERICK, Alexandre A., 2016; EMERICK, Alexandre A.; REYNOLDS, 2013b), except for the systematic error reduction term, which will be detailed in section A.4.3.

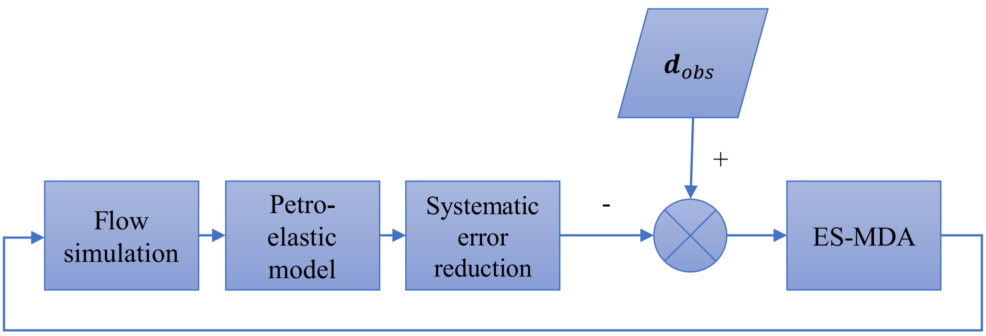


Figure A.2: Data assimilation workflow.

A.4.2 Model diagnosis

There are three basic pieces of evidence of the important influence of model errors in the data assimilation results: large deviations between simulated and observed responses after the calibration, spatial or temporal correlations in these deviations, and resulting implausible parameter values (OLIVER, D.S.; ALFONZO, 2018a). Moreover, it is possible to define quantitative diagnostic metrics based on the comparison of calibrated model responses and the observed data (OLIVER, D. S.; REYNOLDS; LIU, 2008; OLIVER, D.S.; ALFONZO, 2018a; TARANTOLA, 2005).

In this work, we suggest starting the evaluation of the data assimilation quality with the analysis of the posterior parameters' values, considering the previous geological knowledge. When strong model errors are present, it is possible to identify implausible parameter distributions in the calibrated ensemble by comparing them with the prior distributions. Furthermore, we recommend the analysis of the minimum and maximum values of the observed and simulated data from the prior and the calibrated ensembles. To avoid misleading results due to outliers, percentile measures could be used. If there are persistent differences that are not associated with physical parameter updates, this indicates that a systematic error reduction function may be beneficial to the data assimilation process.

A.4.3 Reduction of systematic errors

When the acoustic impedance data obtained directly from the forward PEM is compared with the observed acoustic impedance data from the inversion process, there is an intrinsic error. This error is caused by the resolution loss due to the seismic wave frequency content and the interference between layers. Moreover, the seismic inversion process is uncertain and may result in a relative-acoustic impedance attribute, whose physical unities are arbitrary (STEPHEN, Karl D.; KAZEMI, 2014). Part of these errors is systematic and may lead to inconsistent model updates and exaggerated uncertainty reduction in the data assimilation. By systematic, we mean that it affects the whole data in the same way, causing a significant distortion.

It is possible to mitigate this resolution effect by improving the models applied in the data assimilation workflow. To do so, one can apply a filter after the PEM that aims to limit the frequency content of the simulated signal to the realistic spectrum, e.g. (FAHIMUDDIN; AANONSEN; SKJERVHEIM, 2010b; OLIVER, D.S.; ALFONZO, 2018b). However, this tends to increase the complexity of the simulation model and the computational costs.

Therefore, other applications neglect this resolution loss in the forward model, e.g. (EMERICK, Alexandre A.; REYNOLDS, 2013b; GOSSELIN *et al.*, 2003; STEPHEN, K. D., 2007; STEPHEN, Karl D.; MACBETH, 2008). Here, we decided to simplify the model and consider the resolution loss as a source of model errors due to missing physics.

One possible way to mitigate the systematic error effect is to add an observation bias term in the data assimilation process (OLIVER, D.S.; ALFONZO, 2018a), which is similar to increasing the dimension of the problem (EVENSEN, G., 2018). Here, it was possible to consider some physically related premises that helped define a systematic error reduction function. This nonlinear transformation, which is applied after the PEM (Figure A.2), can be interpreted as a resolution distortion proxy. However, since there is a level of subjectivity in the definition of such function, we included an uncertain parameter that defines its format. Thus, this methodology also increased the dimension of the data assimilation problem, similarly to the previously cited methodologies.

It is important to mention that this procedure aims to reduce systematic model errors. However, it is clear that there will be relevant remaining errors that motivate the application of the "total" error covariance matrix estimation, which will be described in the following section.

A.4.4 "Total" error covariance matrix estimation

In a general sense, the total error covariance matrix estimation follows the iterative procedure proposed by Oliver and Alfonzo (2018a) and is shown in equation (A.9). Nevertheless, as expected in most practical applications involving seismic data, $N_d \gg N_e$. Therefore, the direct application of equation (A.9) would lead to a low-rank covariance matrix. To overcome this issue, if it is possible to assume stationarity, one can apply the procedure proposed by Oliver and Alfonzo (2018b). They generated vectors of perturbed data by shifting and recombining the deviations from the previous data assimilation. On the contrary, if the problem is non-stationary, one can apply the shrinkage algorithm of Ledoit and Wolf (2004) or the one presented by Chen et al. (2010), which outperforms the former when $N_d \gg N_e$. Nevertheless, this latter approach may require the storage and inversion of huge matrices, especially in cases that involve large reservoir models.

Here, we propose to fit covariance models to the experimental covariance matrix, enabling heterogeneous variance of the residuals, reducing the storage requirements, and also simplifying the inverse computations. In summary, considering the workflow of Figure A.1, we start the first data assimilation with an arbitrary diagonal covariance matrix, whose elements can be the variance of the measurement errors or inflated variances (SUN; VINK; GAO, 2017; VINK; GAO; CHEN, 2015). After the first data assimilation, we compute the residuals, $\boldsymbol{g}(\boldsymbol{M}^k) - \boldsymbol{d}_{\boldsymbol{obs}}$, and estimate their variances. Then, we obtain the correlation lengths by fitting an analytical model to the experimental correlation matrix of the residuals. Any correlation model could be applied at this step, but the exponential anisotropic function, equation (A.10), fitted well in all of our cases. In equation (A.10), Δx, Δy, and Δz are physical distances of residuals points, while L_1, L_2, and L_3 are calibrated correlation lengths. Alternatively, one could fit an experimental semivariogram of the residuals, leading to equivalent results. Finally, we apply the new covariance matrix, with the estimated diagonal elements and correlation lengths, in a new data assimilation process. These steps are repeated until the covariance model changes are negligible.

$$c(h) = \exp\left(-3\sqrt{\left(\frac{\Delta x}{L_1}\right)^2 + \left(\frac{\Delta y}{L_2}\right)^2 + \left(\frac{\Delta z}{L_3}\right)^2}\right) \quad \textbf{(A.10)}$$

A.5 Application

After describing our methodology, in this section we detail the realistic application through which we generated our results. We start by describing the Benchmark case, following with our PEM. Then, we detail the seismic data generation process and compare the different seismic data that we used. Finally, the computed data assimilation cases are listed.

A.5.1 Case description

We applied the methodology described in section A.4 to the realistic Benchmark case UNISIM-I-H, which is based on the data from the Namorado Field, an offshore turbidite sandstone reservoir in the Campus Basin, Brazil (AVANSI; SCHIOZER, 2015). All the observed data, which may be referred to as “measured”, was generated from the fine-scale model called UNISIM-I-R, that mimics a real reservoir in our application. This reference model has 3,408,633 active cells, whose approximated dimensions are $25 \times 25 \times 1$ m. The application of our methodology in this realistic Benchmark allows us to compare the results with the reference for validation purposes.

The simulation models were built based on data from the wells, 14 producers and 11 injectors, which means that only limited information about the reference was available during the geological modeling. Furthermore, the simulation models are at a coarser scale, having 38,466 active cells with 100 × 100 × 8 m each (Figure A.3).

Despite the existence of other uncertain attributes in this benchmark case (MASCHIO, C. *et al.*, 2013), we chose to use the synthetic 3D seismic data to calibrate the porosity. There is a strong correlation of this variable with the acoustic impedance. Moreover, we have included the permeabilities in the 3 main directions in the data assimilation process, because its distribution is correlated with the porosity. All these spatially distributed attributes are represented by a prior ensemble of 500 3D models that is available in (MASCHIO, C. *et al.*, 2013).

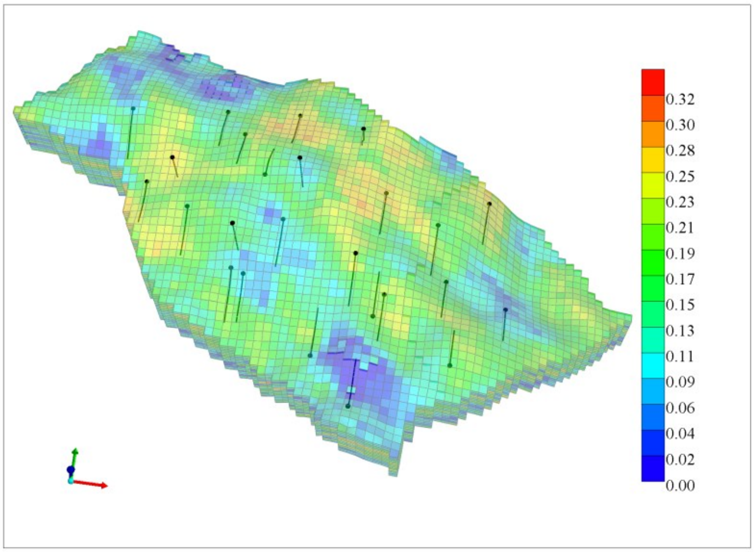


Figure A.3: Simulation model example (porosity).

A.5.2 Petro-elastic model

We used the same PEM in the observed data generation and in the data assimilation workflow because the analysis of model errors caused by petro-elastic uncertainties is outside the scope of this work. This PEM is based on the popular Gassmann equation (GASSMANN, 1951). We considered two minerals in the reservoir, shale and quartz, and their proportions were described by the net-to-gross ratio. This is a common simplification (e.g. (EMERICK, Alexandre Anozé *et al.*, 2007; STEPHEN, K. D., 2007)). Furthermore, the Hashin-Shtrikman bounds (HASHIN; SHTRIKMAN, 1963; apud AVSETH; MUKERJI; MAVKO, 2005) were used to compute the equivalent mineral elastic properties. The dry rock properties were modeled by polynomial equations (EMERICK, Alexandre Anozé *et al.*, 2007), whose pressure and porosity dependencies were similar to the Hertz-Mindlin model (MAVKO; MUKERJI;

DVORKIN, 2009). Finally, the fluid properties were obtained using the Batzle-Wang correlations (BATZLE; WANG, 1992) and Wood's formula (MAVKO; MUKERJI; DVORKIN, 2009).

A.5.3 Seismic data

The synthetic seismic data that we used in this study were generated using the workflow presented in Figure A.4. First, we ran the reference model at the fine-scale and obtained the pressure and saturation fields at the start of the simulation. These results, together with the rock and fluid properties, were used in the PEM to generate the fine-scale P-wave impedance field. Thereafter, this data was transferred to a regular seismic grid, where the seismic forward model, 1D convolution, was computed to obtain the amplitude data. A colored (coloured in the original manuscript) inversion algorithm (LANCASTER; WHITCOMBE, 2000) was applied to obtain the relative I_P data again, which was then transferred to the coarse simulation scale. Thus, a deterministic seismic inversion was considered. For a detailed explanation of this process, refer to (DAVOLIO; SCHIOZER, 2019; DE SOUZA, 2018).

The complete process of Figure A.4 was used to generate the so-called INV data (inverted impedance – step 5), which represents the lower resolution information in our application, mimicking observed real data. We also used, as observed data in our tests, the impedance obtained directly from the PEM (step 2) applied to the output of the upscaled reference model. In this study, this is called the PEM data and it represents the best seismic impedance (as if the seismic data were almost perfect). This kind of data does not exist in real applications, but it was used for validation and comparison purposes. Nevertheless, it is important to mention that when this information is used to calibrate the simulation models, this cannot be considered a model error-free case. The simulation models were built from limited information from the reference and none of the elements of the ensemble is able to reproduce its entire complexity. In other words, some heterogeneities are not represented by the simulation models.

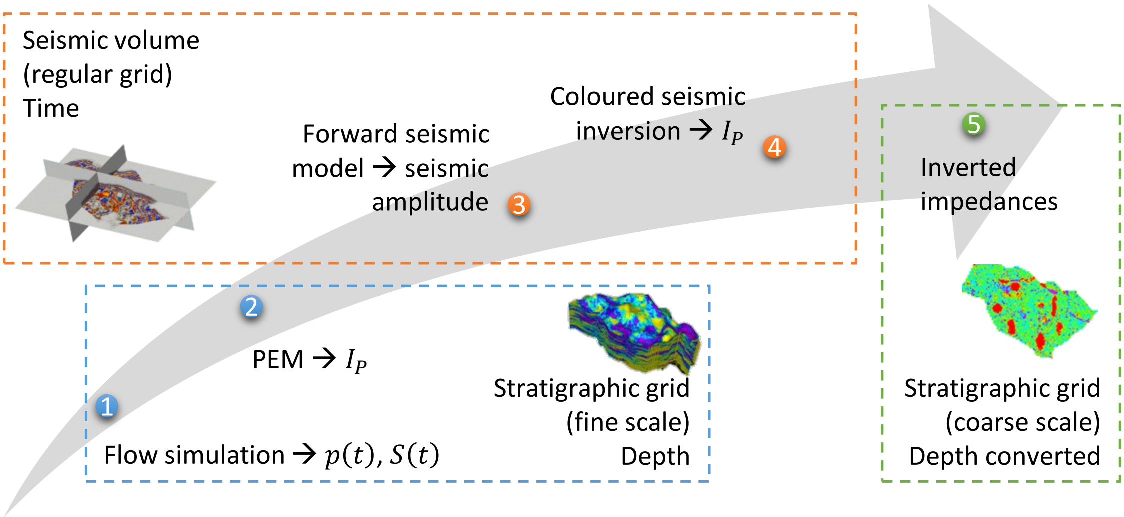


Figure A.4: Synthetic seismic data generation workflow.

Figure A.5 compares the fine-scale reference (a), the PEM (b) and the INV (c) observed impedance data. The differences of the former and the PEM are due to the upscaling of the reference model, which causes subtle value changes and geometric alterations due to discretization. The two main effects of the resolution loss between the PEM (b) and the INV (c) data are the reduction of the high impedance values, due to averaging, and the distortion of the images, due to the inter-layer interference. These deviations were treated in this work as systematic errors and part of the "total" observation errors in the data assimilation process.

A.5.4 Data assimilation cases

We used the PEM and INV data to compute several data assimilations, whose results will be discussed in the next section. To facilitate the reading, we apply the nomenclature presented in Table A.1, where the cases are classified based on the observed data, the presence of the systematic error reduction function and the arbitrary diagonal initial error covariance matrix. We give further detail about the latter in section A.6.1. In Table A.1, DA-PRIOR means the initial dataset before any data assimilation, while all the other cases are assimilating 3D seismic. For instance, DA-INV-SER corresponds to the data assimilation using INV as observed data with the application of the systematic error reduction procedure.

In addition to this nomenclature, we use the term iteration 0 (it. 0) to refer to first data assimilation in the workflow of Figure A.1, in which an arbitrary diagonal error covariance matrix is applied. Furthermore, iterations 1, 2 and 3 refer to the iterative estimation of the covariance matrix.

In all of the data assimilations discussed in section 0, we have applied the ESMDA algorithm discussed in section A.3.1. We used 4 data assimilations in the ESMDA formulation, with decreasing inflation factors: $\alpha^1 = 9.333$, $\alpha^2 = 7$, $\alpha^3 = 4$ and $\alpha^4 = 2$, which are the same used by Emerick and Reynolds (2013b). Due to a large amount of data, we also applied subspace inversion (EMERICK, Alexandre A.; REYNOLDS, 2012), maintaining 99% of the sum of the nonzero singular values. Moreover, we used Kalman gain localization (ZHANG, Yanfen; OLIVER, 2011) with a correlation length of $1000\ m$ in the horizontal plane using the fifth-order distance-based correlation function proposed by Gaspari and Cohn (1999).

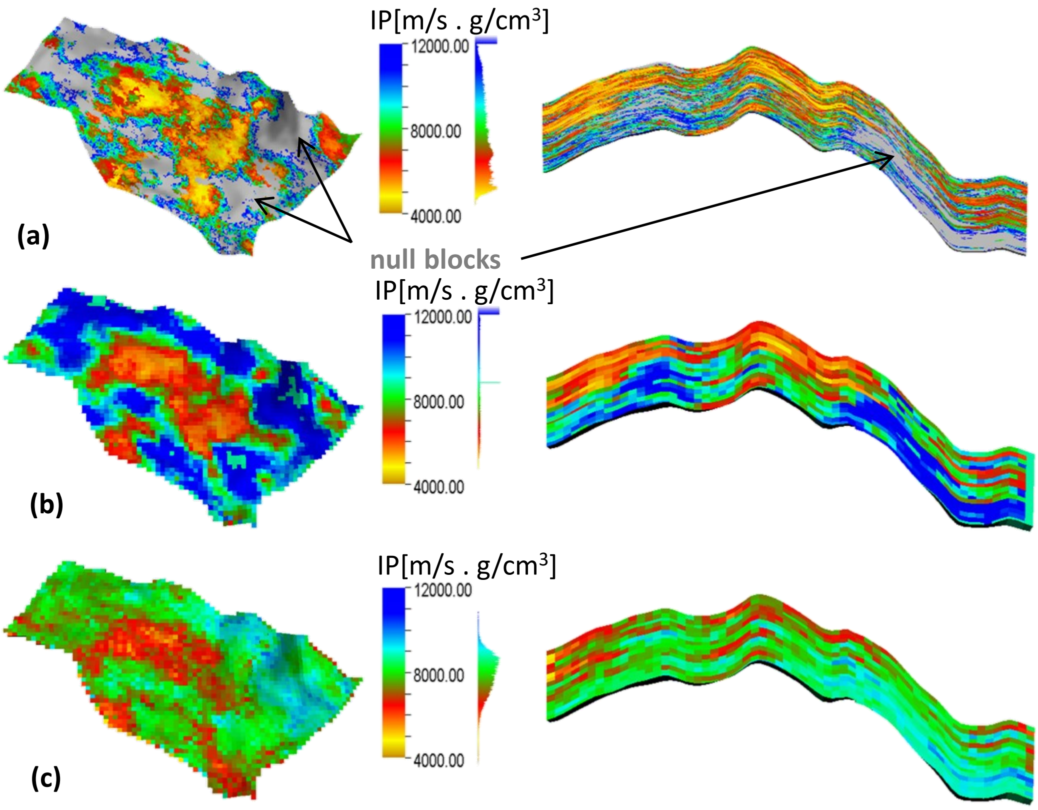


Figure A.5: Impedance observed data comparison: (a) reference fine-scale, (b) PEM data and (c) INV data. The layer 12 on the simulation scale and the correspondent position on the fine-scale model are shown on the left. A cross section of each case is shown on the right.

Table A.1: Data assimilation nomenclature.

Nomenclature	Observed data	Systematic error reduction	Initial covariance
DA-PRIOR	No	No	Usual
DA-PEM	PEM	No	Usual
DA-INV	INV	No	Usual
DA-INV-SER	INV	Yes	Usual
DA-INV-INFL	INV	No	Highly inflated

A.6 Results and discussions

In this section, we show the results of our application and discuss how the proposed methodology helped in reservoir characterization through 3D seismic data assimilation. The results are presented following the methodology order. We compare the results using the PEM data and the INV data to evidence the impact of the resolution loss in the latter. We start by discussing the initial data assimilation, which are the first step in Figure A.1. Thereafter, we address a model diagnosis based on the first data assimilations results. Since we detect the need for a systematic error reduction function, the definition of this function is described subsequently. After that, we present our approach to update the "total" observation error covariance matrix. Finally, the results from the complete methodology are discussed.

Throughout this section, we compare our data assimilation results in terms of the volume of oil in place (VOIP) distribution. The VOIP represents the total volume of oil at the start of the flow simulation, measured at standard conditions (15 ºC and approximately 1 atm). The notation m^3std indicates that the value is in cubic meters measured at standard conditions. This variable was chosen because it reflects the overall change in the porosity field, and it is correlated to the field production performance. Nevertheless, sometimes we represent our results in terms of porosity maps to highlight specific spatial changes.

A.6.1 Initial data assimilation

Following the general workflow of Figure A.1, we start by performing an initial data assimilation using the available observed data. In this first computation, no specific model error treatment is applied. However, it is necessary to specify an initial error covariance matrix that will be used in the ESMDA method. Therefore, we analyzed the influence of the initial value of the diagonal elements of an uncorrelated covariance matrix on the final uncertainty quantification, after the application of the iterative update of a correlated matrix.

In Figure A.6 we can see the comparison of the results of DA-INV and DA-INV-INFL. The first starts with a low variance uncorrelated error covariance matrix (red curves), initially leading to increased VOIP with smaller variance in comparison with the first iteration of the latter. In this case, the iterative procedure (dashed red curves) slightly increased the final VOIP variance and reduced its expected value. The other test (purple curves) started with a diagonal matrix with standard deviations about 15 times greater than the former. As expected, the initial iteration led to higher uncertainty in the final VOIP. However, after 2 iterations of the "total" observation error covariance matrix update, we obtained a response which was approximately equal to the former case. This result was obtained using the INV data, but we experienced similar behavior using the PEM data.

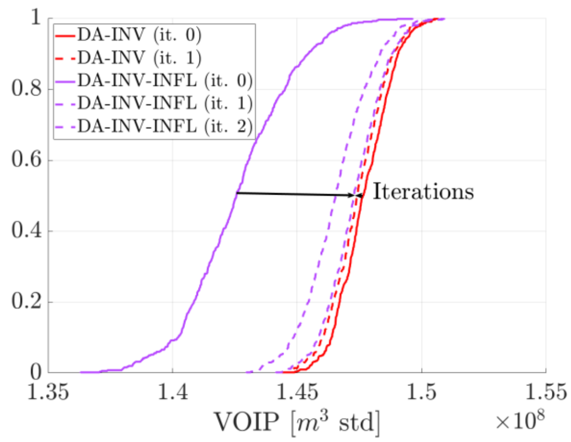


Figure A.6: Influence of the initial variance of the observation errors on the final uncertainty quantification, using the iterative update of the "total" error covariance matrix. The VOIP distributions of the DA-INV cases are in red and the distributions of the DA-INV-INFL cases are in purple. The iterations are shown in dashed lines.

The results of Figure A.6 indicate that the initial value of the diagonal elements of the observation error covariance matrix is not a critical parameter in this methodology. Therefore, we recommend starting with the magnitude of the measurement errors, or inflated values, and perform some steps of the "total" observation error covariance matrix update. In all our applications, 3 iterations were enough to reach approximately stable values of the posterior uncertainty quantification. For the initial data assimilation in each case hereafter (it. 0), an arbitrary uncorrelated error covariance matrix with the standard deviation corresponding to 50% of the standard deviation of the data was applied.

Figure A.7 shows the results of DA-PEM (it. 0) and DA-INV (it. 0). In both cases, the information reduced the VOIP variance and changed its averages to higher values, when compared to the DA-PRIOR. This indicates that our prior ensemble is somewhat pessimistic. In a field application, this could be the final result, if no correlated error impact is detected. However, in the following subsection, we show that our simplified forward model impairs our data assimilation results in the DA-INV (it. 0) case.

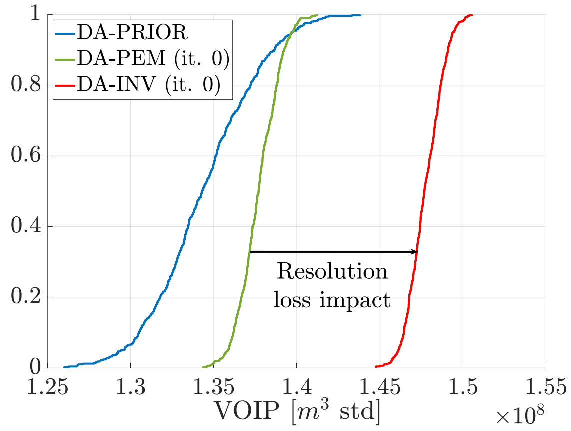


Figure A.7: First data assimilations results.

A.6.2 Model diagnosis

We start the model diagnosis from the results of the first data assimilations, which are summarized in Figure A.7. The calibration DA-INV (it. 0) resulted in volumes higher than all the models from the prior ensemble. This result could imply the poor parametrization of the prior ensemble or the influence of observation errors, due to the difference between the simplified forward model and the seismic data generation process (Figure A.4). This latter is the main impact in our case, which resulted in implausible volume distributions.

In order to check the necessity of systematic error reduction and confirm the last conclusion, we compare the histograms of the observed INV data and the simulated impedance from cases DA-PRIOR and DA-INV (it. 0) in Figure A.8. Note that the full seismic data generation process reduced the high impedance values, that reached 10000 $\frac{m}{s}.\frac{g}{cm^3}$ in the observed INV data and almost 12000 $\frac{m}{s}.\frac{g}{cm^3}$ in both the simulated data. Although the high impedance points frequency was reduced in the DA-INV (it. 0), due to the data assimilation process, the discrepancy of the maxima remains. Since we expect the occurrence of low porosity regions in both the reservoir and the simulation models, this analysis indicates the distortion caused by the resolution loss process, that is not included in our forward model. Since this effect was more pronounced in higher impedances, it led to optimistic and inconsistent model updates (see Figure A.7). This corroborates the application of a systematic error reduction function in the data assimilation involving the INV observed data, to avoid inconsistent parameters updates. This will be addressed in the following subsection.

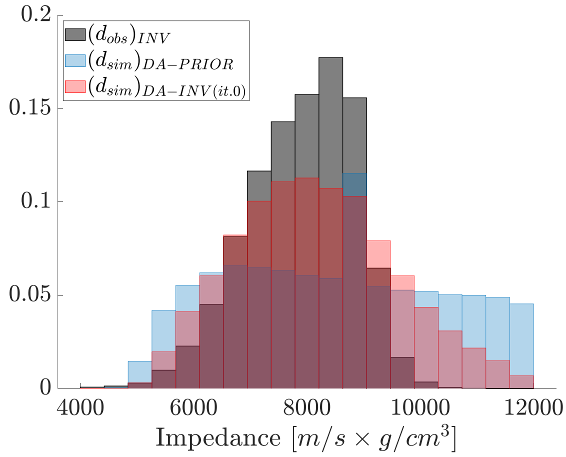


Figure A.8: Comparison of the observed and simulated impedance histograms. The ordinate is the relative frequency, with a maximum of 1 if all the data belonged to the same interval.

A.6.3 Reduction of systematic errors

Our systematic error reduction may be interpreted here as a resolution distortion proxy that was applied to the forward model. To define the format of this function, we assumed two physically related premises, as follows.

Our first assumption is that the minimum and maximum values of the baseline impedance should be approximately the same in the observed and the simulated data. This is analogous to consider that the range and the combinations of the PEM input variables and parameters, such as porosity, saturation, mineral content, and pressure, are the same in the reservoir and the simulation model. Therefore, we can obtain one point of the transformation function from the maximum values of the simulated and observed data. To reduce the influence of outliers, we took the mean of the 0.5% highest values in the DA-PRIOR forward simulations and the observed data to define one point of the transformation function.

The second assumption is that most of the distortions caused by resolution loss happened at the high impedances or worse reservoir regions because the heterogeneities, including non-reservoir rocks, are more frequent in those areas. Apart from the geological knowledge, there is evidence that supports this premise. First, the minimum values from the simulated responses are close to the minimum values in the observed data, despite the deviations in the high impedance points. Second, when the logs from the wells are compared to the observed data in those regions, the deviations are significantly higher at the worst reservoir regions, even when both the log and the seismic data indicate relatively high impedances (see Figure A.9).

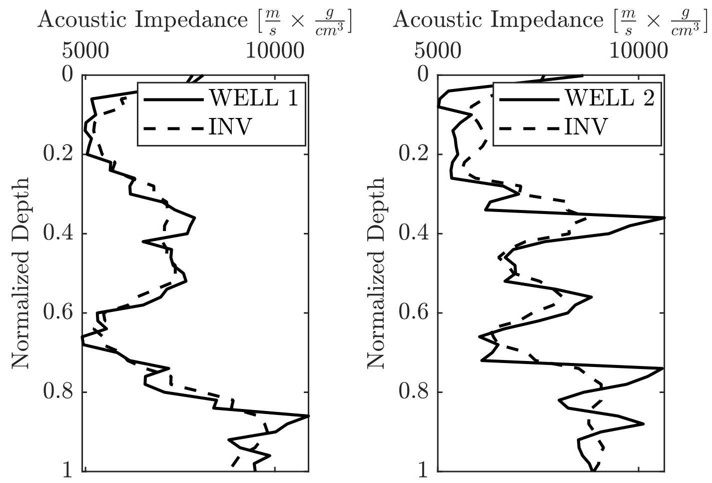


Figure A.9: Comparison of the INV data with two well logs.

Following the previous premises, our systematic error reduction function tends to the original simulated value at low impedance and equalizes the maximum of the observed and simulated data at the higher values. We opted here to use a simple nonlinear function composed of two straight lines of different slopes. The point where the function deviates from the $x = y$ line is used to define the systematic error reduction nonlinear function. It is important to mention that there is a level of subjectivity in this parameter, which justifies a definition of an uncertain range and its inclusion in the data assimilation process. This uncertain range was defined based on the comparison between the simulated DA-PRIOR and the INV observed data.

The three points that define the 2-line function are presented in Table A.2. The function converts the simulated result, x, to a distorted value, y, trying to mimic the resolution loss effect in the data. The first point is a small impedance value that pertains to the $y = x$ line. The second point represents where the attenuation begins, and it is uncertain. The third point represents the relation of the maxima of the simulated and observed data.

Table A.2: Points that define the systematic error reduction function.

$x\left[\frac{m}{s}\cdot\frac{g}{cm^3}\right]$	$y\left[\frac{m}{s}\cdot\frac{g}{cm^3}\right]$	**Source**
5000	$y = x$	Any small impedance value in the $y = x$ line.
$\sim N(8800{,}350)$	$y = x$	Uncertain point where the impedance distortion begins.
11100	10000	0.5% highest values in simulated and observed data.

Taking for instance the mean of the midpoint distribution, the systematic error reduction function follows the equation (A.11).

$$y = ax + b, \text{where } \begin{cases} a = 1 \text{ and } b = 0 \text{ if } x \leq 8800 \\ a = 0.522 \text{ and } b = 4.206 \times 10^3 \text{ if } x > 8800 \end{cases} \quad \textbf{(A.11)}$$

Once the systematic error reduction function was defined, following the methodology illustrated in Figure A.1, we studied a way to update the "total" observation error covariance matrix, which will be addressed in the following subsection.

A.6.4 "Total" observation error covariance matrix update

We started our analysis of the update of the "total" observation error covariance matrix by testing the hypothesis of stationarity. This could simplify our methodology, as reported by Oliver and Alfonzo (2018b). In Figure A.10, we show the behavior of the variance of the residuals of the final ensemble from DA-INV (it. 0) case, where it is noticeable that these residuals' second moment changes regionally.

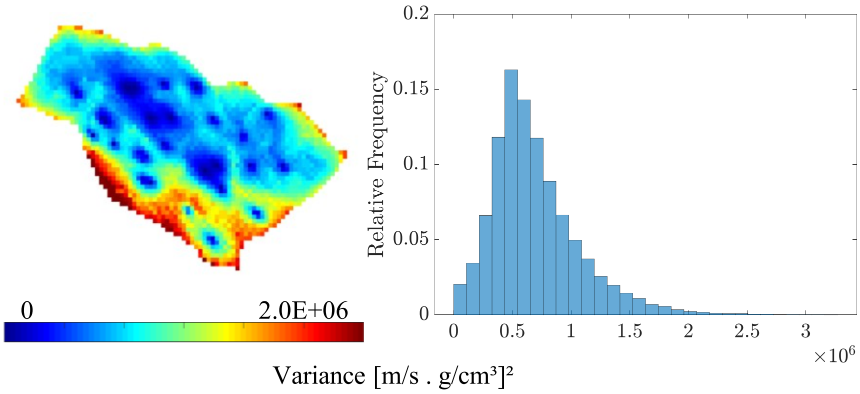


Figure A.10: Variance of the difference between the simulated and observed acoustic impedance data from DA-INV (it. 0) case. The map on the left shows the 2nd layer of the simulation model. The histogram on the right shows the distribution of the variance in the entire model.

The variance distributions shown in Figure A.10 motivated the use of heterogeneous variances in all applications. Therefore, we have decided to update the "total" error covariance matrix using a simplified analytic covariance model fitted to the experimental data, enabling heterogeneous variance. In all our cases, exponential functions with horizontal lengths of roughly $400\,m$ and vertical lengths of approximately $10\,m$ fitted well to the residuals' behavior. One example using the residuals from DA-INV (it. 0) case is shown in Figure A.11.

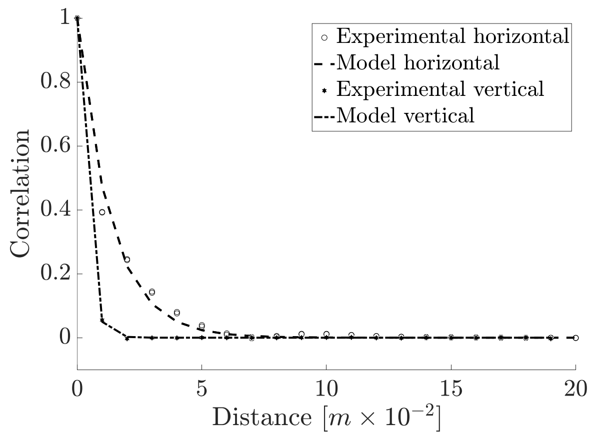


Figure A.11: Example of correlation model match for DA-INV (it. 1).

A.6.5 Complete methodology results

We applied the full methodology, including systematic error reduction and "total" observation error covariance update in the DA-INV-SER case. The "total" observation error covariance matrix update was also applied in the DA-PEM case. Nevertheless, the systematic error reduction was not necessary for this latter because this data is quasi-ideal, and there is no bias to be corrected. The results, in terms of VOIP distributions, are presented in Figure A.12. Note that more plausible volume distributions ensued from the leveling of the high impedance values in DA-INV-SER, even on the initial iteration. Furthermore, the following iterations slightly increased the variability of the posterior ensemble and brought the VOIP expected value closer to the reference. This latter result was also noticed in the DA-PEM and, in different degrees, on all other tests that were performed. This indicates that the iterative update of the "total" observation error covariance matrix contributes to improving the definition of the residual weights, mitigating the effects of the errors in the data assimilation process. It is worth mentioning that the reference value is a possible outcome in all distributions in Figure A.12.

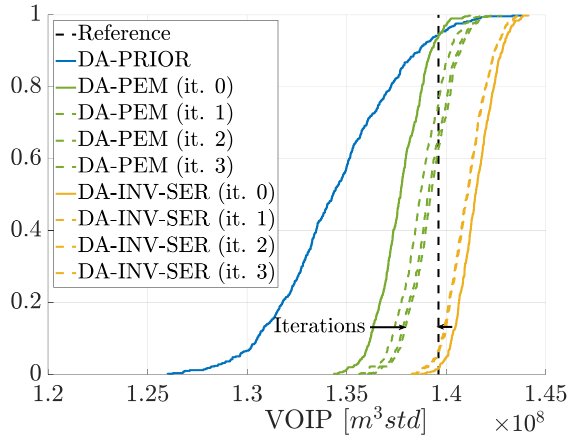


Figure A.12: VOIP cumulative distribution curves of DA-PRIOR (light blue), DA-PEM (light green) and DA-INV-SER (yellow). Three iterations of the "total" observation error covariance matrix estimation are shown in dashed lines for both assimilations.

In Figure A.13, we show the mean porosity maps of the simulation model's 12th layer, which was chosen based on the high contrast of good and bad reservoir regions. The mean porosity maps of DA-PRIOR, DA-PEM, DA-INV, and DA-INV-SER are compared. There are important distortions caused by the seismic forward simulation and the inversion process in our data, which cause the major differences observed between the porosity estimations from DA-PEM and the DA-INV or DA-INV-SER, especially in the northwestern region (red arrows), where lower impedance values occur in $(IP_{obs})_{INV}$ data, due to the interference between layers (Figure A.5).

When DA-INV results are compared to DA-INV-SER in Figure A.13, we observe that the full methodology application reduces the mean parameter changes in comparison with the prior distribution. This is more pronounced in the eastern region (black arrows), where the porosity is kept relatively low in DA-INV-SER, despite the reduction of the impedance values in that region in the INV data (Figure A.5). Another example of a weaker model update occurs in the southwest (blue arrows), where the porosity is maintained in intermediate values in DA-INV-SER, despite the low impedance in the observed INV data (Figure A.5). We refer to weaker model updates in the sense that the values are kept closer to the prior ensemble. Even though it is not shown in Figure A.13, the standard deviation of the porosity in DA-PEM is reduced to about 56% of the standard deviation of the initial ensemble, from 6.1% to 3.4% on average. Moreover, the DA-INV porosity distribution has, on average, 52% of the standard deviation of the DA-PRIOR. This number is slightly increased to 55% in the DA-INV-SER case.

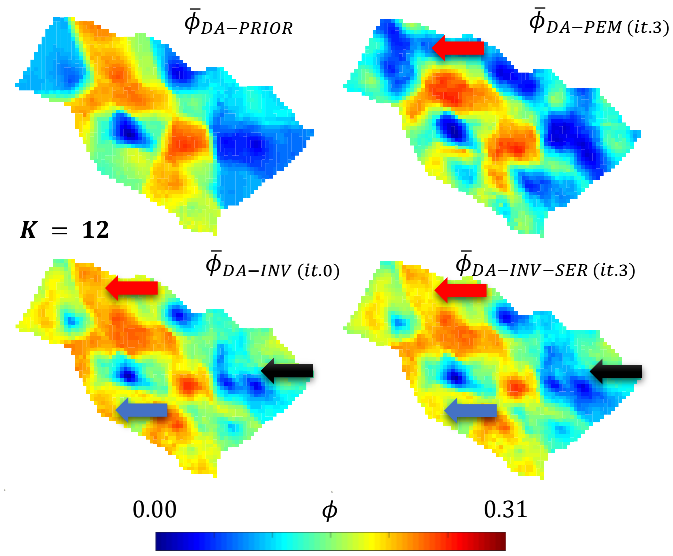


Figure A.13: Mean porosity maps of layer 12.

To check the usefulness of the 3D seismic data assimilation in helping the well data history matching, we analyzed the deviations during the 2618 days history of the DA-PRIOR, DA-PEM (it. 3), and DA-INV-SER (it. 3) cases in Figure A.14. These two last cases were selected because they were the best results to be used as prior ensembles in the following production and time-lapse seismic data assimilation studies. We considered the oil rate, water-cut, gas-oil-ratio, and bottom-hole pressure of the 14 producers and the water injection rate and bottom-hole pressure of the 11 injectors as the available data. The sum of quadratic deviations OF was normalized using the variance of the measurement errors, as reported in (MASCHIO, C. *et al.*, 2013). We can see that the 3D seismic data assimilation alone helped reduce the overall well data OF from around 6.7 to around 3.1 on average in DA-PEM (it. 3). However, the resolution loss limited this reduction to about 4.9 (DA-INV-SER (it. 3)).

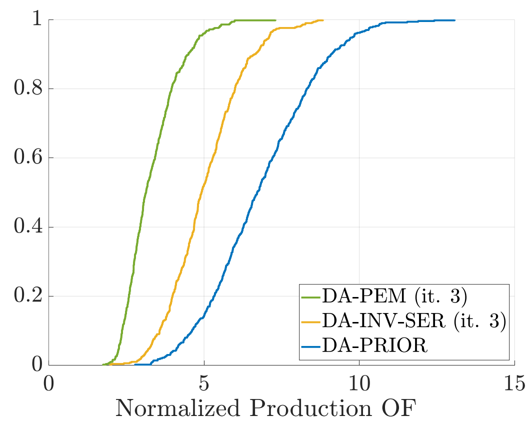


Figure A.14: Well data normalized OF.

A.7 Final remarks

This work proposed a methodology to deal with model and observed data errors when assimilating 3D seismic data. The calibrated porosity models output from this procedure should be used as prior models for a complete data assimilation of well production data and 4D seismic data (if available), which is a topic for future work. The application of the full methodology in the case studied presented promising results. It provided estimations of volumes of oil in-place that were more consistent with the reference (Figure A.12) and it also yielded lower well data OF (Figure A.14).

Nevertheless, there are still some distortions in the final ensemble's models that were caused by the observation errors in the DA-INV-SER (it. 3) case (Figure A.13). This was expected as INV is a very realistic observed data. Despite some simplifications assumed in the modeling process, it does present important features of real data, such as the poor vertical resolution. Indeed, the whole dataset used here presents challenges of real cases: distortions on observed data and lack of information when building prior models. Moreover, an ideal case (free of any model or observation errors) is not comprised in the dataset.

The ideal case considered here (DA-PEM cases) also showed that our methodology positively contributes to the reservoir characterization in the presence of moderately low model errors (Figure A.12).

It is worth highlighting that this work brings an important contribution regarding the definition of the covariance matrix $\boldsymbol{C}_e$, which is a difficult task, especially for seismic data. In Figure A.6, we show that one can try different values when defining this matrix, that the iterative updating of the "total" error covariance matrix will converge to plausible estimations after few iterations.

The promising results of the proposed methodology in the realistic benchmark used here for 3D seismic data assimilation encourage its application on more complex cases involving different data types in future work.

A.8 Conclusions

We addressed the 3D seismic data assimilation process in the presence of model errors caused by the resolution loss. We showed the possible impacts of this kind of error in the reservoir volume characterization. Moreover, we have proposed a methodology to deal with model errors, which considers the non-stationary behavior of seismic data errors. The "total" observation error covariance update is performed here by fitting an analytical correlation model, enabling heterogeneous variance. We also included a systematic error reduction function that is calibrated during the data assimilation process and whose format is defined based on physical-related premises. The key specific conclusions of this work are:

- The iterative update of the “total” error covariance matrix provided a way to improve the weights of the data assimilation method that was practically insensitive to the initial covariance estimate, a parameter that is hard to define in some applications.
- By using the heterogeneous experimental variance and fitting an analytical correlation model, we were able to update the “total” error covariance matrix. This methodology improved the reservoir volume characterization in all of our tests, reducing the impacts of the model errors.
- The inclusion of an uncertain systematic error reduction function was necessary to reduce the bias in the simulated data and adequately calibrate our models’ parameters.
- The systematic error reduction function, which was calibrated during the data assimilation process, tends to weaken the model parameters’ updates, in a sense that they are kept closer to the prior ensemble.
- When the resolution loss was neglected in the 3D seismic data assimilation, we have obtained implausible volume distributions, according to our prior geological knowledge.
- This iterative methodology may be too costly, from a computational perspective, for some practical cases. However, 3D data assimilation is necessary just once in a closed-loop workflow and it has the benefit of providing more geologically consistent prior models to be used in well production (and 4D seismic data) assimilation.
- The computational cost may be reduced by only considering the first couple of iterations, which already improves the final uncertainty representation.
- We obtained stable uncertainty quantification results after three iterations in all of our tests.

In future work, we will extend this methodology to 4D seismic data assimilation and consider other sources of model errors, such as PEM uncertainties.

A.9 Acknowledgments

This work was conducted with the support of Petrobras, Shell, Total, CNOOC, CNPC and PPSA within the ANP R&D tax as "commitment to research and development investments" and Energi Simulation. The authors are grateful for the support of the Center of Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil) and Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil). In addition, special thanks to CMG and Schlumberger for software licenses. The authors would also like to thank Alexandre A. Emerick (Petrobras) for providing the EHM tool to UNISIM and Rafael Medeiros de Souza for generating the synthetic seismic data.

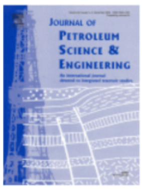
APPENDIX B - LICENSE AGREEMENTS

The following license agreements grant permission to publish the previously published manuscripts that comprise this thesis.

B.1 Assimilating time-lapse seismic data in the presence of significant spatially correlated model errors

CCC RightsLink®

Home | Help | Live Chat | Sign in | Create Account



Assimilating time-lapse seismic data in the presence of significant spatially correlated model errors

Author: Gilson M. Silva Neto,Alessandra Davolio,Denis J. Schiozer

Publication: Journal of Petroleum Science and Engineering

Publisher: Elsevier

Date: Available online 22 June 2021

Journal Author Rights

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW

B.2 Subspace Ensemble Randomized Maximum Likelihood with local analysis for time-lapse seismic data assimilation

Copyright Clearance Center Marketplace™

Society of Petroleum Engineers (SPE) - License Terms and Conditions

This is a License Agreement between Gilson Moura Silva Neto ("You") and Society of Petroleum Engineers (SPE) ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Society of Petroleum Engineers (SPE), and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	07-Apr-2021	**Type of Use**	Republish in a thesis/dissertation
Order license ID	1110263-1	**Publisher**	THE SOCIETY,
ISSN	1086-055X	**Portion**	Chapter/article

LICENSED CONTENT

Publication Title	SPE JOURNAL -RICHARDSON-	**Country**	United States of America
Author/Editor	SOCIETY OF PETROLEUM ENGINEERS (U.S.)	**Rightsholder**	Society of Petroleum Engineers (SPE)
Date	01/01/1996	**Publication Type**	Journal
Language	English		

REQUEST DETAILS

Portion Type	Chapter/article	**Rights Requested**	Main product
Page range(s)	1-21	**Distribution**	Worldwide
Total number of pages	21	**Translation**	Original language of publication
Format (select all that apply)	Print, Electronic	**Copies for the disabled?**	No
Who will republish the content?	Author of requested content	**Minor editing privileges?**	Yes
Duration of Use	Life of current and all future editions	**Incidental promotional use?**	No
Lifetime Unit Quantity	More than 2,000,000	**Currency**	USD

NEW WORK DETAILS

Title	Seismic Data Assimilation in Reservoir Models: Improving Production Forecast in Complex Applications	**Institution name**	University of Campinas
Instructor name	Denis José Schiozer	**Expected presentation date**	2021-06-02

ADDITIONAL DETAILS

Order reference number	N/A

The requesting person / organization to appear on the license	Gilson Moura Silva Neto

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	SPE-205029-PA
Editor of portion(s)	Society of Petroleum Engineers (SPE)
Volume of serial or monograph	N/A
Page or page range of portion	1-21
Title of the article/chapter the portion is from	Subspace Ensemble Randomized Maximum Likelihood with Local Analysis for Time-Lapse-Seismic-Data Assimilation
Author of portion(s)	Gilson Moura Silva Neto ; Ricardo Vasconcellos Soares ; Geir Evensen ; Alessandra Davolio ; Denis José Schiozer
Issue, if republishing an article from a serial	N/A
Publication date of portion	2021-02-10

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL

RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.
8. Miscellaneous.

    8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

    8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

    8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

    8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

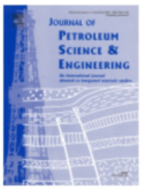
    8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1

B.3 Improving fluid modeling representation for seismic data assimilation in compositional reservoir simulation

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Sign in | Create Account



Improving fluid modeling representation for seismic data assimilation in compositional reservoir simulation

Author: Gilson M. Silva Neto,Victor de Souza Rios,Alessandra Davolio,Denis J. Schiozer

Publication: Journal of Petroleum Science and Engineering

Publisher: Elsevier

Date: November 2020

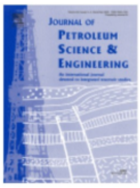
Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW

B.4 3D seismic data assimilation to reduce uncertainties in reservoir simulation considering model errors

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Sign in | Create Account



3D seismic data assimilation to reduce uncertainties in reservoir simulation considering model errors

Author: Gilson M. Silva Neto,Alessandra Davolio,Denis J. Schiozer

Publication: Journal of Petroleum Science and Engineering

Publisher: Elsevier

Date: June 2020

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW