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

GERMANO SCARABELI CUSTÓDIO ASSUNÇÃO

A METHODOLOGY TO COMPARE PROBABILISTIC DATA FROM 4D SEISMIC AND RESERVOIR SIMULATION MODELS

METODOLOGIA PARA COMPARAR DADOS PROBABILÍSTICOS DA SÍSMICA 4D E DE MODELOS DE SIMULAÇÃO DE RESERVATÓRIO

CAMPINAS
2016

GERMANO SCARABELI CUSTÓDIO ASSUNÇÃO

A METHODOLOGY TO COMPARE PROBABILISTIC DATA FROM 4D SEISMIC AND RESERVOIR SIMULATION MODELS

METODOLOGIA PARA COMPARAR DADOS PROBABILÍSTICOS DA SÍSMICA 4D E DE MODELOS DE SIMULAÇÃO DE RESERVATÓRIO

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 Master in Petroleum Sciences and Engineering in the area of Reservoirs and Management.

Dissertação de Mestrado apresentada à Faculdade de Engenharia Mecânica e Instituto de Geociências da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do título de Mestre em Ciências e Engenharia de Petróleo, na área de Reservatórios e Gestão.

Orientador: Prof. Dr. Denis José Schiozer

Coorientadora: Dra. Alessandra Davolio

Este exemplar corresponde à versão final da Dissertação defendida pelo aluno Germano Scarabeli Custódio Assunção e orientada pelo Prof. Dr. Denis José Schiozer

Assinatura do Orientador

CAMPINAS

2016

Agência(s) de fomento e nº(s) de processo(s): CAPES, 33003017

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Luciana Pietrosanto Milla - CRB 8/8129

As79m

Assunção, Germano Scarabeli Custódio, 1990-
A Methodology to compare probabilistic data from 4D seismic and reservoir simulation models / Germano Scarabeli Custódio Assunção. – Campinas, SP : [s.n.], 2016.

Orientador: Denis José Schiozer.
Coorientador: Alessandra Davolio.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Mecânica e Instituto de Geociências.

1. Engenharia de petróleo - Métodos de simulação. 2. Reservatórios (Simulação). 3. Ondas sísmicas. 4. Incerteza. 5. Integração. I. Schiozer, Denis José,1963-. II. Davolio, Alessandra,1980-. III. Universidade Estadual de Campinas. Faculdade de Engenharia Mecânica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Metodologia para comparar dados probabilísticos da sísmica 4D e de modelos de simulação de reservatórios
Palavras-chave em inglês:
Petroleum engineering - Simulation methods
Reservoirs (Simulation)
Seismic waves
Uncertainty
Integration
Área de concentração: Reservatórios e Gestão
Titulação: Mestre em Ciências e Engenharia de Petróleo
Banca examinadora:
Denis José Schiozer [Orientador]
Alexandre Anozé Emerick
Osvair Vidal Trevisan
Data de defesa: 26-07-2016
Programa de Pós-Graduação: Ciências e Engenharia de Petróleo

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

DISSERTAÇÃO DE MESTRADO ACADÊMICO

A METHODOLOGY TO COMPARE PROBABILISTIC DATA FROM 4D SEISMIC AND RESERVOIR SIMULATION MODELS

Autor: Germano Scarabeli Custódio Assunção
Orientador: Prof. Dr. Denis José Schiozer
Coorientadora: Dra. Alessandra Davolio

A Banca Examinadora composta pelos membros abaixo aprovou esta Dissertação:

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

Prof. Dr. Osvair Vidal Trevisan
DE/ DEM / UNICAMP

Dr. Alexandre Anozé Emerick
CENPES / PETROBRAS

Campinas, 26 de julho de 2016.

DEDICATION

To my mother, Maura

& to my love, Júlica.

ACKNOWLEDGMENTS

The completion of this work could not have been possible without the assistance of many people:

I would like to sincerely acknowledge the supervision of Professor Denis J. Schiozer, for his inspiring motivation, immense knowledge, meticulous scrutiny and systematic guidance during these years.

I also wish to thank my co-advisor Alessandra Davolio for giving me the initial idea of this study and reading in detail the entire manuscript. Moreover, thanks for all the discussions and teachings concerning seismic interpretation and reservoir simulation. I could not have imagined having a better co-advisor for this work.

I thank the members that took part in the dissertation committee evaluation, for the time and dedication appreciating this dissertation and for their interest on this work: Osvair Vidal Trevisan (FEM/UNICAMP), Alexandre A. Emerick (CENPES/PETROBRAS) and Célio Maschio (CEPETRO/UNICAMP).

I thank the Brazilian funding agency CAPES and the company BG Brasil for the master scholarships that allowed me to undertaking this research and to attend conferences. I also thank UNISIM and CEPETRO for supporting this work and CMG and Schlumberger for software licenses.

Likewise, I would like to thank the members of staff at Energy Department and UNISIM for all their support, mainly in technical and administrative issues. A special thank you to all my friends from the master courses, which have helped me through this journey, in many different ways: Bordeaux, Caaso, Cathy, Darlan, Felipe, Forlan, Juliana, Luz, Pazetti, Ricardo Moreno, Ricardo Vasconcellos, Robson e Santiago.

A special note of gratitude to my old and dear friends: Aneci, Brunot, Carol, Denis, Edmur, Gabrielzim, Graciete, Jaque, Jiló, Joe, Júlia, Júnio, Magnão, Nayara, Rafinha, Renata, Renatíssima, Rick, Thamires, Wallace, Weslen e Zizi. Love you all.

I thank my beloved mother, Dona Maura, for all her immense love and care, despite the geographic distance. You will be always my inspiration! Moreover, I must express my very profound gratitude to my sisters, Feliciana and Luciana, for providing me continuous encouragement throughout my years of study. Thanks to them, for being always there for me in highs and lows in my life.

Finally, thank you my darling, Júlica, for her continued support and encouragement. I'm so lucky to have you in my life!

ABSTRACT

The focus of this work was to develop a methodology to compare probabilistic data from two technologies, numeric reservoir simulation and 4D seismic. The first part of this study presents the proposed methodology that can be seen as a diagnostic tool to compare, guide and better understand simulation and seismic information. In this work the comparison between the two data sets was done in the pressure and saturation domain. So, multiple dynamic change maps (water saturation and pressure change maps) yielded from multiple simulation models and from probabilistic synthetic seismic inversion were the input of the methodology, which generates a "diagnostic map" that allows quantifying the agreement between simulation and seismic data. The information acquired from this "diagnostic map" was then applied to select the best simulation models using 4D seismic data, in regions where seismic was more precise than simulation. It was shown that the selected dynamic change maps from simulation models better matched the expected answer than the initial ones. In the second part of this study, a comparison between the proposed methodology and a traditional methodology was performed. It was considered that the traditional methodology uses probabilistic information from simulation and deterministic from seismic. Thus, the additional information that a probabilistic seismic integration in reservoir modelling workflow could bring was evaluated. Although satisfactory results were observed in both methodologies, proposed procedure showed to be more robust than the traditional method. The third part was performed to quantify which method was the most accurate to calculate the probability distribution of a certain data set. The estimation of probability distributions from seismic and simulation data sets was required in step one of the proposed methodology. Thus, the goal of the third part was to guarantee that the methodology developed in this work had used the most appropriate statistical tool. It was found that the kernel density estimator was the most accurate among the three methods studied. The fourth and last part of this dissertation presents a complementary study of the methodology developed in the first part. The aim of this last part was to evaluate the robustness of the proposed methodology when different qualities of simulation and seismic data were available, in other words, the applicability of the methodology in different (synthetic) cases. Despite parts two and three of the present work had shown significant results, it is important to highlight that the main contributions of this work is the "diagnostic map", which integrate

probabilistic data from simulation models and from 4D seismic from an innovative perspective.
Keywords: history matching; 4D seismic; reservoir simulation; probabilistic integration; reservoir monitoring.

RESUMO

O foco do presente trabalho foi desenvolver uma metodologia para comparar dados probabilísticos de duas tecnologias, simulação numérica de reservatórios e sísmica 4D. A primeira parte deste estudo apresenta a metodologia proposta, que pode ser vista com uma ferramenta de diagnóstico para comparar, guiar e melhor compreender as informações da simulação e da sísmica. Neste trabalho a comparação entre os dois conjunto de dados foi feita nos domínios da pressão e da saturação. Assim, múltiplos mapas de mudanças dinâmicas (mapas de mudança de pressão e saturação de água) obtidos de múltiplos modelos de simulação e de uma inversão sísmica sintética probabilística foram os dados de entrada da metodologia, que gera um "mapa diagnóstico" que permite quantificar a concordância entre dados de simulação e sísmica. A informação obtida deste "mapa diagnóstico" foi então aplicada para selecionar os melhores modelos de simulação usando os dados da sísmica 4D, em regiões onde a sísmica era mais precisa que a simulação. Foi mostrado que os mapas selecionados dos modelos de simulação honraram melhor a resposta esperada que os mapas iniciais. Na segunda parte deste estudo, uma comparação entre a metodologia proposta e uma metodologia tradicional foi realizada. Considerou-se que a metodologia tradicional usa dados probabilísticos da simulação e determinísticos da sísmica. Assim, foi avaliado a informação adicional que a integração sísmica probabilística no fluxograma de modelagem de reservatório poderia trazer. Embora resultados satisfatórios tenham sido observados em ambas as metodologias, o procedimento proposto mostrou-se mais robusto que o método tradicional. A terceira parte deste trabalho foi realizada para identificar o método mais acurado para calcular a distribuição de probabilidade de um conjunto de dados. A estimativa das distribuições de probabilidade da sísmica e da simulação é requerida no passo um da metodologia proposta. Assim, o objetivo deste estudo foi garantir que a metodologia desenvolvida neste trabalho usasse a ferramenta estatística mais adequada. O estimador kernel de densidade de probabilidade foi o método mais acurado entre os três estudados. A quarta e última parte desta dissertação apresenta um estudo complementar da metodologia desenvolvida na parte um. O objetivo desta última parte foi avaliar a robustez da metodologia proposta quando diferentes qualidades de dados da simulação e da sísmica estão disponíveis, ou seja, a aplicabilidade da metodologia em diferentes casos (sintéticos). Apesar das partes dois e três do presente trabalho terem mostrado resultados significantes, é importante destacar que a principal contribuição do presente trabalho foi o "mapa diagnóstico”, que integra com um perspectiva inovadora dados probabilísticos de modelos de simulação da sísmica 4D.

Palavras Chave: Ajuste de histórico; sísmica 4D; simulação de reservatório; integração probabilística; monitoramento de reservatório.

TABLE OF CONTENTS

1. INTRODUCTION

One of the main objectives of the reservoir engineer is to predict the overall behavior of the reservoir. In this context, simulation models have an important role, allowing the professional to integrate several measured properties from different areas and numerically simulate the reservoir behavior to evaluate field performance: predicting reserves, forecasting reservoir production, performing reservoir characterization studies, analyzing risks under different production strategies, managing the reservoir.

However, the simulation models are simplifications of the real reservoirs and contain limitations such as numerical errors and uncertainties in the properties that describe the reservoir. The first limitation is inherent to any computational analyses and strongly depends of improvements in numerical methods to calculate complex equations, scale transfer from a certain data etc. The second one, uncertainties in the reservoir properties, occurs mainly because most of reservoir parameters are obtained indirectly, through correlation and interpretation (Maschio and Schiozer, 2013).

In order to minimize these limitations, a history matching procedure is applied. Basically, the idea of this procedure is to change uncertain reservoir properties of the models until the simulation results match the measured (observed) dynamic data, such as pressure change and fluids rates. Therefore, history matching is an inverse problem, where the expected answer is known (production rates and pressure), and reservoir properties to reach this answer are unknown. This procedure has been developed since the late 60’s, but the last decade has seen notable progress with several published works (Sarma *et al.*, 2005; Gervais and Roggero, 2010; Olivier and Chen, 2011; Emerick and Reynods, 2013; Maschio and Schiozer, 2016).

In general, the observed dynamic data used is the production, injection and pressure measured at the wells (Ida, 2009). Nonetheless, the wells are located in sparse regions, which call for addition areal dynamic data, such as 4D seismic.

4D seismic is the repetition of two (or more) 3D, or 2D seismic surveys acquired at different times over the same area. It gives areal dynamic reservoir information, that which can be used as additional data in history matching. Moreover, it can provide the fluid contact with time, estimate the fault seal and locate the bypass fluids (Yan, 2014).

The incorporation of seismic information to update reservoir models can be performed in three domains: (1) seismic amplitude, *i.e.*, seismic original domain, (2) seismic impedances and (3) saturation and pressure domain. As presented in several works in literature (Risso and Schiozer, 2008; Fahimuddin, 2010; Roggero *et al.*, 2012; Riazi *et al.*, 2013), there is not standard domain, since the choice of the domain used depends of the study performed and the available information. Figure 1 presents the seismic domains and the possible matches with simulation data. The present work used seismic information in saturation and pressure domain. The reason for this choice is described in Section 1.3. More details about the seismic domains can be found in the work of Sagitov and Stephen (2012).

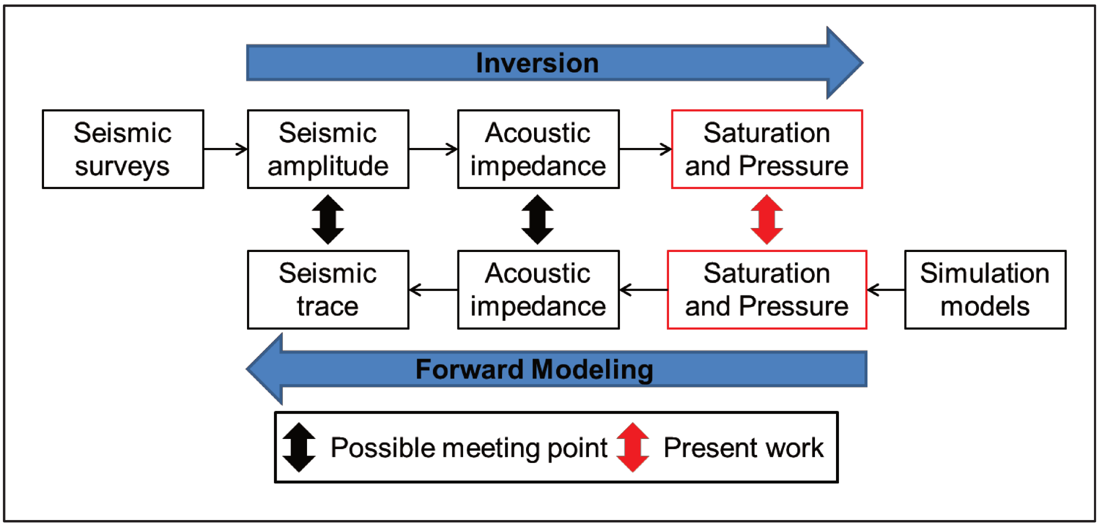

Figure 1–Possible domains to compare 4D seismic and simulation data (Modified from Landa and Kumar, 2011).

Other issue to be considered is how to apply 4D seismic data in reservoir characterization: qualitatively or quantitatively. Traditionally, 4D seismic information is used qualitatively and used to visualize reservoir conditions under production: detecting the reservoir fluid changes during production, identifying undrained oil areas, possible faults etc. Then, the simulation models are updated in order to reproduce the behavior qualitatively observed in 4D seismic data. The works developed by Behrens *et al.* (2001), Shi *et al.* (2006), de Brito *et al.* (2010) and Sagitov and Stephen (2012) are some examples of this type of application.

Quantitative use of seismic emerged from the demand to use more effectively and directly the 4D seismic data within assisted (or automatic) history matching procedures. In this situation, seismic information is usually incorporated inside objective functions (that measure the quality of the matching), which must be minimized. One of the most significant studies in this context is the one presented by Gosselin *et al.* (2003), who used a gradient-based methodology to optimize an objective function with 4D seismic information integrated. This method was named History Matching Using Time Lapse Seismic (HUTS) and showed significant improvement in reservoir characterization and reduction in the range of uncertain properties. Later, Portella and Emerick (2005) showed that HUTS works fine whether seismic data of good quality is available. In the same year, Stephen *et al.* (2005) presented a history matching method that used 4D seismic and production data simultaneously, on an integrated workflow, with some improvements in relation to HUTS, such as in the gradient-based method used.

More recently, Almeida *et al.* (2014) used the indicator called Normalized Quadratic Deviation with Sign (NQDS) to incorporate a water saturation change map from 4D seismic in a history matching procedure.

Beyond the type of the seismic data set used (amplitude, impedance or saturation and pressure domain) and how to use it (qualitatively or quantitatively) there is still another challenge, involving the quality of 4D seismic data. As emphasized Castro (2007) and Bosch *et al.* (2010), seismic signals and the seismic data processing (such as inversions) have several uncertainties, which make seismic responses non-unique. Besides that 4D interpretation requires the knowledge of rock and fluid properties which can be very uncertain too.

Traditionally, seismic and simulation data are used deterministically, as illustrates Figure 2a. In this case, the dynamic change map (saturation and pressure change or impedance variation, etc.) obtained from a unique reservoir simulation model is compared (qualitatively or quantitatively) with the dynamic change map from a deterministic 4D seismic. In these cases, the uncertainties from simulation and seismic data are not quantified.

However, reservoir simulation and 4D seismic data contain several uncertainties and limitations, since both technologies deal with several unknown properties and with solution of complex inverse problems, such as history matching (reservoir simulation data) and seismic inversion (4D seismic data).

In order to mitigate inaccurate conclusion that can be obtained from studies with deterministic simulation model, the number of works that uses multiple models (probabilistic) instead of a single simulation model has increased. For problems with a large number of parameters taken probabilistically, there are some methods found in the literature: (1) the ensemble smoother (ES) method proposed by Leeuwen and Evensen (1996) and applied in reservoir history matching in works (Skjervheim *et al.*, 2011; Emerick, 2016); (2) genetic algorithms, used to reservoir calibration, as shown in Romero *et al.* (2000) and Xavier *et al.* (2013); (3) neighborhood algorithms, applied for history-matching problems (Suzuki, 2007; Jin *et al.*, 2012) and others.

Nonetheless, even when probabilistic simulation data are considered, if 4D seismic information is used, it is usually considered as a deterministic ("exact") data that the multiple simulation models should match, as illustrates Figure 2b. Although it would be more reasonable contemplate the uncertainties from both technologies (since both have several limitations), there are still few works that used probabilistic seismic information from seismic and simulation in the same time (Figure 2.c).

We can find some studies regarding seismic uncertainties; for instance, Grana and Mukerji (2014) proposed a Bayesian inversion to carry out 4D seismic data. The probabilistic estimates from this inversion could then be integrated in a history matching procedure. However, it is notorious the lack of works that proposed methodologies to integrate this probabilistic 4D seismic data with probabilistic reservoir simulation information.

Emerick (2016) incorporated 4D seismic impedances using an approximated data-error covariance. His work showed that the incorporation of 3D and 4D seismic provided some improvements in the data matches and great reductions in the variability of predicted water rate and in the permeability distribution of the field.

Landa and Kumar (2011) presented a methodology where the reservoir models are calibrated using production and 4D seismic data simultaneously, through the same workflow. This procedure was performed in a probabilistic scenario and 4D seismic data were used in the amplitude domain. A probabilistic seismic modeling was required inside a history matching procedure, which allows accessing the pressure and saturation (or impedances) scenarios that provided the best final solution. This could be seen as a procedure that integrates probabilistic data from production and from 4D seismic. However this can be a very complex task to perform and good parameterization is necessary to guarantee the success of the implementation. Moreover, this probabilistic joint procedure (history matching and seismic inversion) can generate statistical bias that would not be observed in seismic inversions without production data integrated (as presented Grana and Mukerji, 2014). Hence, to avoid these drawbacks, it is interesting to develop a method that integrates probabilistic data from production and seismic but without perform this kind of joint process.

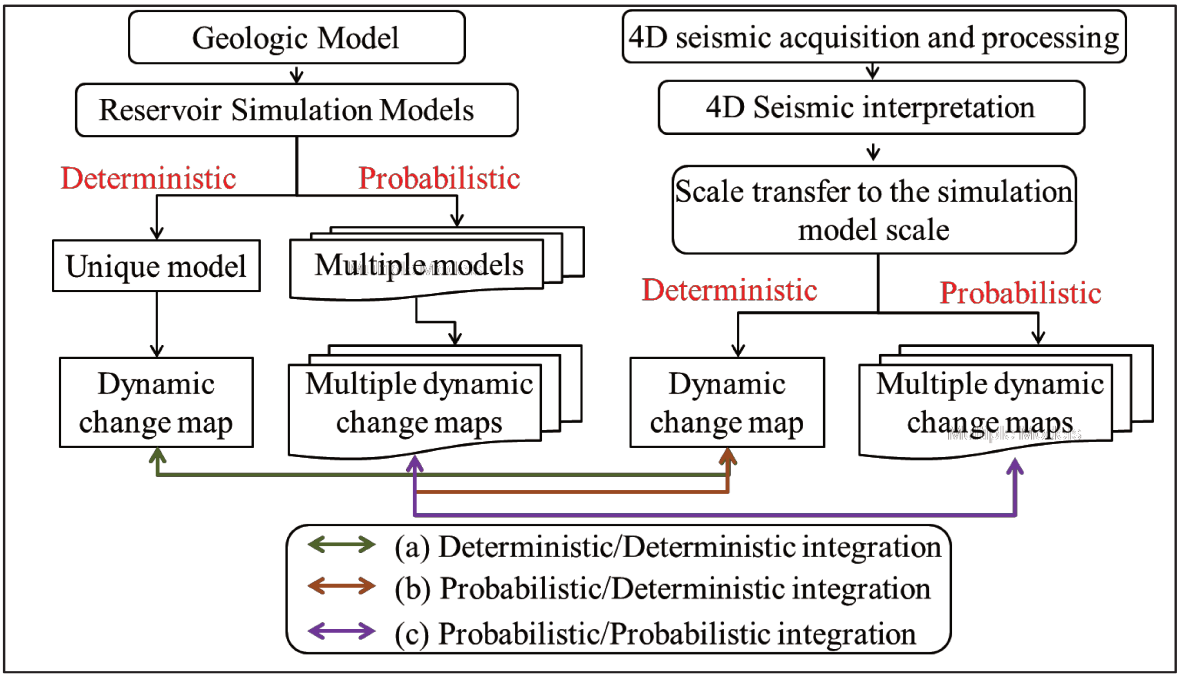

Figure 2–Possible ways to compare 4D seismic and reservoir simulation data: (a) considering both deterministically, (b) probabilistic simulation data with deterministic observed 4D seismic data and (c) both probabilistically with uncertainties.

1.1. Motivation

Although 4D seismic data can be used to improve reservoir characterization, their uncertainties are usually not considered in seismic history-matching procedures. In the present scenario, the lack of methodologies that uses 4D seismic probabilistically is remarkable. Moreover, recently published studies (Davolio *et al.*, 2013a and Tian *et al.*, 2014) propose the use of available simulation data (engineering data) to constraint and acquire more consistent information from seismic inversions. However, simulation models are also not completely reliable.

So, it becomes important to develop a way to perform an integration considering the uncertainties from both technologies (simulation and seismic) and to identify which information is most reliable in each reservoir location.

1.2. Objectives

The present work introduces a methodology to compare and integrate probabilistic information from two technologies: 4D seismic and reservoir simulation. The methodology generates a diagnostic tool, that can be used to guide the probabilistic integration between seismic and simulation and to select the most representative information from both.

1.3. Premises

- The study was developed using synthetic data. There is a reference model that represents the true earth model, in other words, the answers to be reached. Thus, this reference model is used to verify the accuracy of the proposed methodology. The reference model is also used to obtain the data usually measured in a field (well log, production data etc.). From these measured data the simulation models and history data are generated.
- Seismic data (P and S impedances) were generated from a petro-elastic model that used input data from the reference model. These impedances were probabilistically inverted to obtain pressure and saturation estimates. No seismic amplitudes were used.
- Pressure and saturation domain is used to compare simulation and seismic data. Pressure and saturation maps are more complex to be obtained from seismic data; however, they are direct responses from the simulation models. Another reason for working on saturation and pressure domains, as explained in Davolio (2013), is because the values of these physical quantities can be better controlled (to establish feasible limits) than the elastic properties of rock (such as impedance).
- All data are in the same scale 110x90 blocks with 9 layers.
- There is no presence of gas (the reservoir pressure is kept above bubble pressure through water injection).

1.4. Description of the work

This dissertation is structured in three papers. The first paper (Chapter 2) describes the methodology proposed to compare the probabilistic information from seismic and simulation, which is the main contribution of the present work. The second paper (Chapter 3) proposes a comparative study between the proposed methodology and one methodology performed with deterministic seismic information (traditionally used). The third paper (Appendix A), presents the statistical study performed to define the most accurate method to estimate the probability distribution of a certain data set. The information from this last paper is used in the first paper (Chapter 2).

The dissertation also comprises Appendix B, which presents a complementary analysis to Paper 1 (Chapter 2) discussing the application of the methodology proposed with different datasets. Finally, Chapter 4 presents the most relevant conclusions and future steps.

A summary of the three papers and the relation between Appendix B and Paper 1 are highlighted in the next section.

1.4.1. PAPER 1: "A Methodology to Integrate Multiple Simulation Models and 4D Seismic Data Considering Their Uncertainties"

Germano S. C. Assunção, Alessandra Davolio, Denis José Schiozer.
This paper was prepared for oral presentation at the SPE Annual Technical Conference and Exhibition held in Dubai, UAE, 26–28 September 2016.

The methodology proposed in this paper is the main contribution of this dissertation, presenting a new mechanism to evaluate the information from 4D seismic and simulation data considering their uncertainties.

As main results, we identified regions in the reservoir where 4D seismic data could bring more information than simulation and vice-versa. Moreover, it is possible to find reservoir locations where both data are providing divergent information (which is a indicative of presence of "unknown unknowns") or convergent information (seismic and simulation are well matched).

Appendix B presents some complementary results to Paper 1. The same methodology from Paper 1 is discussed in Appendix B, however different sets of data from simulation and seismic are tested, in order to evaluate the applicability of the proposed methodology.

1.4.2. PAPER 2: "A Comparative Study of two Methodologies to Integrate Reservoir Simulation and 4D Seismic Data"

Germano S. C. Assunção, Alessandra Davolio, Denis José Schiozer.
This Technical Paper was prepared for presentation at the Rio Oil & Gas Expo and Conference 2016, held between October, 24-27, 2016, in Rio de Janeiro.

In this work, the methodology proposed in Chapter 2 is compared with a methodology traditionally used (which integrates 4D seismic deterministically). The methodologies are performed to select the most representative pressure and saturation changes maps from simulation models using 4D seismic data, mimicking an iteration of a seismic history matching procedure where only 4D seismic is used in the objective function. The models selected by the application of each methodology are compared and the differences between them are analyzed.

The contribution of this study to the dissertation is to show the most relevant differences brought between determinist and probabilistic use of seismic data to support reservoir characterization.

1.4.3. PAPER 3: "Quantitative Comparison of Non-Parametric Methods to Handle Probabilistic Data From 4D Seismic and Reservoir Simulation"

Germano S. C. Assunção, Alessandra Davolio, Denis José Schiozer.
To be submitted to a journal.

A critical part in probabilistic analyses is the choice of the most accurate statistical method to estimate the probability distribution of the data sets used. There are several methods to estimate the probability distribution of a certain data set, but they present some drawbacks. Thus, this work aims to evaluate three of the most used methods: histogram, empirical cumulative frequency curve and kernel density estimator.

After comparing the three methods we concluded that the kernel density estimator is the more accurate. The results from the present paper contribute to the development of methodology proposed in Chapter 2, since the estimation of the probability distribution is the first step of the proposed methodology.

2. PAPER 1: "A Methodology to Integrate Multiple Simulation Models and 4D Seismic Data Considering Their Uncertainties"

Germano S. C. Assunção, Alessandra Davolio, Denis José Schiozer.

This paper was prepared for oral presentation at the SPE Annual Technical Conference and Exhibition held in Dubai, UAE, 26–28 September 2016.

SPE-181602-MS

A Methodology to Integrate Multiple Simulation Models and 4D Seismic Data Considering Their Uncertainties

Germano S. C. Assunção, Alessandra Davolio, Denis J. Schiozer, State University of Campinas (UNICAMP)

This paper was prepared for presentation at the SPE Annual Technical Conference and Exhibition held in Dubai, UAE, 26–28 September 2016.

Abstract

Traditionally, integration between 4D seismic (4DS) and simulation data has been performed considering the 4DS data deterministically. However, there are uncertainties in the response of seismic. The goal of the methodology presented in this work is to compare the changes of dynamic properties estimated from 4DS and simulation models considering the uncertainties inherent to both data.

The relevant reservoir uncertainties can be combined to generate multiple simulation models, which provide maps of dynamic changes, such as pressure change (Δp) and water saturation variation (ΔS_w). Available 4DS can also be used to map dynamic changes. Through a stochastic seismic inversion, multiple ΔS_w and Δp maps can be obtained from 4DS. After selecting a proper scale (scale transference), we compare the dynamic maps from seismic and simulation data using probabilistic density functions (PDFs), establishing levels of agreement/disagreement between 4DS and simulation data.

To validate the methodology we use a synthetic dataset, with moderate complexity and seven uncertainties mapped, such as fault transmissibility, porosity, facies, and permeability. 500 maps of ΔS_w and Δp from 4D seismic were generated from prior probabilistic seismic inversion. 500 simulation models previously calibrated using well production data generated the set of 500 maps of ΔS_w and Δp from simulation. Applying the methodology, we identify four regions: (1) reservoir locations where both estimates (seismic and simulation) are similar, showing regions properly calibrated, (2) locations where simulation estimates are more precise than 4D seismic, (3) reservoir locations where the data sets indicate divergent estimates, and (4) 4DS estimates are more precise than simulation.

This information can be very useful to guide data integration. As an example, we show that region (4) can be used to select the simulation models that reproduce ΔS_w or Δp behavior from 4DS, since 4D seismic data is more precise than the simulation estimates in this region. Other useful information from the proposed methodology is that the reservoir zones identified as region (2) can be used as a constraint to reinterpret 4D seismic data, as simulation estimates

are more precise.

The methodology is a new way to evaluate the information from 4D seismic and simulation data considering uncertainties. The identification of these four regions can be useful in the parametrization phase of the history matching procedure (a complex process), as an additional tool to understand the properties in this procedure. The methodology also indicates possible locations to use reservoir engineering constraints to improve seismic interpretation, in regions where estimates from simulation are more precise than 4D seismic data. Moreover, we can use the methodology to determine critical reservoir locations to be reevaluated, those presenting disagreement between the two data source.

Introduction

Reservoir modeling is a complex task, involving a large number of uncertain properties that characterize the reservoir, such as water-oil contact, fault transmissibility, permeability and porosity distributions. A common practice to reduce the uncertainty of these properties is to evaluate the difference between dynamic observed data and simulated data. The properties in the simulation models are then updated until this difference reaches an acceptable value. This procedure, history matching, plays a key role in reducing the uncertainties, providing more reliable simulation models to manage the real reservoir, explore different locations for infill wells, study potential benefits of smart wells, optimize well distribution, and more (Oliver and Chen, 2011).

According to Morell (2010), dynamic observed data is usually divided into two categories: production and seismic data. Production data, such as oil flow and bottom hole pressure, are obtained from well measurement but as the wells are in specific reservoir locations, they lack areal information. Seismic data, on the other hand, is spatially valuable, providing dynamic information in inter-well regions. 4DS is the difference observed over two or more 3D (or 2D) seismic data sets of the same area acquired at different times. In history matching, 4DS can be used as dynamic information to reveal unknown characteristics of the reservoir such as fluid displacement and pressure changes.

In addition, the history matching can be performed deterministically or probabilistically. Using the deterministic approach a single simulation model attempt to reproduces the real reservoir. In this case, the properties are updated using available dynamic observed data and then the calibrated model guides the management of the real reservoir. This approach is statistically weak because it ignores the uncertainties of the properties: production forecast is performed using a single model. Considering these uncertainties is important because they can range widely. Thus, over the last years, research has focused on developing probabilistic history matching procedures, to handle several models simultaneously, for a more reliable analysis, as shown in Evensen *et al.* (2007), Elsheikh *et al.* (2012), and Maschio and Schiozer (2014).

In this context, 4D seismic information is usually used as deterministic dynamic observed data to improve the reservoir models (probabilistic history matching) or model (deterministic history matching). Thus, we can compare the simulation result from a single model with the 4D seismic data (e.g. study proposed by Almeida *et al.*, 2014) or compare the results from multiple simulation models with the available 4D seismic data (e.g. works of Stephen *et al.*, 2004 and Riazi *et al.*, 2013). These single or multiple simulation models may have been previously calibrated with a history matching procedure using production data from wells or not, depending on the study.

Nonetheless, seismic data also have some uncertainties due to noise and possible errors caused by resolution, acquisition and processing the data. Deterministic 4D seismic information does not consider these uncertainties: the changes interpreted from 4D seismic data are assumed "exact", as Castro (2007) explained. Thus, following the current trend for using probabilistic history matching procedures to handle reservoir simulation models, using 4D seismic information probabilistically would improve the accuracy of the 4D seismic data.

Some works integrate 4D seismic data and simulation data following a probabilistic approach, such as the joint history matching/inversion. The basic idea of this procedure is to update well production data and 4D seismic data simultaneously. An objective function is defined to measure the mismatch between the measured dynamic data (production and 4D seismic-related data) and the corresponding simulated responses. Then, an optimization algorithm is run to minimize this function and update seismic and simulation data (Gervais-Couplet *et al.,* 2010 and Landa and Kumar, 2011).

However, the joint history matching/inversion are a complex non-linear inverse problem, sometimes involving expensive computational simulations. One possible drawback of this approach is the generation of inconsistent geological behavior when minimizing the objective function. The success of this type of method (and others related to inverse problems) depends on defining the uncertain properties to be modified and how to change them, to achieve an acceptable match through an optimization procedure.

The objective of this work is to compare probabilistic data from 4D seismic and reservoir simulation, but without an integrated optimization process. The main idea is to evaluate the agreement/disagreement of the two data sets (seismic and simulation) through the modified overlapping coefficient (OVL). We propose the use of the OVL as a tool to compare seismic and simulation data considering the uncertainties for both. This methodology is a diagnostic tool, which can be used to understand the uncertainties from both data sets, to guide the seismic and simulation integration in any process (for instance in joint history matching/inversion) and/or to select the most representative data sets concerning both sides: seismic and simulation. Previous work performed by Davolio and Schiozer (2015) proposed a methodology to compare maps of dynamic changes from 4D seismic and simulation data, this work presents a more robust way to compare them.

Methodology

This methodology integrates reservoir simulation data and 4D seismic data, quantitatively and probabilistically. The workflow in **Fig. 1** shows the domain of integration between 4D seismic and simulation data for the proposed methodology.

Considering *n* maps of Δp estimated from a probabilistic 4D seismic inversion transferred to simulation scale and *m* maps of Δp yielded from multiple simulation models. Two pressure change vectors are built for every reservoir grid block: $SEIS_p = [\Delta p_1^{SEIS}, \Delta p_2^{SEIS} ... \Delta p_n^{SEIS}]$ and $SIM_p = [\Delta p_1^{SIM}, \Delta p_2^{SIM} ... \Delta p_m^{SIM}]$.

The same procedure can be performed to ΔS_w maps, using the following vectors: $SEIS_{Sw} = [\Delta S_{w1}^{SEIS}, \Delta S_{w2}^{SEIS} ... \Delta S_{wn}^{SEIS}]$ and $SIM_{Sw} = [\Delta S_{w1}^{SEIS}, \Delta S_{w2}^{SEIS} ... \Delta S_{wm}^{SEIS}]$.

The overlapping coefficient (OVLC) refers to the area under two probability density functions. Thus, OVLC determines the relative closeness of the two datasets, where 0% represents complete disagreement and 100%, represents data agreement.

Using the kernel density estimator (KDE) proposed by Botev *et al.* (2010), we can generate the probability density functions (PDFs) from both data.

PDF_{SEIS} and PDF_{SIM} denote the probability distribution functions of vectors $SEIS_p$ and SIM_p, respectively. The overlapping coefficient is defined according to Weitzman (1970) apud Schmid and Schmidt (2006) by **Eq. 1** and it can be used to identify the coincidence interval of PDF_{SEIS} and PDF_{SIM}:

$$\text{OVLC} = \int_0^1 \min\{PDF_{SEIS}; PDF_{SIM}\}dp, \quad \ldots\ldots(1)$$

The OVLC method is usually used to compare two probability density functions, with applications in other fields. Further explanation about the OVLC can be found in Bradley (2006), and Al-Saleh and Samawi (2007).

In this work we propose the modified overlapping coefficient (OVL). In this approach, two parameters are defined: (1) OVL_{SIM} is the area of PDF_{SIM} within the OVLC interval and (2) OVL_{SEIS} is the area of PDF_{SEIS} with the OVLC interval. These parameters are shown in **Eqs. 2** and **3:**

$$OVL_{SIM} = \frac{Area\ under\ PDF_{SIM}\ within\ OVLC\ interval}{Overall\ area\ under\ PDF_{SIM}}, \quad \ldots\ldots(2)$$

$$OVL_{SEIS} = \frac{Area\ under\ PDF_{SEIS}\ within\ OVLC\ interval}{Overall\ area\ under\ PDF_{SEIS}}, \quad \ldots\ldots(3)$$

Fig. 2 illustrates the parameters presented above. The green dashed lines represent the overlapping interval (OVLC interval). We first identify the interval **(Fig. 2a)** and then compute the OVL_{SIM} and OVL_{SEIS} through **Eqs. 2** and **3**, as presented in **Figs. 2b** and **2c**, respectively.

The procedure is performed for every grid block. The OVL information is then gathered in two maps, the OVL_{SEIS} and OVL_{SIM} maps, to observe the overall behavior of the estimates from seismic and simulation data. We can identify the four different arrangements through cross plotting the OVL_{SIM} and OVL_{SEIS} maps, as **Fig. 3** illustrates.

Based on the cross plot, we defined four different regions indicating: (1) locations with agreement between simulation and seismic data, i.e., regions properly calibrated, (2) locations where simulation data is more precise than seismic, thus, simulation can be used to improve 4D seismic interpretation[1], (3) areas where simulation and seismic data disagree, indicating regions where the uncertainties related to both data should be re-evaluated, and (4) reservoir locations where seismic data is more precise than simulation, so 4D seismic data can be used to calibrate simulation models following traditional history matching practices.

In the first phase of integration between seismic and simulation data, we considered 80% agreement between OVL_{SEIS} and OVL_{SIM} as an acceptable value. However this tolerance is user defined and as the responses from 4D seismic and simulation data become more precise, this parameter can be adjusted to a greater value. Note that precision in data analysis is associated with the amount of variation or dispersion of the data set. A precise data set indicates that the data points tend to the mean, while imprecision is indicated by data points spread out over a wider range of values, precision is not the same as accuracy.

[1] In this case, it is highly recommended to run a previous history matching using well production data to have (more) reliable models to evaluate the 4D seismic interpretation.

Fig. 4 presents an overall workflow of the proposed methodology, divided into three steps: (1) generate PDFs, (2) calculate OVLs in every grid and (3) cross plot OVL_{SIM} and OVL_{SEIS} and identify the 4 regions. Thus, identifying the 4 regions is the main contribution of our methodology.

Instead of Δp or ΔS_w maps, the proposed methodology can utilize any other 4D seismic attribute, such as impedances. The choice of which attribute to use depends on the data available.

Application

Reservoir description

This study used a synthetic dataset to generate a 3D clastic model (Beta model), on which our approach is tested. We used a reference model representing the true earth model and corresponding to a system with two facies: sandstone and shaly-sandstone, with high horizontal continuity and different porosity, permeability and net-to-gross (NTG) distributions. The structural framework of the reservoir is represented by an anticline comprising 4 major faults with different transmissibility **(Fig. 5a)**. The reference model contains about 1,600,000 cells (270x330x18). Nineteen wells are active during the flow simulation (Black-oil): eleven vertical producers and five vertical injectors **(Fig. 5b)**.

Model uncertainties

Uncertainty in reservoir data is mainly caused by uncertainties in measurements, data handling coming in second. The two goals of uncertainty analysis are to quantify and reduce uncertain reservoir properties, aiming to generate more accurate models. There are seven uncertain attributes considered in this work: (1) relative permeability for the two facies, (2) ratio between permeability vertical and horizontal (Kz/Kx), (3) transmissibility of the four faults, (4) facies distribution, (5) porosity, (6) absolute permeability and (7) NTG distributions. Correia *et al.* (2016) presents further details about these uncertainties.

Simulation models

The history data was obtained from the reference model considering a five years flow simulation. Combining all the uncertain properties previously mentioned and using a coarser grid block (90x110x9), 500 simulation models (m= 500) was created through the DLHG technique (Schiozer *et al.,* 2014). After the generation of the models, a well history matching was performed and the data here comprises 500 simulation models yielded from an intermediate step (step 2 of 4). The details of the well history matching are presented by Almeida *et al.* (2014).

4D modeling

The reference model provides the 4D seismic data used in this study. The first step of the seismic modeling consists of a flow simulation to predict fluid saturations and pressure in the reservoir at the time of the seismic surveys. There are a base survey before production start and a monitor after 5 years of production. The reservoir properties are converted to seismic attributes, such as P- and S-wave velocities and density (elastic domain) using the petro elastic model presented in Pazetti *et al.* (2015). In this work the "observed" seismic data are the P and S impedances computed from the forward modeling, no seismic amplitudes are generated. To produce a more realistic dataset a random noise was added to seismic impedances as described in Davolio and Schiozer (2014). These disturbed impedances are used as input to a probabilistic inversion procedure, based on Latin Hypercube, estimating pressure and saturation changes described in Davolio and Schiozer (2015), consequently, multiple scenarios (*n=500*) of water saturation and pressure distribution are computed from 4D seismic impedances.

ΔS_w and Δp maps

Fig. 6 illustrates the estimates of ΔS_w used in this work. All the maps are obtained from layer 3 (from 9 layers). In **Fig. 6a**, the reference map presents the expected value, that is, the true answer. **Fig. 6b** shows the mean of the saturation change maps from the 500 simulation models and **Fig. 6c** presents the mean of estimates from the 500 maps of ΔS_w provided by 4D seismic inversions. Reference and seismic information were scale transferred, therefore, available data (reference, simulation and seismic) are in the simulation scale (90x110x9).

Likewise, **Fig. 7a** presents pressure change indicated by the reference model. The mean of 500 Δp maps from simulation models is presented in **Fig. 7b** and **7c** shows mean of 500 Δp maps yielded from the 4D seismic inversion.

Results

Pressure analysis

As we can see in **Fig. 7b**, the average Δp estimate from the simulation models is homogeneous into the reservoir zones bounded by the faults. For example, the drained area of wells P1, P2, P9 and P11 indicates a mean value for Δp of -14 MPa. The reservoir area between faults A and B, presents a mean value of Δp equal to -8 MPa. In the region from fault C to D, Δp is roughly zero. Hence, the Δp estimate from history-matched models (using well data) is delimited by the presence of faults. Observing the reference map **(Fig. 7a)** we can see that the reference pressure behavior is homogenous throughout the reservoir. It indicates that the 500 models considered here, extracted from an intermediate step of the well history matching, do not have a proper pressure calibration.

From a qualitative and visual analysis, we can see that the estimates provided by 4D seismic data **(Fig. 7c)** are closer to the reference map **(Fig. 7a)** than estimates from the simulation models. So the Δp estimates from 4D seismic can be used to calibrate estimates from history-matched models.

The above analysis is deceptively simple, as it is impossible to be performed for a real case due to the lack of the reference response of the reservoir. Thus, two issues must be highlighted: (1) how to guarantee correct analysis of uncertainties without a reference model and (2) how to quantitatively integrate probabilistic information from 4D seismic and simulation data using more information than just the mean values from data sets.

The methodology proposed here, uses the overlapping coefficient to address these issues. The OVL-based methodology uses probability density functions, PDFs, to represent the available data. It brings more detail about distribution of studied data than a simple analysis of the mean average. Thus, it might convey some information that the mean from probabilistic data might not.

By computing the parameters OVL_{SEIS} and OVL_{SIM} for every grid block and gathering this information in the OVL_{SEIS} and OVL_{SIM} maps, as shown in **Fig. 8**, we can observe the overall behavior of pressure change from 4D seismic and simulation models taking into account their variability

Three different zones are visible in **Fig. 8a:** (1) where the estimations show low precision (OVL_{SIM} < 80%), (2) where the precision of simulation estimates is high (OVL_{SIM} >80%) and (3) a heterogeneous zone in the north, where OVL_{SIM} varies from 0% to 100%.

Regarding OVL_{SEIS} map, **Fig. 8b**, the estimates from 4D seismic appear to be precise in most grid blocks. A few grid blocks, inside the rectangle (wells P1, P2, P9, P11, I1 and I2), presented OVL_{SEIS} lower than 80%.

Different from the mean maps in **Fig. 7b** and **7c**, the OVL_{SEIS} and OVL_{SIM} maps do not show the mean behavior of pressure change estimates, but how those estimates are distributed in every grid block and how precise these distributions are. OVL_{SEIS} or OVL_{SIM} greater than 80% show precise estimates of Δp, whereas values under 80% indicate estimates with lower precision, i.e. greater variability.

We cross plotted the maps of **Fig. 8** and grouped the OVL information for every grid block in the four proposed regions, as presents **Fig. 9a**. The points concentrate in region (1), where 4D seismic and simulation distributions look very much alike and in region (4), where seismic distribution is more precise than simulation distribution. Some points in region (3) show locations where seismic and simulation estimates indicate different ranges and a few points in region (2) when simulation is more precise than seismic. **Fig. 9b** illustrates the map from the cross plot information, allowing us identify the four regions in the reservoir configuration.

To validate the procedure, **Fig. 10** presents some examples of grid blocks where the PDF of Δp from 4D seismic and from simulation models identified regions (1), (2), (3) and (4). At region (1) both PDFs show the same trend, indicating properly calibrated reservoir grid blocks (data agreement). In region (2), simulation distribution varies less than seismic distribution while in region (4) seismic distribution varies less. In region (3) seismic and simulation distributions present different ranges, therefore, as a future step, a more detailed analysis of the characteristics from those grid blocks must be carried out to improve the uncertainty mapping and definition of all data.

The first application of the 4 regions map is to use the information from 4D seismic data at region (4). In these areas, estimations from 500 Δp maps provided by 4D seismic data presented lower variability than estimates from well history matched models; therefore, seismic is useful to reduce uncertainties in reservoir simulation.

We selected simulation models that reproduce the range indicated by 4D seismic distribution. **Fig. 11** presents an example of selection using 4D seismic data for one grid block. Initial distribution from simulation is more variable, i.e. less precise, than selected simulation models within 4D seismic range.

Fig. 11 presents the selection for one grid block. We applied it for every grid block in region (4), then gathered the 10% most frequently selected simulation models (considering all blocks), and then computed the mean from those models to finally obtain the Δp map in **Fig. 12c.**

The selected simulation models **(Fig. 12c)** are closer to the expected answer **(Fig. 12a)** than the initial models **(Fig. 12b)**. Note that the reservoir locations in blue in **Fig. 12b**, where the ratio between the mean of the initial Δp estimates from simulation and the reference value **(Fig. 12a)**, was greater than 2.5. The selected models presented values closer to the reference maps, with a ratio between Δp estimates from selected models and reference value of roughly 1.3 (where ratio 1 indicates identical values). These regions showed the greatest improvement although the others also improved. The selection generated estimates very close to the reference value as well as reduced uncertainties, since selected models present more precise distributions.

Also note that we chose models according to the information from a small number of blocks (only those classified in region 4). So, the application of the OVL map allows the targeted use of 4D seismic information, differing from traditional methods that compare the quadratic difference between simulation and seismic data (for every grid block of the reservoir) as presented by Almeida *et al.* (2014). Assunção *et al.* (2016) compares the application of both procedures to select the simulation models that best honor the observed 4D seismic data. They show that the OVL can provide a set of selected simulation models more statically consistent with the 4D seismic data than the quadratic difference method.

Saturation analysis

Fig. 6c presents the average of the estimates of water saturation change from 4D seismic data. The mean value shows some variation of water in regions far from injectors, which were unexpected according to the true answer, representing zones with problematic seismic signals. Nonetheless, 4D seismic data show lower "rings" of water fronts than simulation, honoring the reference response.

Similarly to the pressure analysis previously presented, we calculated the parameters OVL_{SIM} and OVL_{SEIS} for every grid block and grouped in OVL_{SIM} and OVL_{SEIS} maps, seen in **Fig. 13.** For regions close to injector wells, both values, OVL_{SIM} and OVL_{SEIS} are high. Nonetheless, the water front OVL_{SEIS} is greater than OVL_{SIM}, as highlighted in the green rectangle.

At the reservoir locations far from injector wells, such as the drainage area of production wells P8 and P9 (green arrows), OVL_{SEIS} is lower than OVL_{SIM}, as a result, we conclude that in these regions information from simulation is more precise than 4D seismic data.

Cross plotting OVL maps **(Fig. 14a)** and mapping the four regions **(Fig. 14b)**, we can identify locations where seismic data is more precise than simulation (region 4) and where simulation data is more precise than seismic (region 2). In region 4, seismic data must be used to reduce uncertainties in water front estimations provided by history matched models and in region 2 the simulation model results can be used as constraints to reduce noise and uncertainties in the probabilistic 4D seismic inversion.

Selecting simulation models within the seismic distribution range, for every reservoir grid block at region 4 can be useful to calibrate simulation estimations. The method here applied is the same performed in the pressure analysis. We selected the simulation models at seismic range for every grid block and gathered the 10% (50 out of 500) most frequently selected models. **Fig. 15** presents these results.

Fig. 15c shows that around injector well I1, the waterfront zone was minimized when seismic information was incorporated. A perfect match of the saturation front was not expected because we are selecting only the best 50 models out of the 500 models available that did not include a perfect model. This type of procedure should be incorporated into a history-matching process (changing the reservoir uncertainties and generating new models) and by doing so we can generate simulation models that yield saturation fronts closer to those observed from 4D seismic data. However, history matching process is not the objective of this study; instead we present a new tool to assist this process.

The methodology proved to identify models that better estimate Δp and ΔS_w than initial models. Moreover, the application of OVL methodology also identified a critical zone, region 3 **(Figs. 9b and 14b)**. In this region, simulation and seismic estimations disagree without any indication of which is more accurate. In this case we must redefine the uncertainties. This methodology provided further useful information about region 2: the simulation estimations are more precise and could be used as a constraint to reinterpret 4D seismic attributes.

Conclusions

In this work, we proposed a tool to evaluate the agreement of probabilistic data from reservoir simulation and 4D seismic data as independent measurements (decoupled). The base of the methodology is the overlapping coefficient that enables identifying reservoir locations with high and low misfit, evaluating which technology, 4D seismic or numerical simulation, is more precise.

We applied the proposed methodology to a synthetic dataset with the following results:

(1) reduced uncertainties in reservoir models using probabilistic 4D seismic inversion, (2) improved Δp estimates from simulation models previously history matched using only well data and (3) calibrated the waterfront nearby injector wells. Moreover, we identified reservoir locations where information from simulation models can be applied to improve 4D seismic interpretation, that is, using simulation data to constrain seismic inversions (engineering-consistent manner), as proposed, by Davolio *et al.* (2013) and Tian *et al.* (2014). We also identified critical reservoir zones for reevaluation, since the high disagreement between the two data can be an indicative of the presence of “unknown unknowns”.

It is important to highlight that the present work integrated information from ΔS_w and Δp maps, but information from other 4D seismic attributes, such as maps of acoustic impedance, can also be used.

The next step is to integrate a history matching procedure, in short: from the selected models, perform a new history-matching procedure, and obtain another set of pressure and water saturation change maps from simulation data. The OVL comparison could then be performed again with the available 4D seismic data.

Acknowledgments

This work was carried out in association with the ongoing Project registered as "BG-07 Redução de incertezas através da incorporação de sísmica 4D na modelagem do reservatório" (UNICAMP/BG Brasil/ANP) funded by BG Brasil under the ANP R&D levy as "Compromisso de Investimentos com Pesquisa e Desenvolvimento". The authors thank also UNISIM, DE-FEM-UNICAMP, CEPETRO, and CAPES for supporting this work and CMG and Schlumberger for software licenses.

References

Almeida, F. L., Davolio, A., and Schiozer, D. J. (2014). A New Approach to Perform a Probabilistic and Multi-objective History Matching. *SPE Annual Technical Conference and Exhibition.* Amsterdan, The Netherlands.

Al-Saleh, M. F., and Samawi, H. M. (2007). Inference on Overlapping Coefficients in Two Exponential Populations. *Journal of Modern Applied Statistical Methods*, 503- 516.

Assunção, G. S., Davolio, A., and Schiozer, D. J. (2016). A COMPARATIVE STUDY OF TWO METHODOLOGIES TO INTEGRATE RESERVOIR SIMULATION AND 4D SEISMIC DATA. *Rio Oil and Gas.* Rio de Janeiro, Brazil.

Botev, Z. I., Grotowski, J. F., and Kroese, D. P. (2010). KERNEL DENSITY ESTIMATION VIA DIFFUSION. *The Annals of Statistics, 38*(5), 2916-2957.

Bradley, E. L. (2006). Overlapping Coefficient . *Encyclopedia of Statistical Sciences*.

Castro, S. A., Caers, J., Otterlei, C. M. et al (2009). Incorporating seismic data into reservoir models while honoring production and geologic data: A case study. *The Leading Edge*(28(12)), 1498-1506.

Correia, G., Davolio, A., and Schiozer, D. J. (2016). Multi-Scale Integration of 4D Seismic and Simulation Data to Improve Saturation Estimations. *SPE EUROPEC.* Vienna, Austria.

Davolio, A., Maschio, C., and Schiozer, D. J. (2013a). A methodology to constrain pressure and saturation estimation from 4D seismic using multiple simulation models and observed data. *Journal of Petroleum Science and Engineering*, 51-61.

Davolio, A., Maschio, C., and Schiozer, D. J. (2013b). Local History Matching Using 4D Seismic Data and Multiple Models Combination. *EAGE Annual Conference and Exhibition.* London, United Kingdom.

Davolio, A., Maschio, C., and Schiozer, D. J. (2014). A methodology to calibrate water saturation estimated from 4D seismic. *Journal of Geophysics and Engineering, 11*, 1-16.

Davolio, A., and Schiozer, D. J. (2015). Joining Two Probabilistic Approaches - 4D Seismic Data and Simulation Models. *77th EAGE Conference and Exhibition.* Madrid, Spain.

Elsheikh, A. H., Jackson, M. D., and Laforce, T. C. (2012). Bayesian Reservoir History Matching Considering Model and Parameter Uncertainties. *Mathematical Geosciences, 44*, 515-543.

Evensen, G., Hove, J., Meisingset, H. et al (2007). Using the EnKF for Assisted History Matching of a North Sea Reservoir Mode. *SPE Reservoir Simulation Symposium.* Woodlands, Texas.

Gervais-Couplet, V., Roggero, F., Feraille, M. D., Le Ravalec-Dupin, M., and Seiler, A. (2010). Joint History Matching of Production and 4D-Seismic Related Data for a North Sea Field Case. *SPE Annual Technical Conference and Exhibition.* Florence, Italy.

Landa, J. L., and Kumar, D. (2011). Joint Inversion of 4D Seismic and Production Data. *SPE Annual Technical Conference and Exhibition* . Denver, Colorado, U.S.A.

Maschio, C., and Schiozer, D. J. (2014). Bayesian history matching using artificial neural network and Markov Chain Monte Carlo. *Journal of Petroleum Science and Engineering, 123*, pp. 62-71.

Morell, E. (2010).History Matching of Norne Field. Master's thesis, Norwegian University of Science and Technology, Norway.

Oliver, D. S., and Chen, Y. (2011).Recent Progress on Reservoir History Matching: A Review. *Computational Geosciences*,15, 185-221.

Pazetti, B. A., Davolio, A., and Schiozer, D. J. (2015). The Impact of Rock and Fluid Uncertainties in the Estimation of Saturation and Pressure from a 4D Petro Elastic Inversion. *Journal of Geophysics and Engineering,*, 686-701.

Riazi, N., Lines, L., and Russell, B. (2013). Integration of time-lapse seismic analysis with reservoir simulation. *Integration geoConvention.* Canada.

Schiozer, D. J., Avansi, G. D., and Santos, A. A. (2014). Methodology for Risk Quantification Combining Geoestatistical Realizations and Discretized Latin Hypercube . *Computers and Geosciences*.

Schmid, F., and Schmidt, A. (2006). Nonparametric Estimation of the Coefficient of Overlapping - Theory and Empirical Application. *Computational Statistics and Data Analysis*, 1583-1596.

Stephen, K. D., Soldo, J., Macbeth, C., and Christie, M. (2004). A MULTIPLE MODEL APPROACH TO HISTORY MATCHING AND UNCERTAINTY ANALYSIS USING TIME-LAPSE SEISMIC. *9th European Conference on the Mathematics of Oil Recovery.* Cannes, France.

Tian, S., Shams, A., and MacBeth, C. (2014). Closing the loops using engineering-consistent 4D seismic inversion. *The Leading Edge, 33*, 188-196.

Weitzman, M. (1970). *Measure of the overlap of income distribution of while and Negro families in the United States.* U. S. Department of Commerce, Bureau of the Census,, Washington, DC.

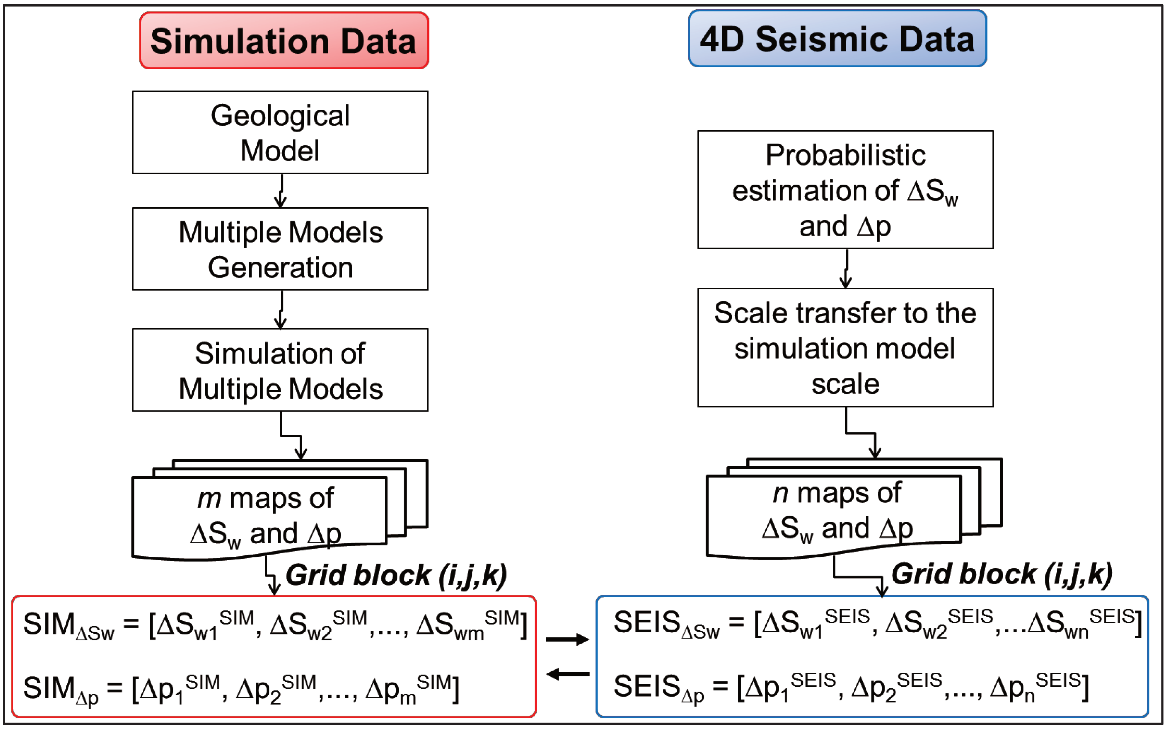

Fig. 1—Workflow to compare 4D seismic and simulation data.

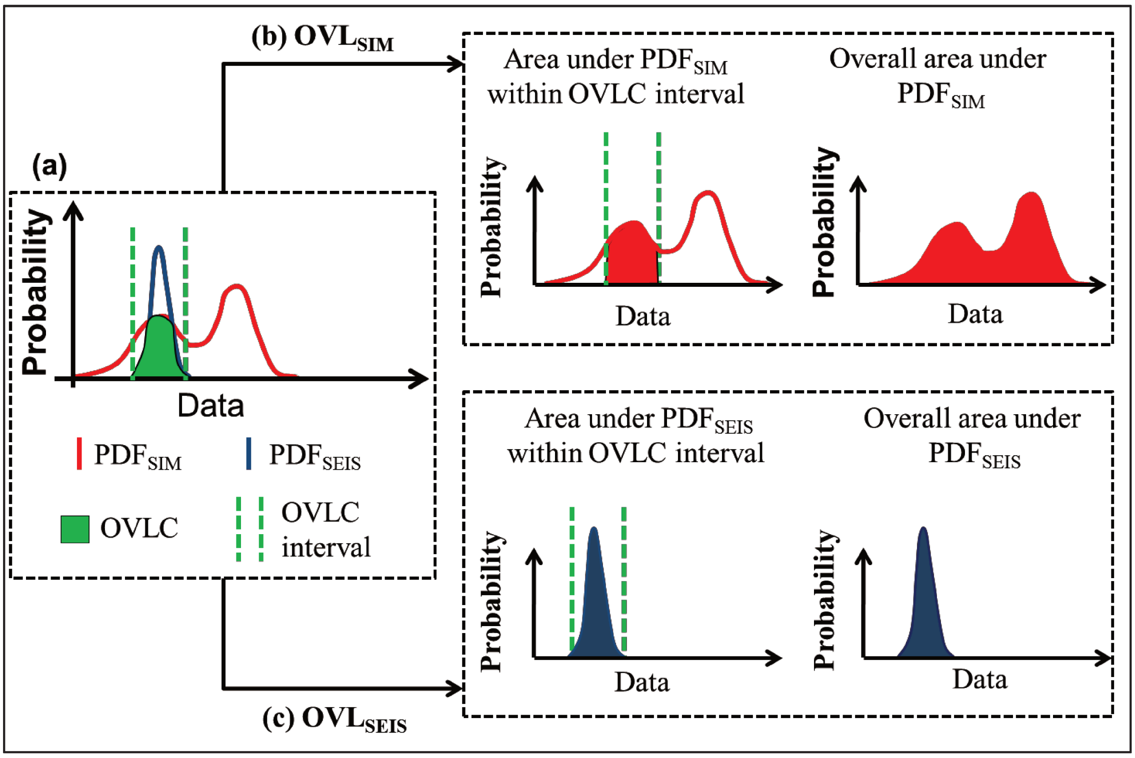

Fig. 2—(a) OVL interval, (b) OVL_{SIM} and (c) OVL_{SEIS}.

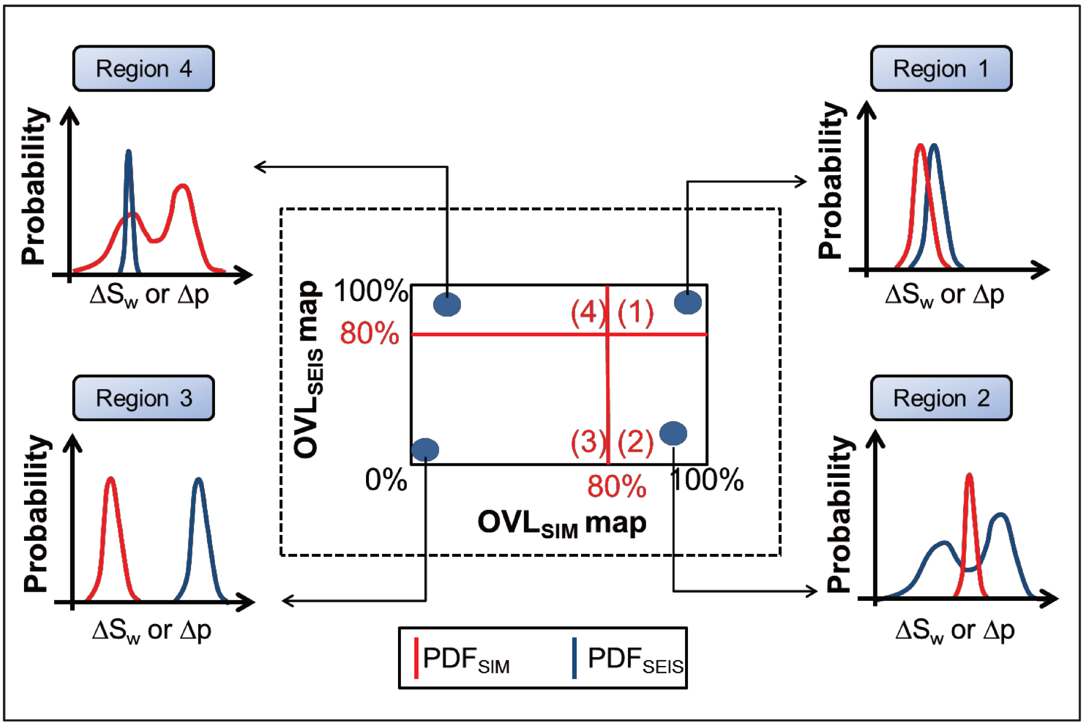

Fig. 3—Cross plot of OVL_{SEIS} and OVL_{SIM} maps and identification of the four regions.

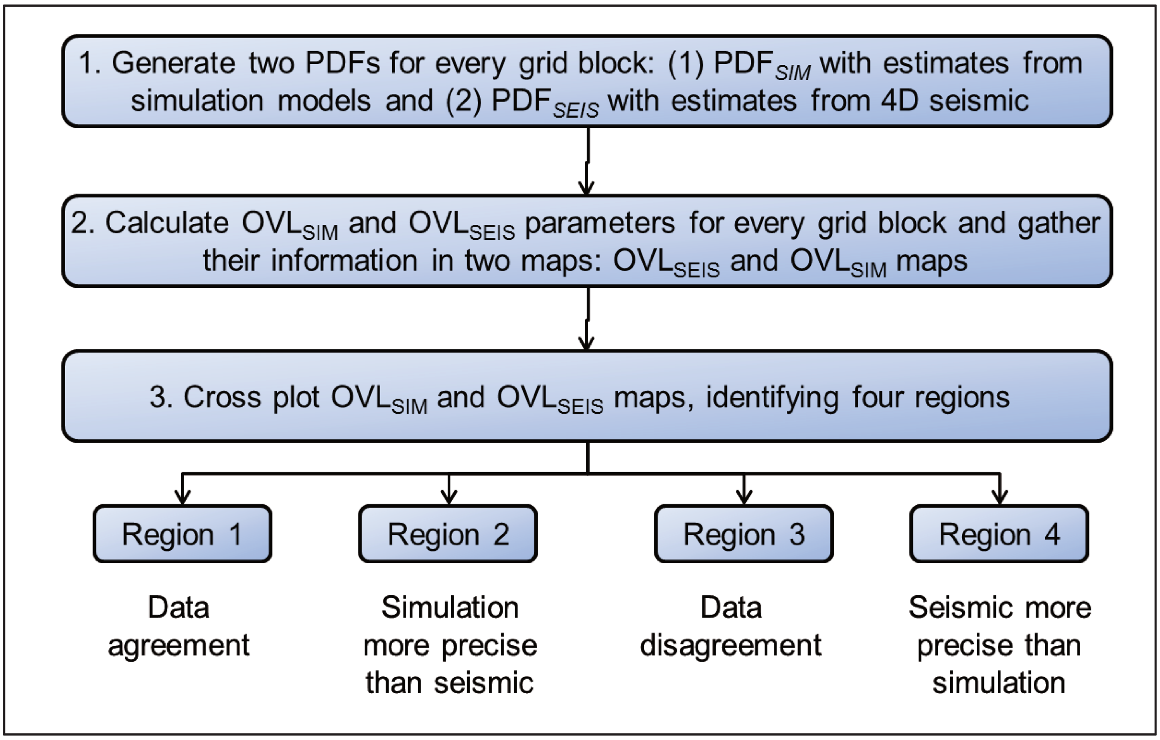

Fig. 4—OVL methodology workflow.

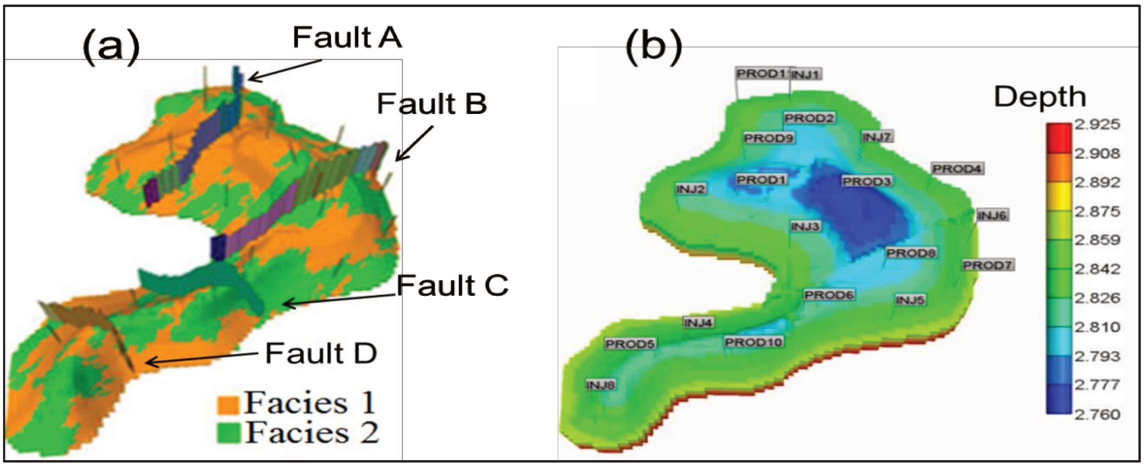

Fig. 5—Reference model: (a) faults and facies and (b) Well locations and depth of reference model.

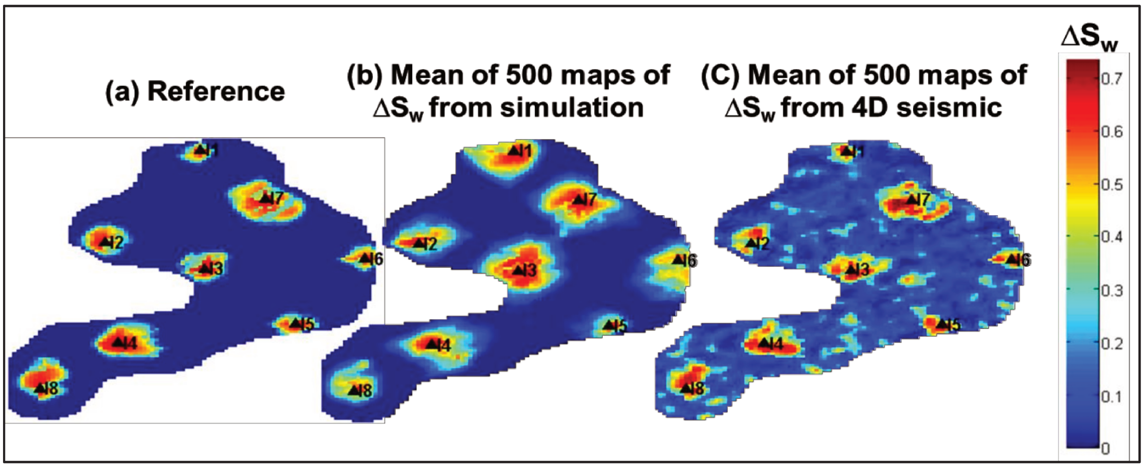

Fig. 6—Maps of water saturation changes - (a) reference map, (b) mean of simulation ΔS_w maps and (c) mean of 4D seismic ΔS_w maps (every map is showing the 3rd layer out of 9).

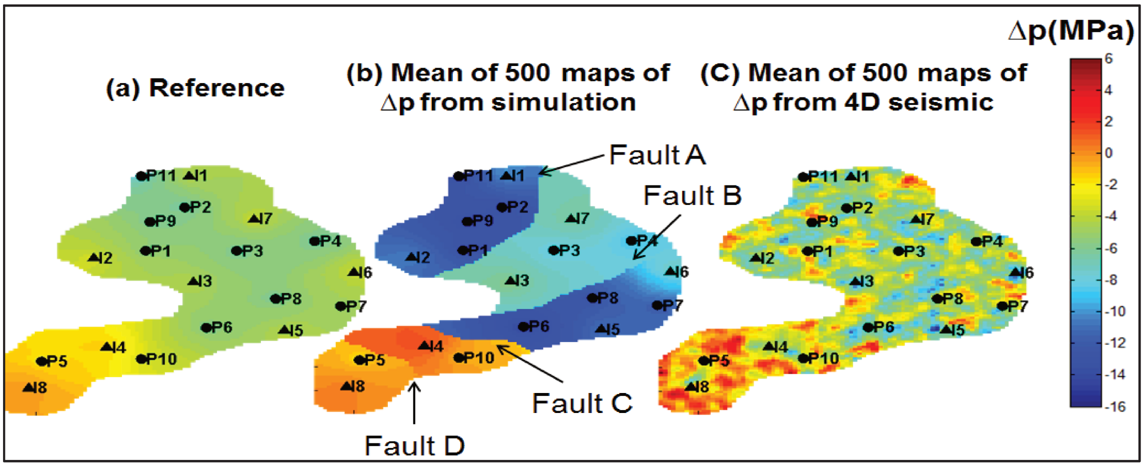

Fig. 7—Maps of pressure changes - (a) Reference map, (b) mean of simulation Δp maps and (c) mean of 4D seismic Δp maps (every map is showing the 3rd layer out of 9).

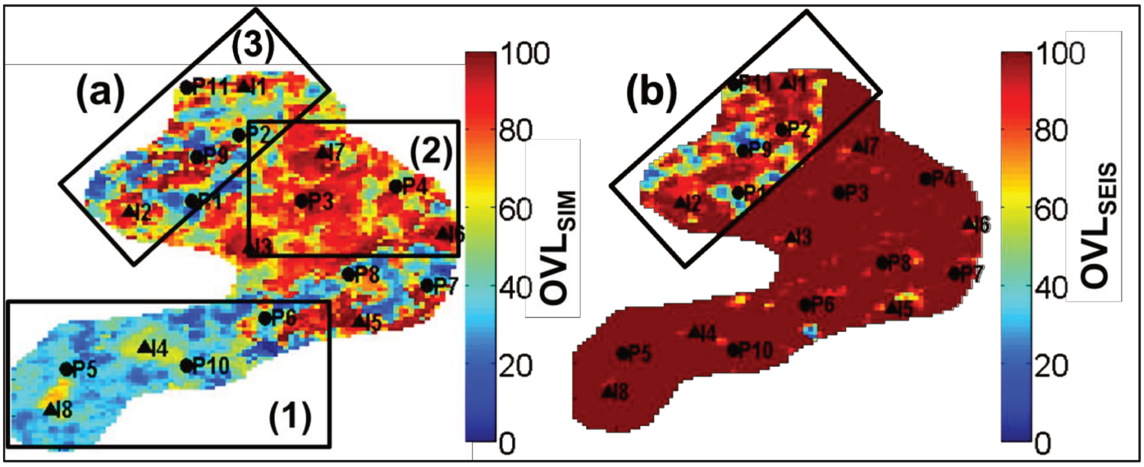

Fig. 8—(a) OVL_{SIM} map and (b) OVL_{SEIS} map for pressure analysis (every map is showing layer 3 out of 9 layers).

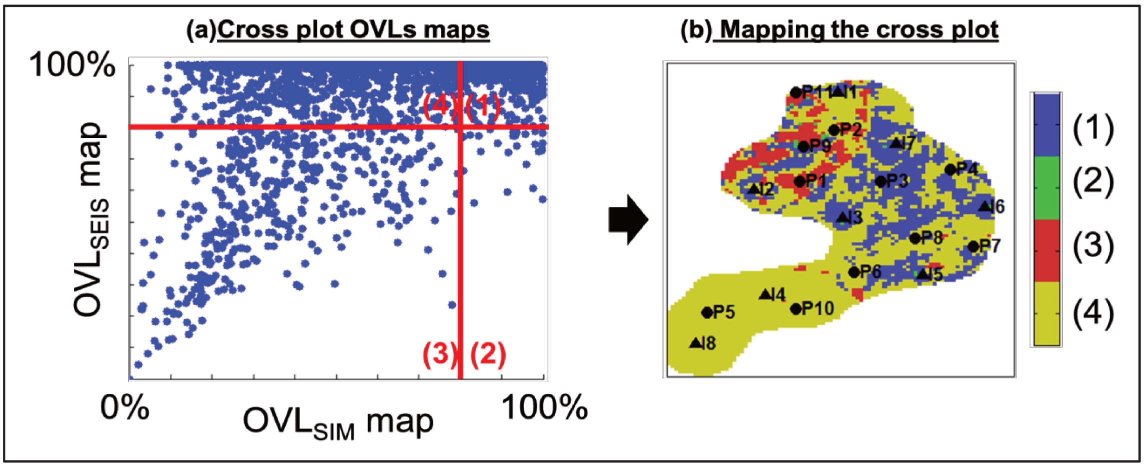

Fig. 9—(a) Cross plot of OVL_{SEIS} and OVL_{SIM} maps and (b) Map of the cross plot (4 regions map is presenting layer 3 out of 9 layers).

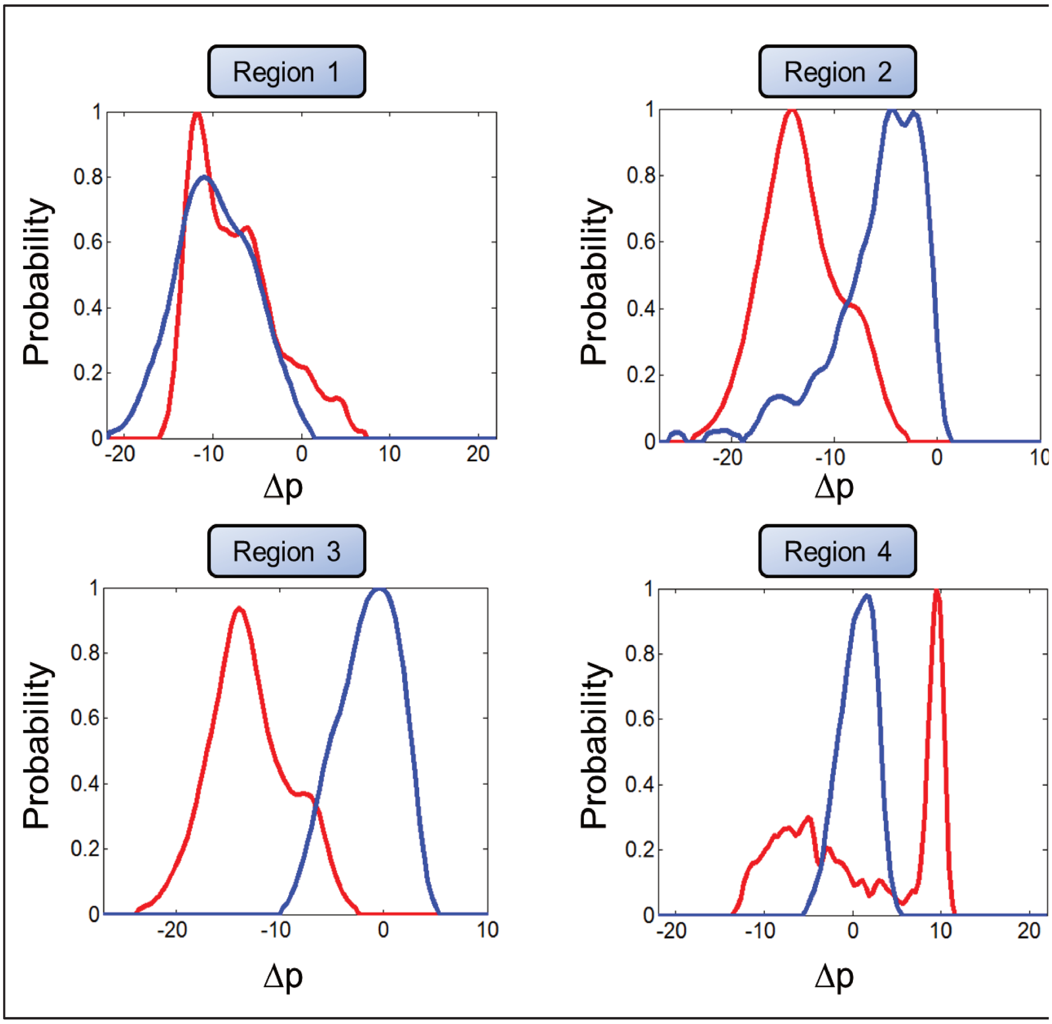

Fig. 10—Analysis of distributions at region (1), (2), (3) and (4).

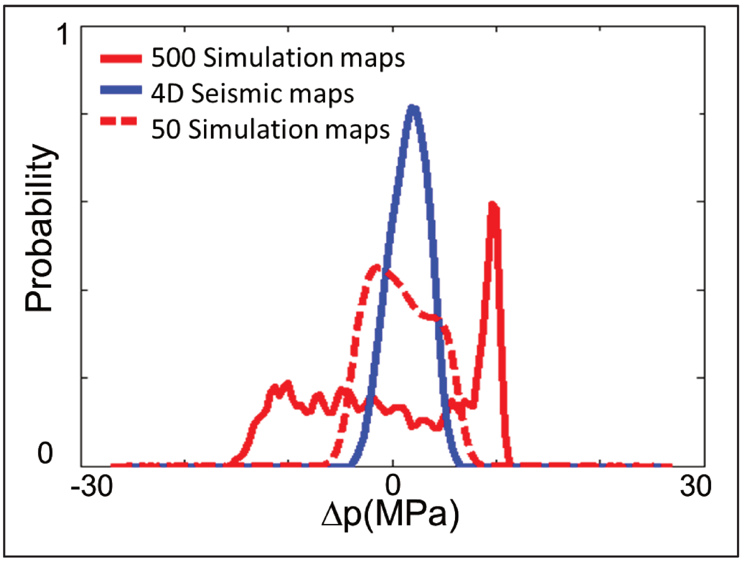

Fig. 11— Example selection performed using 4D seismic information.

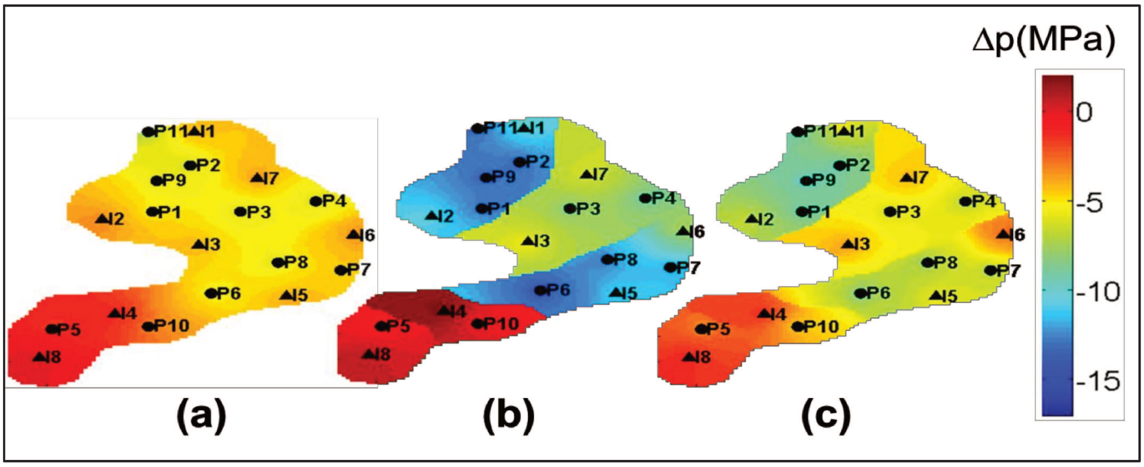

Fig. 12— (a) Expected value, (b) Mean of estimates of pressure change from initial simulation models and (c) Mean of selected models using 4D seismic (every map is showing layer 3 out of 9 layers).

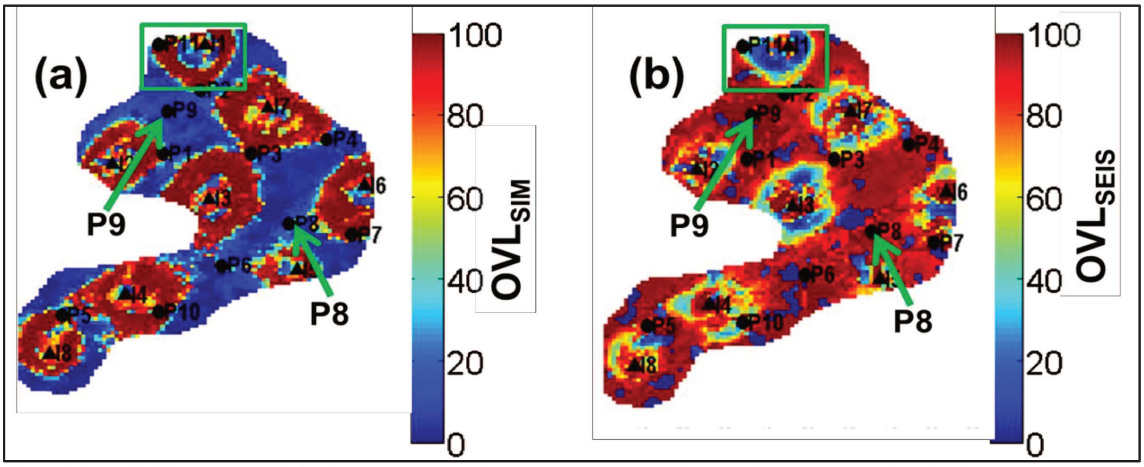

Fig. 13— OVL_{SEIS} and OVL_{SIM} maps in water saturation change analysis (every map is showing layer 3 out of 9 layers).

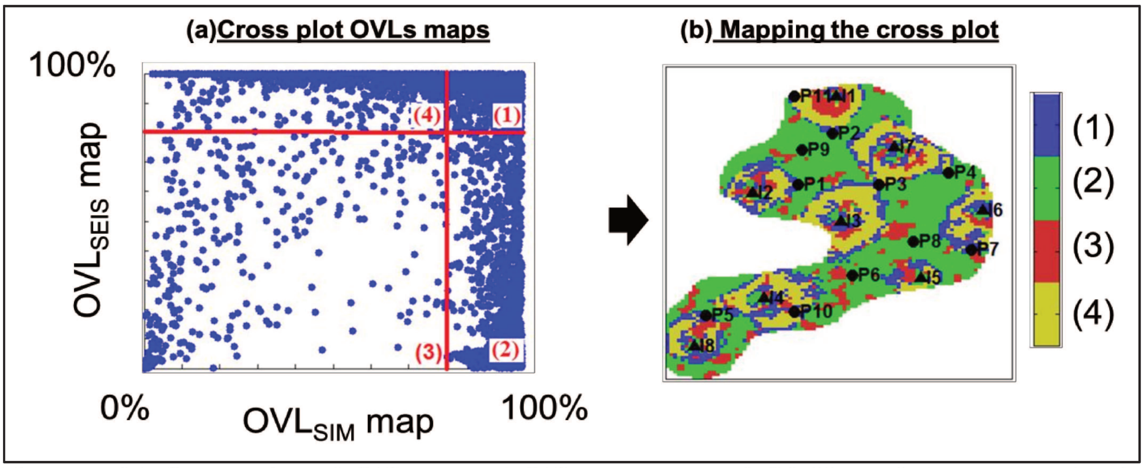

Fig. 14—(a) Cross plot of OVL_{SEIS} and OVL_{SIM} maps and (b) Map of the cross plot (4 regions map is presenting layer 3 out of 9 layers).

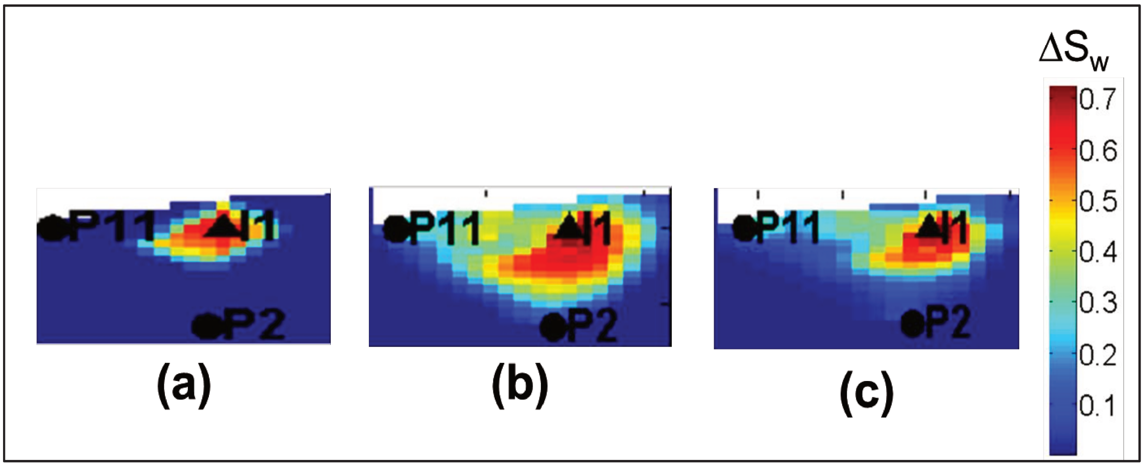

Fig. 15—Water saturation changes around injector I1. (a) Reference model, (b) Mean of estimates from initial simulation models and (c) Mean of selected models using 4D seismic (every map is showing layer 3 out of 9 layers).

3. PAPER 2: "A Comparative Study of two Methodologies to Integrate Reservoir Simulation and 4D Seismic Data"

Germano S. C. Assunção, Alessandra Davolio, Denis J. Schiozer.

This Technical Paper was prepared for presentation at the Rio Oil & Gas Expo and Conference 2016, held between October, 24-27, 2016, in Rio de Janeiro.

RIO OIL & GAS 2016 EXPO AND CONFERENCE
24-27 de outubro de 2016
Riocentro | Rio de Janeiro - Brasil
Promoção: ibp

IBP1456_16

A COMPARATIVE STUDY OF TWO METHODOLOGIES TO INTEGRATE RESERVOIR SIMULATION AND 4D SEISMIC DATA

Germano S.C. Assunção[1], Alessandra Davolio [2], Denis J. Schiozer[3]

Abstract

In the present study, two methodologies to integrate 4D seismic and reservoir simulation data were compared. The first one, named quadratic difference (QD), follows the traditional way, using deterministic seismic information. In such method, 4D seismic information is considered an additional observed dynamic data that the simulation models must honor. The second method, called overlapping (OVL) methodology, considers the seismic information obtained from a probabilistic inversion; instead of using the 4D seismic as an observed accurate data, their uncertainties are also considered.

This study was performed in a synthetic reservoir, with intermediate complexity and several uncertainties. Maps of pressure and water saturation changes from reservoir simulation models and 4D seismic data were the information considered.

The methodologies were used to select the most representative pressure and saturation changes maps from simulation models. After the selection of simulation models using both methodologies separately, the differences between them were analyzed. In the pressure analysis, the most pronounced difference was in the pressure change maps: OVL methodology showed to be more statistically consistent than QD in incorporating seismic information, since OVL only used the seismic data in locations where it were more precise than simulation data. In addition, comparing the trends of the values for transmissibility of the faults, we found that OVL was more accurate than QD to select the simulation models.

In saturation analysis, the same trend was noticed, but as such analysis was performed locally, the results were less pronounced than the results from pressure change maps. Moreover, instead of comparing the transmissibility of the faults, we compared the difference between the porosity and permeability images from selected models from QD and OVL. We found relevant differences in selected permeability and porosity maps from both methodologies, highlighting locations where OVL and QD are showing different trends that should be considered in an optimization process.

1. Introduction

History matching is a procedure that integrates dynamic information in the reservoir modeling workflow, aiming to match simulation and observed data, to obtain reliable models to manage the reservoir performance and provide more realistic production forecast. This dynamic information is usually obtained from wells, e.g. production and pressure data. However, wells are located in specific reservoir regions; therefore, there is a lack of spatial information, which can generate limited and inaccurate calibrations. Thus, new sources of dynamic data have been studied and integrated into the calibration process of the simulation models, notably the 4D seismic (S4D) data.

4D seismic is the repetition of 3D (or 2D) seismic surveys at different times in the reservoir. When comparing the difference between these surveys, it is possible to obtain a spatial view of the fluids displacement as well as the pressure changes in regions that are not covered by the wells.

Traditionally, the history matching process is performed deterministically, i.e., the forecast and production strategies are based on a single model. In this situation, the S4D information (as changes in impedance maps, or saturation and pressure) is integrated as observed data that the simulation model must honor. Risso and Schiozer (2008) show an example of such integration.

Nevertheless, the history matching procedure is an inverse problem and should have multiple possible responses. Combinations of different values of the parameters that characterize the reservoir can generate the observed data, which requires probabilistic approaches to handle these non-uniqueness responses (Xavier *et al.*, 2013; Maschio and Schiozer, 2014; Emerick, 2016). However, even when the history matching is performed probabilistically, 4D seismic is still integrated deterministically, so it is considered as a unique observed data that multiple simulation models must honor, as presents Riazi *et al.*, 2013 and Almeida *et al.*, 2014.

Although uncertainties in seismic data are not usually considered, the acquisition, processing and interpretation of seismic has several limitations that should be taking into account. Using 4D seismic as observed data (deterministic) may not generate consistent and accurate results, since the impact of the seismic errors are neglected. Castro (2007) and Bosh *et al.* (2010) present some drawbacks of deterministic seismic inversion. In the present scenario, it is remarkable the lack of studies that integrate, probabilistically, both data: simulation models and 4D seismic. Landro and Kumar (2011) and Assunção *et al.* (2016) present some of examples found in the literature.

The idea of this work is to compare the information acquired from two methodologies and evaluate the possible differences between them. The first uses the 4D seismic information in a deterministic way (classical approach), while the second uses 4D seismic probabilistically, considering its uncertainties.

2. Objective

The objective of this study is to evaluate the differences between two methodologies that incorporate the 4D seismic data within the history matching workflow. The first methodology, called quadratic difference (QD), uses 4D seismic deterministically following the traditional history matching practices such as the one proposed by Almeida *et al.* (2014). The second methodology, named OVL, considers an approach that handles probabilistic estimates from 4D seismic data and reservoir simulation models.

3. Methodology

In the present study we used a synthetic 4D seismic data that are represented by maps of pressure and saturation changes. Thus, available 4D seismic is n maps of ΔS_w and Δp, yielded from a probabilistic synthetic seismic inversion and transferred to the scale of the simulation models. The information from simulation models are m maps of ΔS_w and Δp, which are the dynamic changes estimated from m simulation models generated from the combination of mapped uncertainties (defined in section 4.2).

The QD methodology, fully presented in section 3.1, is used to select the simulation models using deterministic 4D seismic data (section 3.1.1). Section 3.2 presents the OVL methodology, which also uses the 4D seismic information to select the most representative simulation models. However, the OVL methodology uses probabilistic 4D seismic information (section 3.2.1). The last section of the methodology, section 3.3, present details about the differences evaluated between both methodologies. We emphasize that, due to the characteristics of the data sets, the analysis of Δp and ΔS_w maps are performed separately.

3.1. Quadratic difference (QD) methodology

This methodology incorporates the seismic information deterministically, calculating the quadratic difference between the single map of Δp (or ΔS_w) obtained from the observed 4D seismic information and the m maps of Δp (or ΔS_w) generated from simulation of m models. There are multiple maps from simulation models that must be compared with the observed (deterministic) map obtained from 4D seismic inversion.

The discrepancy between the m maps of Δp from simulation and the single Δp map from 4D seismic is measured by the quadratic difference in every grid block, according to Equation (1).

$$\mathrm{QD}_p(\mathrm{s}) = \sum_{k=1}^{K}\sum_{n=1}^{N}\left(\Delta \mathrm{p}_{SIM_s}{}^{k,n} - \Delta \mathrm{p}_{SEIS}{}^{k,n}\right)^2 \; for\ s = 1, 2 \ldots m \quad (1)$$

Where N is the number of blocks in every layer and K is the number of layers of the models. Variable s represents the simulation model studied. SIM is the estimate from the simulation model and SEIS the estimate yielded from 4D seismic inversion.

In the ΔS_w analysis, due to the data variability in depth, the models comparison is made for each layer separately, as shown Equation (2). Thus, at layer 1 the selected models can be different of the selected models from layer *k*. For the pressure analysis, on the other hand, the selected models are the same concerning all layers.

$$\mathrm{QD}_{Sw}(\mathrm{s},\mathrm{k}) = \sum_{n=1}^{N}\left(\Delta \mathrm{Sw}_{SIM_s}{}^{k,n} - \Delta \mathrm{Sw}_{SEIS}{}^{k,n}\right)^2 \begin{cases} for\ k = 1, 2 \ldots K \\ for\ s = 1, 2 \ldots m \end{cases} \quad (2)$$

3.1.1. Selecting simulation models using QD

After calculating the quadratic difference values, $\mathrm{QD}_p(\mathrm{s})$ and $\mathrm{QD}_{Sw}(\mathrm{s},\mathrm{k})$ the simulation models are placed in ascending order and the ones which presented the lowest values of $\mathrm{QD}_p(\mathrm{s})$ are selected. The same procedure is performed with $\mathrm{QD}_{Sw}(\mathrm{s},\mathrm{k})$.

3.2. Overlapping (OVL) methodology

In this methodology, the *n* maps of Δp (or ΔS_w) from seismic are compared, simultaneously, with the *m* maps of Δp (or ΔS_w) from simulation. The methodology consists in computing the overlapping coefficient (OVL) that shows the percentage of overlapping between the probability density functions (PDFs) generated with the Δp (or ΔS_w) estimates from simulation and seismic data. The PDF are generated using the kernel density estimator proposed by Botev *et al.* (2010) and the OVL is calculated for every grid block independently. From the OVL comparison it is possible to identify four different regions: (1) both dynamic changes estimates from seismic and simulation are similar, (2) simulation estimates are more precise than 4D seismic, (3) the data sets indicate divergent estimates and (4) 4D seismic estimates are more precise than simulation. Figure 1 illustrates the methodology and the "4 Regions Map" that can be obtained gathering the information of every grid block. The methodology can be used as a diagnostic tool (for instance, define the agreement/disagreement of the two data sets and define which data is more precise) or/and to select the most representative models. This second application is explained in section 3.2.1.

3.2.1. Selecting simulation models using OVL

The selection step is presented at the bottom of Figure 1. The idea is to select only the simulation models that present the same dynamic change estimates of the seismic, using only the reservoir locations where seismic is more precise than simulation (Region 4). The selection of the simulation models is performed for every grid block and in the end of the procedure, the most frequently selected models considering all blocks are filtered as good models. The selection here is performed separately, for ΔS_w and Δp, thus selected simulation models from Δp analysis are not the same of ΔS_w. Assunção *et al.* (2016) provide further details about this methodology.

3.3. Evaluating differences between QD and OVL methodology

After the selection of simulation models using both methodologies separately, the differences between them are analyzed. In the pressure analysis, we compared the average map of Δp maps selected by each methodology with the reference (true) map, to observe the improvements. We also compared the 4 Regions Map of the selected maps to observe how each methodology used the 4D seismic information available. Differences in some of the uncertain reservoir parameters such as faults transmissibility were also observed in this analysis.

In saturation analysis, the average map and 4 Regions Map were also used to compare QD and OVL, however the saturation difference was observed locally. Instead of comparing the faults transmissibility, we compared the difference between the porosity and permeability maps from selected models from QD and OVL, as these properties are more related to saturation changes.

4. Application

The present work uses a synthetic reservoir with real characteristics, named Beta. There is a reference model, which was modelled in a fine grid and represents the real reservoir response. The information usually measured in the field (production data, well logging, etc) is extracted from the reference model. The synthetic seismic response are also obtained from the reference model, through forward modelling that uses its simulation results and petrophysical properties.

Based on the information "measured in field" (acquired from the reference model), multiple simulation models were generated, in a courser scale and considering the reservoir uncertainties described below. The following sections bellow presents more details about the reference model and the main uncertainties mapped, as well as the simulation and seismic data sets used.

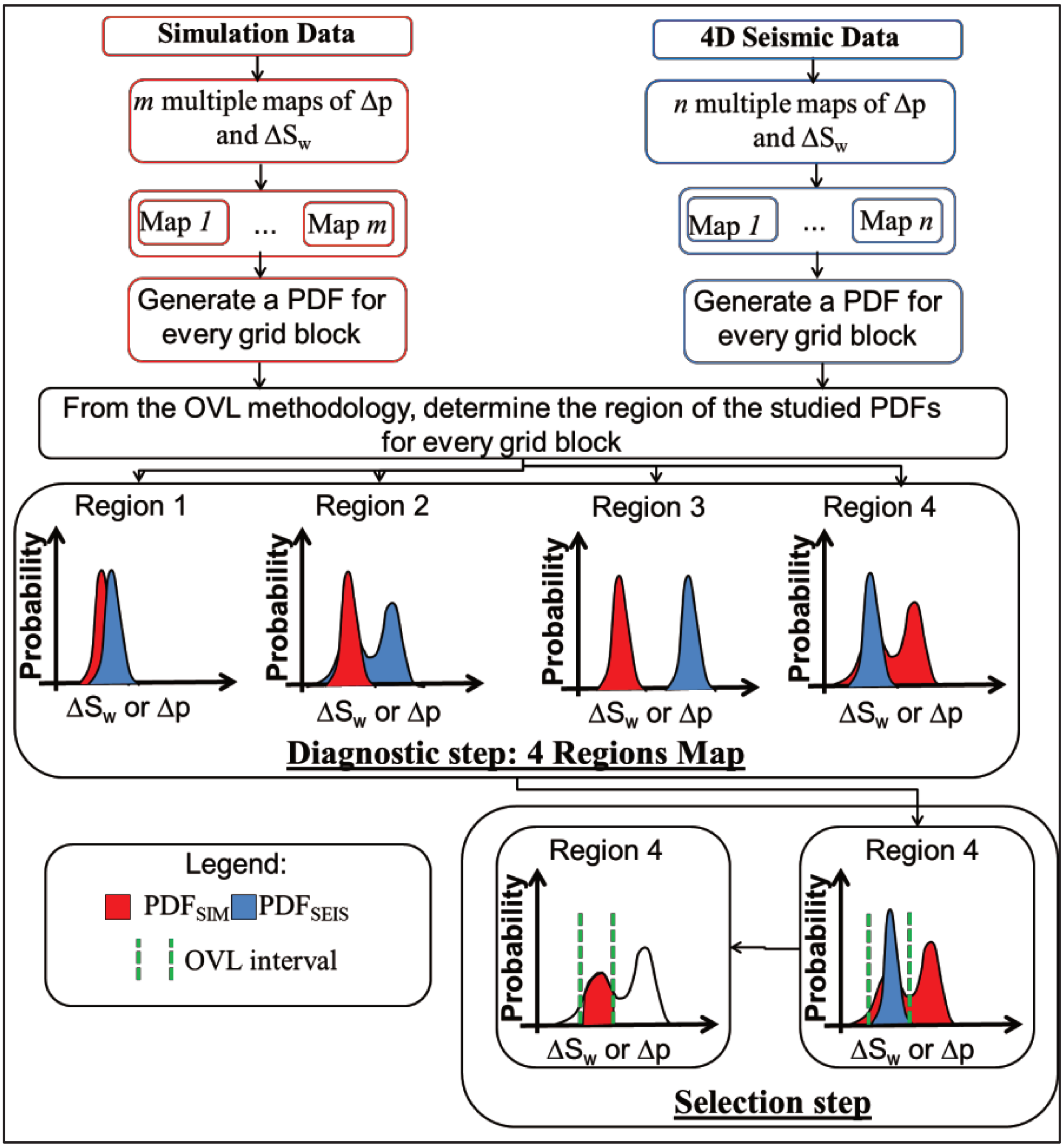

Figure 1–Workflow of the OVL methodology.

4.1. Reference model

The reference model was created from a previous synthetic geological modelling, in a fine grid scale, 270x330 blocks and 18 layers. Its structural framework comprises four seismic faults, 13 sub-seismic sealing faults and two facies (sandstone and shaly-sandstone alternations). It has 11 vertical producers and 8 water injector wells. Figure 2 presents some characteristic of the reference model. More details can be found in Gil *et al.* (2016).

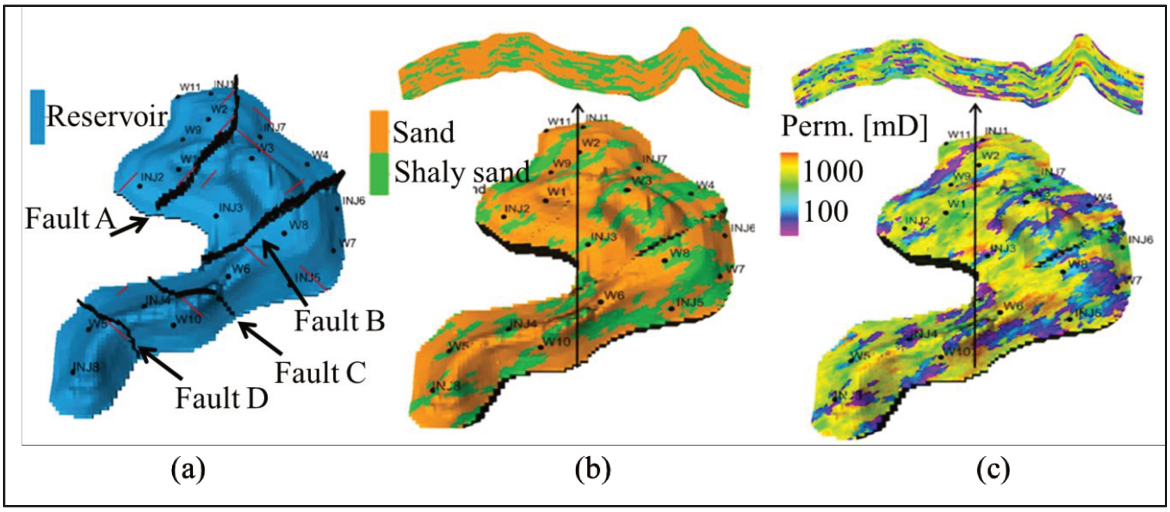

Figure 2–Characteristics of the reference model: (a) Faults and sub-seismic faults (red straight lines), (b) Facies distribution and (c) permeability (Gil *et al.*, 2016).

The reference model is used to validate the response from the methodologies, since it represents the true reservoir answer. The Δp e ΔS_w maps obtained from the reference model were scale transferred to the simulation grid (90x110 blocks and 9 layers). The seismic response was also transferred to the simulation scale. Therefore, all dynamic chance maps used and compared are in the same scale. Figure 3 illustrates the pressure and water saturation changes for layer 1 from the reference model with scale transferred.

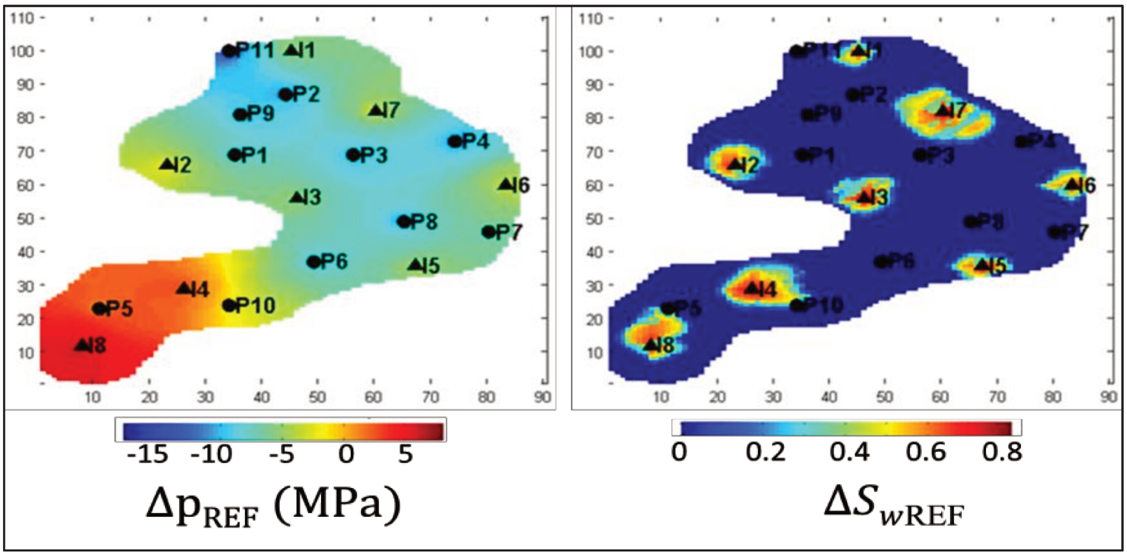

Figure 3–Δp and ΔSw from reference model (layer 3 out of 9).

4.2. Model uncertainties

Seven main uncertainties are considered to generate multiple simulation models: (1) relative permeability of the two facies, (2) ratio of vertical and horizontal permeabilities (Kz /Kx), (3) transmissibility of four faults, (4) facies distribution, (5) porosity, (6) absolute permeability and (7) net-to-gross (NTG). Correia *et al.* (2016) shows the intervals considered for each of the uncertainties listed above.

4.3. Reservoir simulation data

The simulation models are generated from the reference model information, such as well logging. It has a coursed scale, 90x110 blocks and 9 layers. The multiple simulation models were generated combining the main reservoir uncertainties mapped; being that, the true answer (reference model) is not included in the set of models.

The seven uncertainties above mentioned are combined using a sampling technique based on the Latin Hypercube, which generated 500 models. Once generated, the mismatch between the 500 models and production data was assessed and a well history matching procedure was applied. Therefore, the simulation data here used are a set of models obtained from an intermediate step of an iterative and probabilistic well history matching process. More precisely, 500 models obtained from step 2 (out of 4) presented in Almeida *et al.* (2014). From these 500 simulation models, we obtained 500 maps of Δp and ΔS_w. These maps are the information of simulation models used in this study and are presented in Figure 4.

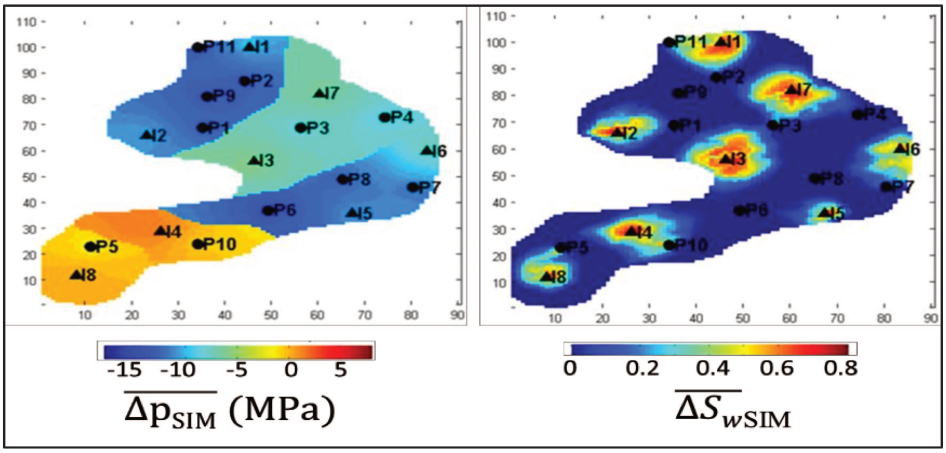

Figure 4–Mean of the 500 maps of Δp and ΔS_w from simulation models (layer 3 out of 9).

4.4. 4D seismic data

Seismic information of this work consists of IP and IS impedances, synthetically built at two different times: pre-production (T0 = 0 days) and after 5 years of production (T1 = 1800 days). This synthetic information was obtained from a petroelastic modelling, which uses the parameters of the reference model as input, such as: porosity, net-to-gross, water saturation and pressure changes at T0 and T1, etc. Further explanation about the petroelatic method used to obtain the synthetic seismic information is presented by Pazetti el al. (2015). Noise is added in these impedance values to produce more realistic synthetic seismic information. This noise addition process is shown in Davolio *et al.* (2014).

Impedances with added noise are the information employed in a probabilistic seismic inversion based on Latin hypercube. This probabilistic inversion generates, from IP and IS impedances, maps with estimates of changes in water saturation and pressure between T0 and T1 (Davolio and Schiozer, 2015). Thus, 500 maps of ΔS_w and Δp are obtained. Figure 5 illustrates the average of the seismic data available.

The deterministic 4D seismic information used in this study is represented by the mean of the ΔS_w and Δp maps as shown in Figure 5.

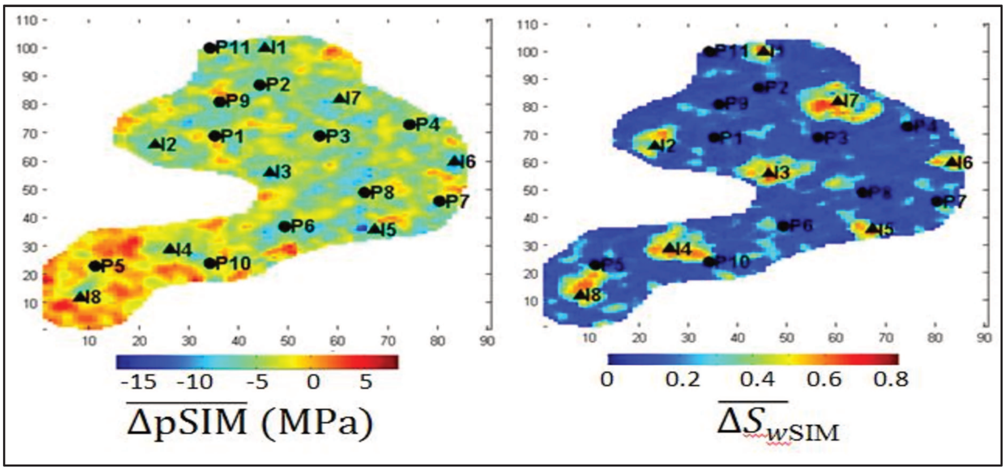

Figure 5–Mean of the 500 maps of Δp and ΔS_w from 4D seismic data (layer 3 out of 9).

5. Results and Discussion

This section introduces the main results obtained. In section 5.1 pressure change maps are studied: we show estimates of the 500 simulation models and then estimates from the selected models from QD and OVL, respectively. The difference between the faults transmissibility are also shown in this section. Section 5.2 presents the results obtained in saturation analysis: saturation changes from the 500 simulation models and the differences between QD and OVL. In addition, in saturation analysis, selected porosity and permeability images from both methodologies are compared.

5.1. Pressure analysis

Figure 6 shows an analysis to access the quality of the 500 simulation models data. In Figure 6a, we can observe the difference between the mean of the 500 Δp maps (Figure 4a) and the reference value of Δp (Figure 3a). There are three critical zones: the first one located in the northwest (wells P1, P2, P9, P11, I1 and I2), the second between the faults B and C and the third between faults C and D. The first and second zones have a difference greater than -9 MPa relatively to the reference value, while the last one shows a discrepancy of more than 3 MPa. This figure also shows the great influence of the faults in the reservoir pressure change, since there is a notable difference in pressure change estimates among the four faults.

By using the OVL methodology as a diagnostic tool, we can compare Δp estimates from the 500 simulation models with the pressure change estimates obtained from the probabilistic 4D seismic. As shown in Figure 6b, there is a huge region in the south of the reservoir where 4D seismic is more precise than simulation, therefore, seismic information can be used and might bring valuable information for the reservoir modelling. In the northwest, there is a predominance of Region 3, that is, seismic and simulation are presenting estimates totally different, therefore this region might be reevaluated. Between faults A and B there are some grid blocks at Region 1, thus, seismic and simulation shows the same trends and are significantly adjusted.

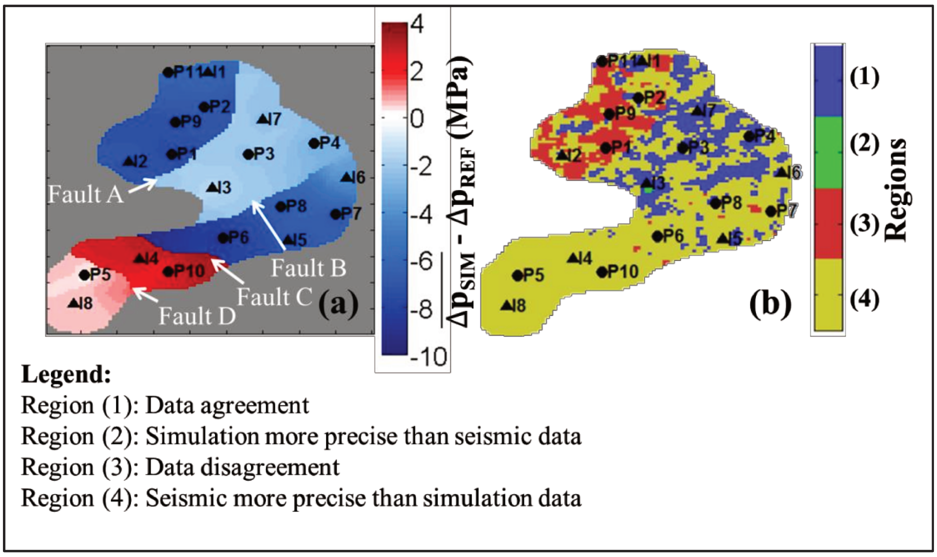

Figure 6–(a) Difference between the mean of the 500 Δp estimates from simulation models and the reference model, (b) 4 regions map obtained from the OVL methodology (layer 3 out of 9).

5.1.1. QD methodology

Performing the quadratic difference methodology to select the most representative simulation models, we obtained the data illustrated in Figure 7. It is important to highlight that this methodology uses deterministic 4D seismic and the 10% most representative models (50 out of 500 simulation models) were selected. This percentage can change depending of the study carried out. Figure 7a shows the average map of these 50 selected models.

The selection using the QD methodology presented good results. The additional dynamic information from 4D seismic reduced considerably the difference between the Δp estimates from simulation models (selected models) and the reference values (compare Figure 6a and Figure 7a). Close to wells P5 and I8 the selected models match the expected value and the three critical zones above described presented considerable improvement: into the first and second zones it has observed a reduction from approximately -9 MPa to around -2 MPa (gain of 78%) and into third zone, between the faults C and D, a reduction from 3 MPa to less than 1 MPa.

Observing the 4 Regions Map (Figure 7b) generated with the 50 models selected, it is noticeable an increase in the number of grid block at Region 1, which indicates other improvement. However, the 4 Regions Map still exhibit regions where 4D seismic is more precise, therefore, could still be used. Moreover, several grid blocks at northwest that initially were indicating Region 3 (Figure 6b) are now showing Region 2. This situation could happen, but we are incorporating 4D seismic data and we expect that the regions where seismic is more precise than simulation (Region 4) change more significantly and not the regions where seismic and simulation were previously indicating to have divergent values.

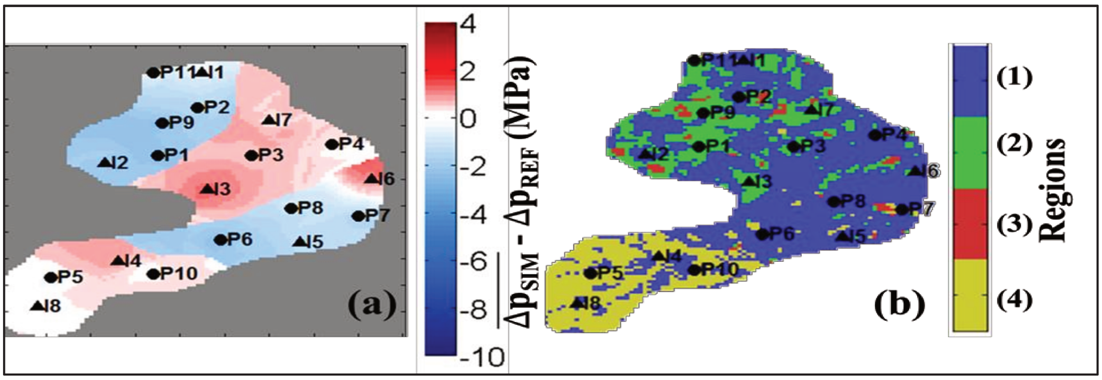

Figure 7–(a) Difference between the mean of the selected Δp maps using the QD methodology and the reference value. (b) 4 regions map obtained from OVL methodology (layer 3 out of 9).

5.1.2. OVL methodology

The second methodology, which is based on the OVL methodology, was used to select the simulation models using probabilistic 4D seismic. It is important to distinguish the two roles of the OVL methodology in present work: the first is the diagnostic tool (where the output of OVL methodology, the 4 Regions Map, is used to compare seismic and simulation data) and the second is the models selection tool (bottom of Figure 1).

Figure 8 presents the results obtained from the 10% most representative selected models from OVL methodology. Comparing Figure 7a and 8a, the OVL methodology shown to be more efficient to incorporate seismic data than QD, mainly in the reservoir locations between faults A and B. Both methodologies presented similar values between faults B and C and in drained area of production wells P1, P2, P9, P11. Close to wells P5, P10, I4 and I8, different trends were observed: OVL presents values between 0 and -1 MPa, while the QD methodology showed values between 0 and 1 MPa in the same region.

The 4 Regions Map (Figure 8b) highlights more differences between both methodologies. We observe that almost all the blocks classified as Region 4 turned to Region 1. Another interesting feature is that reservoir locations that initially were Region 3 (Figure 6b) remained in the same region, showing a more consistent use of the seismic information than the quadratic difference methodology, since just the regions where 4D seismic could be useful changed significantly, while the others, where seismic indicates no valuable information, did not change.

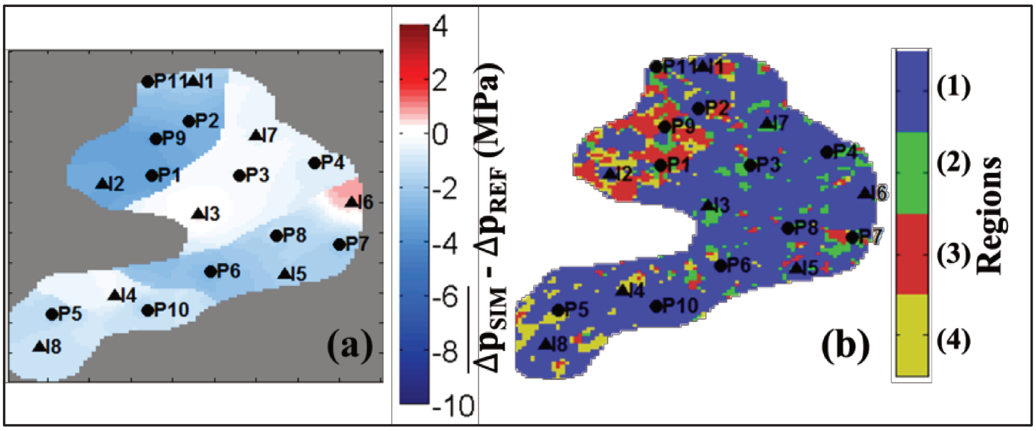

Figure 8–(a) Difference between the mean of the selected Δp maps using probabilistic 4D seismic incorporation and the reference value for Δp, (b) 4 regions map obtained from OVL methodology (layer 3 out of 9).

Figure 9 shows the histograms of the values of transmissibility of the four faults extracted from the simulation models selected by applying both methodologies, QD and OVL. Faults B and C presented similar results for both methodologies. However, the statistical trends of the transmissibility of the faults A and D are different. The most probable level according to the OVL methodology is closer to the reference value than the one pointed by the QD methodology. The difference here observed was not so remarkably because no optimization process was performed. We expect a more pronounced difference if both methodologies were performed jointly within an optimization process.

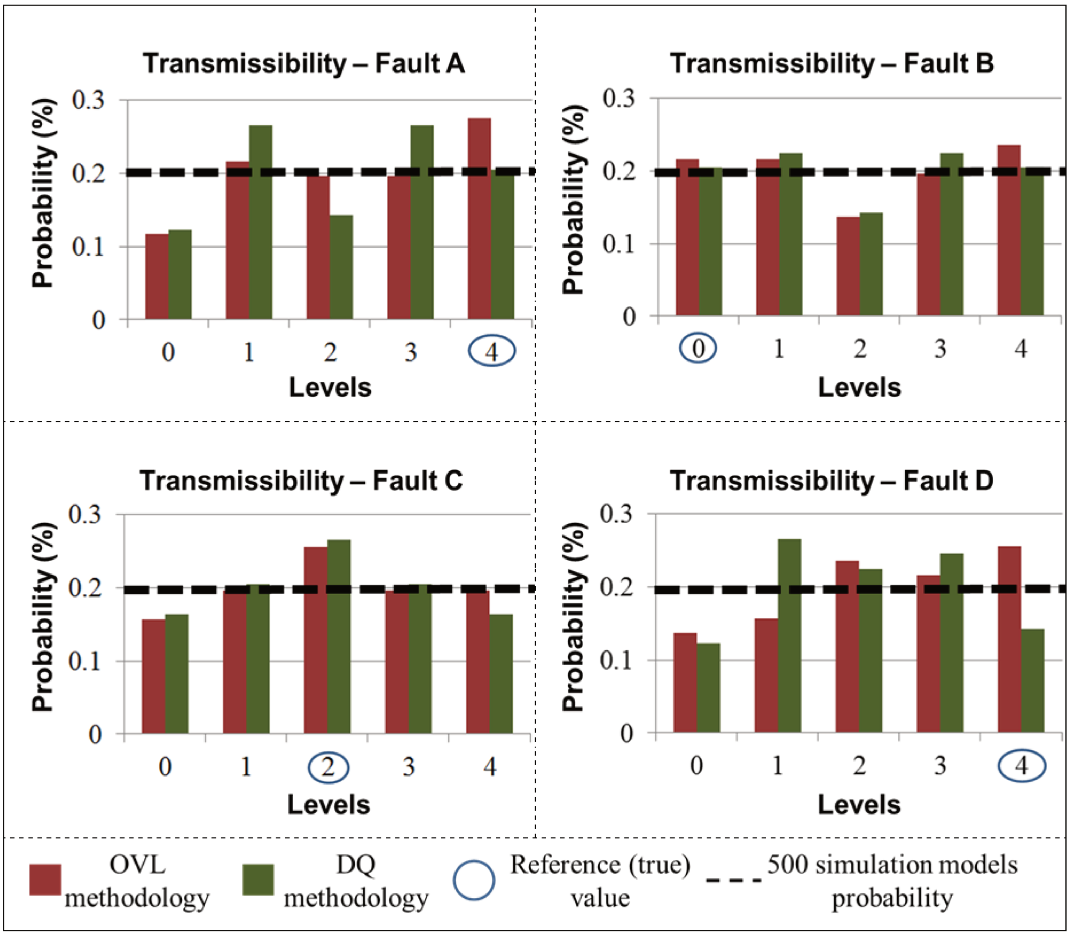

Figure 9–Transmissibility of the four reservoir faults: 500 simulation models probability (black dashed lines) and transmissibility from the selected models.

5.2. Saturation analysis

The same analysis of the pressure change was performed to water saturation change; nonetheless, the ΔS_w analysis was performed locally, since it brings more local information than global. The idea of the local analysis is to try to get information from 4D seismic to better quantify and understand the water front displacement.

Figure 10a shows the reservoir location considered in the present analysis, nearby production wells P2 and P11 and water injector I1. Such location was chosen because there are a considerable difference between the ΔS_w estimated from the 500 simulation models and the reference value. Moreover, there is no influence of any other injector, than I1, in the water saturation change on this region. Figure 10b presents the difference between the mean of ΔS_w estimated from the 500 models and the reference value. We can see a discrepancy between the expected water front and the estimated one, with some grid blocks having a difference in saturation greater than 0.3.

5.2.1. QD and OVL methodologies

The selection performed brought better results when the methodologies were applied, as presented by Figure 11. The most pronounced improvement is found in selected models from OVL methodology (Figure 11c), which presented more grid blocks with the difference between estimated and reference close to zero than the selected models from QD methodology (Figure 11a). The black arrows in Figure 11a and 11c highlight the main differences obtained between QD and OVL. In addition, comparing the 4 Regions Map (Figure 11b and 11d), we can notice that OVL used more consistently 4D seismic information (Region 4) than QD, following the same trend shown in pressure analysis (number of grid blocks in Region 4 after the selection using OVL is lower than QD).

We also compared differences between the porosity and permeability images of the selected models from both methodologies and found some locations where the differences are significant. It is calculated the difference between porosity and permeability from the 50 selected maps from OVL and the 50 selected maps from QD (e.g, for porosity analysis, we calculated $\overline{POR_{OVL}} - \overline{POR_{QD}}$). In order to normalize the results, we divided such difference by the maximum value (among selected maps using OVL and QD) of each parameter studied: porosity, permeability direction I and J and permeability direction k. The values in Figure 12 are shown in percentage and for layer 3. The most pronounced normalized difference was found in images from permeability at direction k, with normalized differences greater than 20% between the selected models from OVL and QD, indicating locations where each methodology showed different trends that could results in pronounced differences if an optimization process had been performed.

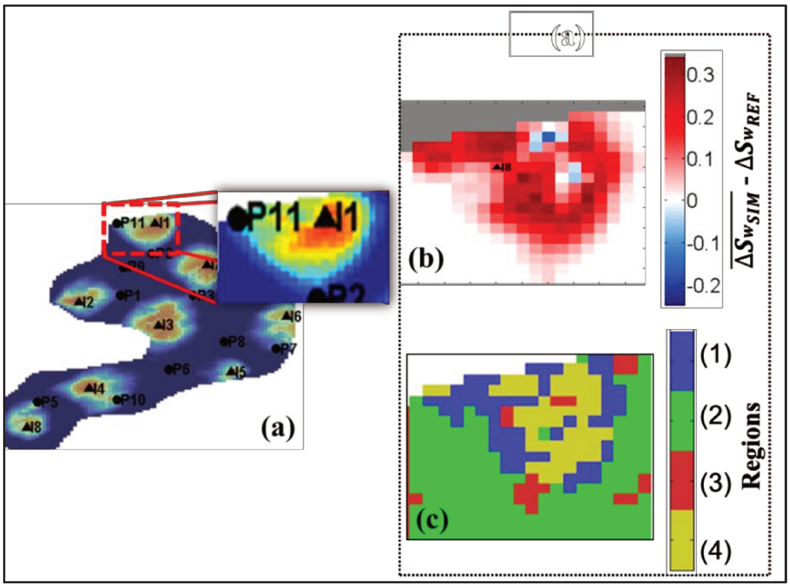

Figure 10–(a) Local region where the analysis was performed, (b) Mean of the 500 ΔSw maps - reference value, (c) 4 Regions Map of the 500 models (layer 3 out of 9).

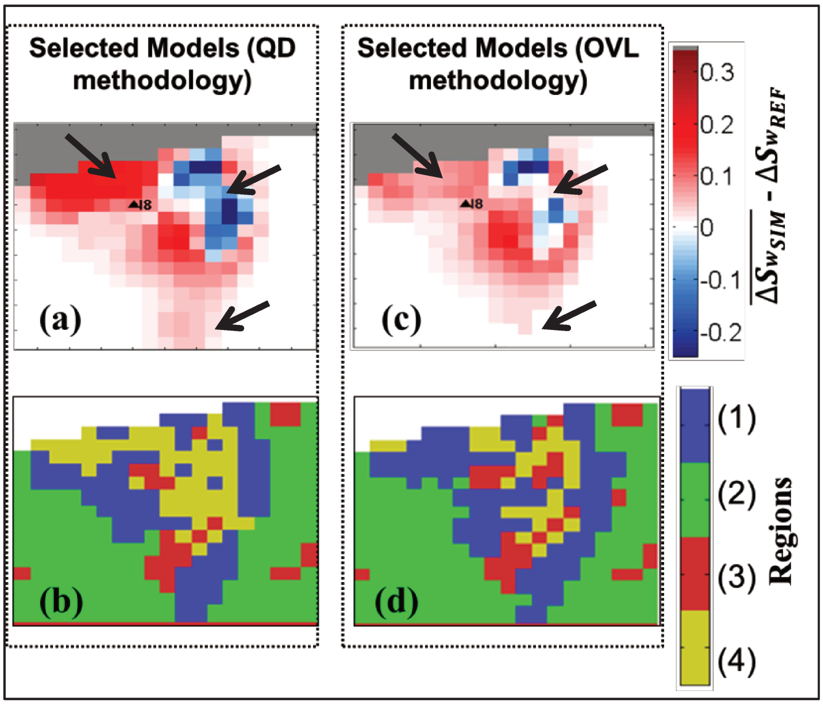

Figure 11–(a) 50 selected ΔS_w maps using from QD - reference value of ΔS_w, (b) 4 Regions Map of the 50 selected models using QD, (c) 50 selected ΔS_w maps using from OVL - reference value of ΔS_w, (b) 4 Regions Map of the 50 selected models using OVL (layer 3 out of 9).

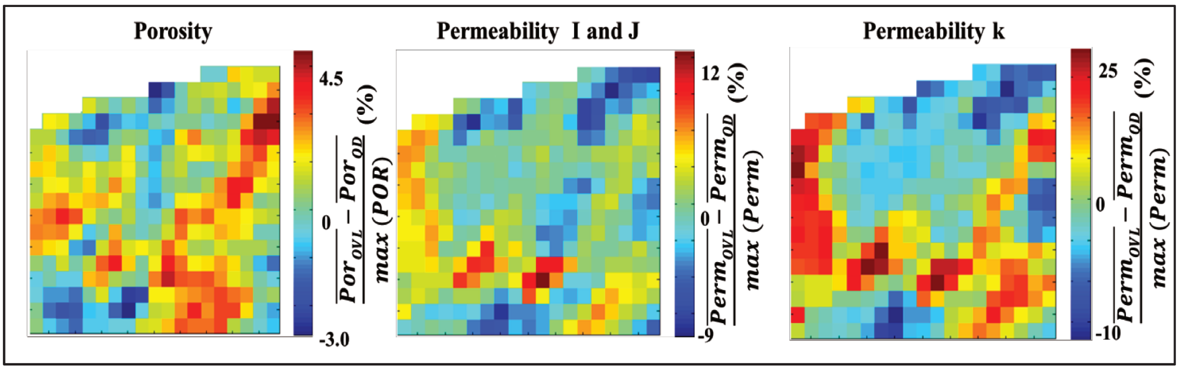

Figure 12–Porosity and permeability maps showing the normalized difference between OVL and QD (layer 3 out of 9).

6. Conclusions

This work presented a comparison of two methodologies that used different 4D seismic information (deterministic and probabilistic) as a constraint to select simulation models.

The methodologies were applied in the selection of the simulation models yielding different results. Regarding Δp maps, the selected models using OVL methodology used more efficiently 4D seismic data available than QD, yielding models properly calibrated in regions where seismic is more precise. The ΔS_w maps were analyzed locally: both methodologies showed valuable improvements to monitor the water front, however, the OVL methodology presented slightly better results.

It is interesting to highlight that QD and OVL methodologies had similar computational cost, even the OVL, which has a bigger data (once the uncertainties are considered). A future work will be conducted to compare the results from both methodologies when they are integrated in a history matching procedure. We expected more remarkable differences between QD and OVL, mainly regarding discretization levels of fault transmissibility and the porosity and permeability values.

7. Acknowledgements

This work was carried out in association with the ongoing Project registered as "BG-07 Redução de incertezas através da incorporação de sísmica 4D na modelagem do reservatório" (UNICAMP/BG Brazil/ANP) funded by BG Brazil under the ANP R&D levy as "Compromisso de Investimentos com Pesquisa e Desenvolvimento". The authors would like to thank UNISIM, DE-FEM-UNICAMP, CEPETRO and CAPES for supporting this work. We also thank CMG and Schlumberger for software licenses.

8. References

Almeida, F. L., Davolio, A., and Schiozer, D. J. (2014). A New Approach to Perform a Probabilistic and Multi-objective History Matching. *SPE Annual Technical Conference and Exhibition* . Amsterdam, The Netherlands.

Assunção, G. S., Davolio, A., and Schiozer, D. J. (2015). A Methodology to Integrate Multiple Simulation Models and 4D Seismic Data Considering Their Uncertainties. *SPE Annual Technical Conference and Exhibition* . Dubai, U.A.E.

Bosch, M., Mukerji, T., and Gonzalez, E. F. (2010). Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics, 75*(5).

Castro, S. A. (2007). A PROBABILISTIC APPROACH TO JOINTLY INTEGRATE 3D/4D SEISMIC, PRODUCTION DATA AND GEOLOGICAL INFORMATION FOR BUILDING RESERVOIR MODELS. *Tese de Doutorado*.

Davolio, A., Maschio, C., and Schiozer, D. J. (2013). A methodology to constrain pressure and saturation estimation from 4D seismic using multiple simulation models and observed data. *Journal of Petroleum Science and Engineering*, 51-61.

Davolio, A., Maschio, C., and Schiozer, D. J. (2013). A Methodology to Constrain Pressure and Saturation Estimation from 4D Seismic using Multiple Simulation Models and Observed Data. *Journal of Petroleum Science and Engineering*, 51-61.

Emerick, A. A., and Reynolds, A. C. (2012). History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 639-659.

Fahimudin, A. (2010). 4D Seismic History Matching Using the Ensemble Kalman Filter (EnKF):Possibilities and Challenges. *PhD Thesis, Department of Mathematics, University of Bergen*.

Kelleher, L. (2014). History Matching Using 4D Seismic. *Department of Earth Science, Imperial College of London*.

Landa, J. L., and Kumar, D. (2011). Joint Inversion of 4D Seismic and Production Data. *SPE Annual Technical Conference and Exhibition* . Denver, Colorado, U.S.A.

Riazi, N., Lines, L., and Russell, B. (2013). Integration of time-lapse seismic analysis with reservoir. *Integration geoConvention.*

Risso, V. F., and Schiozer, D. J. (2008). Utilização de sísmica 4D e de mapas de saturação no ajuste de histórico de reservatórios petrolíferos. *Revista Brasileira de Geociências, 38*, 172-187.

Stephen, K. D., MacBeth, C., and Christie, M. (2005). Multiple model seismic and production history matching: a case study. *SPE EUROPEC/EAGE Ann. Conf. Exhibition.* Madrid, Spain.

Xavier, C., dos Santos, E., Vieira, V. d., and dos Santos, R. W. (2013). Genetic Algorithm for the History Matching Problem. *Procedia Computer Science*, 946-955.

Yan, T. (2014). History Matching of Production and 4D Seismic Data. *PhD thesis, Norwegian University of Technology and Science*.

4. CONCLUSIONS AND FUTURE WORKS

The current work presented a methodology to compare probabilistic dynamic changes maps (water saturation and pressure change maps) from 4D seismic and reservoir simulation models. The proposed method was fully described in Chapter 2 of this dissertation.

The main contribution of this work was to generate a diagnostic tool, which measures the quality of available seismic and simulation data in every reservoir grid block, classifying them in four possible regions: (1) both data in agreement, (2) simulation models data more precise than 4D seismic data, (3) both data in disagreement and (4) 4D seismic data more precise than simulation data. This tool is a way to understand the uncertainties and to guide the integration between probabilistic seismic and simulation data.

Moreover, in the results section from Chapter 2, it was shown that the information from the methodology can be useful to select the most representative dynamic maps from simulation using seismic information, when the latter was more precise than the former: from the 500 initial simulation maps of dynamic changes, the 50 (10%) selected maps presented more accurate results than the initial ones. To measure the accuracy of the initial and final maps it was used a reference model, which represents the expected (true) response.

An application of the methodology to select the most representative seismic and simulated maps of dynamic changes was also tested in Appendix B, where different combinations of seismic and simulation data are studied. The main contribution of this section was to validate the methodology as a diagnostic tool and also to evaluate its application to identify in the selection of the most precise pressure change maps from seismic and simulation.

In Chapter 3, a comparative study was performed to observe advantages and drawbacks between a traditional way to incorporate seismic data into reservoir modelling workflow and the proposed methodology. The main idea was compare the differences found in selected simulation models when 4D seismic information was used. This study was carried out separately for pressure and water saturation changes maps, presenting as main results:

- Δp maps: selected models using the methodology proposed in Chapter 2 presented different information form the selected models using the traditional methodology. Both methodologies reduced the differences between simulated data and expected response (from reference model), generating more accurate results than initial models. However, the proposed methodology used more efficiently the 4D seismic information, since after the selection performed, simulation and seismic data presented more regions in agreement than before. Selected models from traditional methodology, nonetheless, still presented reservoir locations where seismic could be used, even after the selection of the simulation models using available seismic data;

- ΔS_w maps: both methodologies showed valuable improvements to monitor the water front, however, the proposed methodology presented slightly better results, when compared with the expected response, than traditional method used.

Furthermore, the differences in transmissibility, porosity and permeability distributions between selected models from traditional and proposed methodologies were observed. Again, different types of information were observed, highlighting a difference of 25% in some parts of the vertical permeability maps. The results from Chapter 3 is a indicative that by using the proposed methodology one can reach better calibrated models than by applying the traditional approach.

An additional statistical study was presented in Appendix A. The main contribution of this section was to compare statistically, three methods that could be used to calculate the probability distributions from seismic and simulation data. As presented in Chapter 2, the first step in the proposed methodology was the probability estimation of the data and it highly depends on the method used, as shown in Appendix A. Thus, depending on the method used, the methodology from Chapter 2 could bring different outputs. It was developed a methodology to identify which statistical method was the most accurate one and kernel density estimator (KDE) was more accurate than the other two studied methods, namely, histogram and empirical cumulative frequency curve.

The following next steps are proposed to complement the presented work:

- Application of the proposed methodology to real datasets, that is, a more challenging one;
- Performing the proposed methodology with a history matching procedure integrated. In this case, the diagnosis map can be used to guide the parameterization and the definition of the seismic objective function (defining regions to match seismic data, for instance). Also, this map can be used to access the quality of matching in each history matching iteration, providing a visual tool that brings spatial information not only about the error between the two data but also their variability.

REFERENCES

Almeida, F. L., Davolio, A., & Schiozer, D. J. (2014). A New Approach to Perform a Probabilistic and Multi-objective History Matching. *SPE Annual Technical Conference and Exhibition.* Amsterdan, The Netherlands.

Al-Saleh, M. F., & Samawi, H. M. (2007). Inference on Overlapping Coefficients in Two Exponential Populations. *Journal of Modern Applied Statistical Methods*, 503- 516.

Anderson, G., Linton, O., & Whang, Y.-J. (2012). Nonparametric estimation and inference about the overlap of two distributions. *Journal of Econometrics*, 1-23.

Assunção, G. S., Davolio, A., & Schiozer, D. J. (2016). A COMPARATIVE STUDY OF TWO METHODOLOGIES TO INTEGRATE RESERVOIR SIMULATION AND 4D SEISMIC DATA. *Rio Oil & Gas.* Rio de Janeiro.

Assunção, G. S., Davolio, A., & Schiozer, D. J. (2016). A Methodology to Integrate Multiple Simulation Models and 4D Seismic Data Considering Their Uncertainties. *SPE Annual Technical Conference and Exhibition* . Dubai, U.A.E.

Behrebs, R., Condon, P., Haworth, W., Bergeron, M., Wang, Z., & Ecker, C. (2001). 4D Seismic Monitoring of Water Influx at Bay Marchand: The Practical Use of 4D in an Imperfect World. *SPE Annual Technical Conference and Exhibition.* Orleans, Louisiana.

Bosch, M., Mukerji, T., & Gonzalez, E. F. (2010). Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics, 75*(5).

Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). KERNEL DENSITY ESTIMATION VIA DIFFUSION. *The Annals of Statistics, 38*(5), pp. 2916-2957.

Bradley, E. L. (2006). Overlapping Coefficient. *Encyclopedia of Statistical Science*.

Brown, A. (1991). Interpretation of 3D Seisimic Data.

Castro, S. A. (2007). A PROBABILISTIC APPROACH TO JOINTLY INTEGRATE 3D/4D SEISMIC, PRODUCTION DATA AND GEOLOGICAL INFORMATION FOR BUILDING RESERVOIR MODELS. *PhD Thesis*.

Castro, S. A., Caers, J., Otterlei, C. M., Meisingset, H., Hoye, T., Gomel, P., et al. (2009). Incorporating seismic data into reservoir models while honoring production and geologic data: A case study. *The Leading Edge*(28(12)), 1498-1506.

Correia, G. F., Davolio, A., & Schiozer, D. J. (2016). Multi-Scale Integration of 4D Seismic and Simulation Data to Improve Saturation Estimations. *SPE EUROPEC.* Vienna, Austria.

Correia, G., Davolio, A., & Schiozer, D. J. (2016). Multi-Scale Integration of 4D Seismic and Simulation Data to Improve Saturation Estimations. *SPE EUROPEC.* Vienna, Austria.

D'Agostinho, R. B., & Stephens, M. (1986). *Goodness-of-fit-tecnhiques (Statistics: a Series of Textbooks and Monographs)* (Vol. 68). Dekker.

Davolio, A., & Schiozer, D. J. (2012). Pressure and Saturation Estimation from P ans S Impedance: A Theorical Study. *Journal of Geophysics and Engineering, 9*, 447-460.

Davolio, A., & Schiozer, D. J. (2015). Joining Two Probabilistic Approaches - 4D Seismic Data and Simulation Models. *SPE EAGE Conference & Exhibition* . Madrid, Spain.

Davolio, A., Maschio, C., & Schiozer, D. J. (2013). A Methodology to Constrain Pressure and Saturation Estimation from 4D Seismic using Multiple Simulation Models and Observed Data. *Journal of Petroleum Science and Engineering*, 51-61.

Davolio, A., Maschio, C., & Schiozer, D. J. (2013). Local History Matching Using 4D Seismic Data and Multiple Models Combination. *EAGE Anual Conference &*

Exhibition . London, United Kingdom.
Davolio, A., Maschio, C., & Schiozer, D. J. (2014). A methodology to calibrate water saturation estimated from 4D seismic. *Journal of Geophysics and Engineering, 11*, 1-16.
de Brito, D., de Moraes, R., & Emerick, A. A. (2010). The Marlim Field: Incorporating Time-Lapse Seismic in the Assisted History Matching. *SPE Latin American & Caribbean Petroleum Engineering Conference.* Lima, Peru.
Elsheikh, A. H., Jackson, M. D., & Laforce, T. C. (2012). Bayesian Reservoir History Matching Considering Model and Properties Uncertainties. *Mathematical Geosciences*, pp. 515-543.
Emerick, A. A. (2016). Analysis of the performance of ensemble-based assimilation. *Journal of Petroleum Science and Engineering*, pp. 216-239.
Emerick, A. A., & Reynolds, A. C. (2012). History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 639-659.
Emerick, A. A., & Reynolds, A. C. (2013). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, 3-15.
Evensen, G. (2002). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*.
Evensen, G., Hove, J., Meisingset, H. C., Reiso, E., Seim, K. S., & Espelid, Ø. (2007). Using the EnKF for Assisted History Matching of a North Sea Reservoir Model. *SPE Reservoir Simulation Symposium.* Woodlands, Texas.
Fahimudin, A. (2010). 4D Seismic History Matching Using the Ensemble Kalman Filter (EnKF):Possibilities and Challenges. *PhD Thesis, Department of Mathematics, University of Bergen*.
Freedman, D., & Diaconis, P. (1981). On this histogram as a density estimator: L2 theory. *Zeit. Wahr. ver. Geb, 57*, 453–476.
Gervais, V., Roggero, F. (2010). Integration of saturation data in a history matching process based on adaptive local parametrization. Journal of Petroleum Science Engineering 73, 86–98.
Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric Statistical Inference.* Springer.
Gosselinh, O., Aanonsen, S. I., Aavatsmark, I., Cominelli, A., Gonard, R., Kolasinski, M., et al. (2003). History Matching Using Time-lapse Seismic (HUTS). *SPE Annual Technical Conference and Exhibition.* Denver, Colorado.
Grana, D., & Mukerji, T. (2014). Bayesian inversion of time-lapse seismic data for estimation of static reservoir properties and dynamic property changes. *Geophysical Prospecting*.
Hansen, B. E. (2009). Nonparametric Statistics. *Lectures Notes on Nonparametrics*. University of Wisconsin.
Härdle, W. (1994). *Applied Nonparametric Regression.* New York: Cambridge University Press.
Ida, M. (2009). Incorporação Quantitativa de Dados de Sísmica 4D no Processo de Ajuste de Histórico. *Master's Dissertation*.
Jin, L., van den Hoek, P. J., Alpak, F. O., Pirmez, C., Fehintola, T., Tendo, F., & Olaniyan, E. E. (2012). A Comparison of Stochastic Data-Integration Algorithms for the Joint History Matching of Production and Time-Lapse-Seismic Data. *Society of Petroleum Engineers.* doi:10.2118/146418-PA
Kelleher, L. (2014). History Matching Using 4D Seismic. *Department of Earth Science, Imperial College of London*.
Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *Journal of Economic*

Perspectives, 143–156.
Landa, J. L., & Kumar, D. (2011). Joint Inversion of 4D Seismic and Production Data. *SPE Annual Technical Conference and Exhibition* . Denver, Colorado, U.S.A.
Maschio, C., & Schiozer, D. J. (2013). A New Procedure to Reduce Uncertainties in Reservoir Models using Statistical Inference and Observed Data. *Journal of Petroleum Science and Engineering*, pp. 7-21.
Maschio, C., & Schiozer, D. J. (2014). Bayesian history matching using artificial neural network and Markov Chain Monte Carlo. *Journal of Petroleum Science and Engineering, 123*, pp. 62-71.
Maschio, C.; Schiozer, D. J (2016). Probabilistic history matching using discrete Latin Hypercube sampling and nonparametric density estimation, *Journal of Petroleum Science and Engineering*, v. 147, pp. 98-115.
Maschio, C., Vidal, A. C., & Schiozer, D. J. (2008). A framework to integrate history matching and geostatistical modeling using genetic algorithm and direct search methods. *Journal of Petroleum Science & Engineering*, pp. 34-42.
Mizuno, S., Yamaguichi, T., Fukushima, A., & Ohashi, Y. (2005). Overlap coefficient for assessing the similarity of pharmacokinetic data between ethnically different populations. *Clinical Trials*.
Olivier, D. S., & Chen, Y. (2011). Recent Progress on Reservoir History Matching: A Review. *Computacional Geosciences, 15*, 185-221.
Pazetti, B. A., Davolio, A., & Schiozer, D. J. (2015). The Impact of Rock and Fluid Uncertainties in the Estimation of Saturation and Pressure from a 4D Petro Elastic Inversion. *Journal of Geophysics and Engineering,*, 686-701.
Peracchi, F. (2014). *Applied Nonparametric Methods.* Lectures Notes. University of Rome Tor Vergata and EIEF.
Portella, R. C., & Emerick, A. A. (2005). Use of Quantitative 4D-Seismic Data in Automatic History Match. *SPE Latin American and Caribbean Petroleum Engineering Conference.* Rio de Janeiro, Brazil.
Raykar, V. C. (2002). *Probability Density Function Estimation by different Methods Vikas Chandrakant Raykar.* Report written for ENEE 739Q SPRING 2002.
Riazi, N., Lines, L., & Russell, B. (2013). Integration of time-lapse seismic analysis with reservoir simulation. *Integration geoConvention.* Canada.
Risso, V. F., & Schiozer, D. J. (2008). Utilização de sísmica 4D e de mapas de saturação no ajuste de histórico de reservatórios petrolíferos. *Revista Brasileira de Geociências, 38*, 172-187.
Roggero, F., Lerat, O., Ding, D. Y., Berthet, P., & Bordenave, C. (2012). History Matching of Production and 4D Seismic Data: Application to the Girassol Field, Offshore Angola. *Oil & Gas Science and Technology - Revue d'IFP Energies nouvelles, Institut Français du Pétrole*, pp. 237-262.
Romero, C.E., Carter, J.N., Gringarten, A.C., Zimmerman, R.W.: A modified genetic algorithm for reservoir characterisation. In: Int. Oil and Gas Conf. and Exhibition. Beijing, China (2000).
Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Annals of Mathematical Statistics, 27*, 832–837.
Sagitov, I., & Stephen, K. D. (2012). Assisted Seismic History Matching in Different Domains: What Seismic Data Should We Compare? *SPE Europec/EAGE Annual Conference.* Copenhagen, Denmark.
Schiozer, D. J., Avansi, G. D., & Santos, A. A. (2014). Methodology for Risk Quantification Combining Geoestatistical Realizations and Discretized Latin Hypercube . *Computers*

& Geosciences.

Schmid, F., & Schmidt, A. (2006). Nonparametric Estimation of the Coefficient of Overlapping - Theory and Empirical Application. *Computational Statistics & Data Analysis*, 1583-1596.

Shalizi, C. (2009). *Estimating Distributions and Densities.* Lecture Notes on Data Mining, Statistics Department Carnegie Mellon University.

Shi, P., Lin, Y.-A., Pastuska, M., Cui, H., & Mackay, J. (2014). Triggered Sorting and Co-Assembly of Generically Enfineered Protein Microdomains in Cytoplasm. *Advanced material*, 449-454.

Silverman, B. W. (1986). *DENSITY ESTIMATION FOR STATISTICS AND DATA ANALYSIS.* London: Chapman and Hall.

Stephen, K. D., MacBeth, C., & Christie, M. (2005). Multiple model seismic and production history matching: a case study. *SPE EUROPEC/EAGE Ann. Conf. Exhibition.* Madrid, Spain.

Stephen, K. D., Soldo, J., Macbeth, C., & Christie, M. (2004). A MULTIPLE MODEL APPROACH TO HISTORY MATCHING AND UNCERTAINTY ANALYSIS USING TIME-LAPSE SEISMIC. *9th European Conference on the Mathematics of Oil Recovery.* Cannes, France.

Skjervheim, J.-A., Evensen, G., Hove, J. & Vabo, J. G. (2011). An ensemble smoother for assisted history matching. *SPE RESERVOIR SIMULATION SYMPOSITION*. The Woodlans, Texas, USA.

Suzuki, S. (2007). Integrated Evaluation of Structural Uncertainty Using History Matching from Seismic Imaging Uncertainty Model, *PhD Thesis*, Stanford University.

Takada, T. (2001). Nonparametric density estimation: A comparative study. *Economics Bulletin*, 1-10.

Tian, S., Shams, A., & MacBeth, C. (2014). Closing the loops using engineering-consistent 4D seismic inversion. *The Leading Edge, 33*, 188-196.

van Leeuwen, P. J., & Evensen, G. (1996). Data Assimilation and Inverse Methods in Terms of Probabilistic Formulation. *Mon. Weather Rev.*, pp. 2898-2913.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S.* Springer.

Wasserman, L. (2005). *All of nonparametric statistics.* Springer.

Weitzman, M. (1970). *Measure of the overlap of income distribution of while and Negro families in the United States.* U. S. Department of Commerce, Bureau of the Census,, Washington, DC.

Xavier, C., dos Santos, E., Vieira, V. d., & dos Santos, R. W. (2013). Genetic Algorithm for the History Matching Problem. *Procedia Computer Science*, 946-955.

Yan, T. (2014). History Matching of Production and 4D Seismic Data. *PhD thesis, Norwegian University of Technology and Science*.

Yeh, T. -H., Jimenez, E., Van Essen, G., Chen, C., Jin, L., Girardi, A., et al. (2014). Reservoir Uncertainty Quantification Using Probabilistic History Matching Workflow. *SPE Annual Technical Conference and Exhibition.* Amsterdam, The Netherlands.

APPENDIX A

PAPER 3: "Quantitative Comparison of Non-Parametric Methods to Handle Probabilistic Data From 4D Seismic and Reservoir Simulation"

Germano S. C. Assunção, Alessandra Davolio, Denis José Schiozer.

To be submitted to a journal.

Quantitative Comparison of Non-Parametric Methods to Handle Probabilistic Data From 4D Seismic and Reservoir Simulation

Assunção, G.S.C, Davolio, A., Schiozer, D.J. – UNICAMP

A research paper submitted in partial fulfillment of the requirements for the Master of Science Degree with a major in Petroleum and Science Engineering.

To be submitted to a journal.

Abstract

4D seismic data is usually used as additional quantitative data to calibrate reservoir simulation models, providing information such as pressure and saturation changes. Traditionally, these data are used as deterministic observed data that the simulation models must match. However, due to the presence of noise and problems related to seismic acquisition and processing, interpreting seismic signals can be uncertain. Using probabilistic seismic analysis we can evaluate and mitigate these uncertainties.

Some recent works have shown that integration of probabilistic seismic data into probabilistic history-matching workflows can provide useful information for model calibration. However, it is still a difficult task, as we are processing uncertain information from two complex technologies, 4D seismic data and reservoir simulation.

Choosing the most accurate statistical method to estimate the probability density of the data sets is critical in some of these published works. Several methods estimate the probability density distribution of a certain data set, but all present some drawbacks. Thus, this work aims to evaluate the accuracy of three methods to estimate the distribution of a data set: histogram, empirical cumulative frequency curve, and kernel density estimator.

This study has two parts: (1) statistical analysis of the methods, using parametric curves (curves with known characteristics) to estimate which is the most accurate, and (2) evaluation of the performance of the three methods when using realistic simulations and seismic data.

The results show that, considering our application, the kernel density estimator is the most accurate, presenting the lowest error in both parts of the study.

1. Introduction

Reservoir simulation is an important tool for reservoir management because it simulates the behavior of a real field, allowing companies to test different production schemes among other activities. However, building reliable simulation models is difficult due to various limiting factors ranging from numerical errors (e.g. computational simplifications, scale transference) to uncertainties in reservoir characteristics (e.g. porosity and permeability distributions, transmissibility of the faults).

To reduce the uncertainties of the parameters that characterize the reservoir and consequently provide more reliable simulation results, it is common practice to use observed data (measured data in the field) as information that the simulation models must reproduce. For the models to honor observed data, the uncertain parameters are modified until the difference between the simulations and observed data reach an acceptable level. This procedure, history matching, can be done using a single simulation model (deterministic approach) or multiple simulation models (probabilistic approach).

Traditionally, deterministic history matching is used. However, probabilistic approaches have been further studied in recent years since it take into account several uncertainties, related to reservoir characterization, to be handled. Examples of probabilistic history-matching methodologies are presented in Maschio *et al.* (2008), Almeida, *et al.* (2014), and Yeh, *et al.* (2014).

The observed data used in deterministic or probabilistic history matching is usually acquired from production wells placed in sparse regions of the reservoir. The lack of areal (spatial) observed data can yield simulation models with low geological consistency, even when using probabilistic history-matching procedures. To address this issue, additional technologies are used, notably 4D seismic data. 4D seismic is the acquisition, processing and interpretation of 3D (or 2D) seismic surveys, at different times during reservoir production. When comparing the difference between these surveys, we can obtain a spatial view of the movement of fluids and the pressure variation in regions between the wells.

Seismic information is classically integrated as deterministic observed data that the simulation models must honor. However, seismic data also may contain uncertainties due to limitations in acquisition, processing, and interpretation. Therefore, once probabilistic approaches have good history-matching results, probabilistic studies can also be more effective to generate reliable 4D seismic data.

Although it seems reasonably to work with probabilistic approaches concerning the two data, simulation models and 4D seismic, the comparison and interpretation of the two probabilistic data sets is still challenging. A few works found in the literature integrated probabilistic seismic information into reservoir modeling workflow, examples of this can be found in Landa and Kumar (2011), Davolio and Schiozer (2015), and Assunção, *et al.* (2016).

Particularly, in the studies of Davolio and Schiozer (2015) and Assunção, *et al.* (2016), the accuracy of the statistical methods estimating the probability density [1] (from seismic and simulation data) was a main limitation. The former work used histograms to compare seismic and simulation information to create a “map of classes”. The information in this case was water saturation change (ΔS_w) and pressure variation (Δp). They showed that the “map of classes” could indicate which information was more useful, seismic or simulation, for every reservoir location. Assunção, *et al.* (2016) expanded the idea of the “map of classes”, developing a tool to compare the overlap of two curves (one from simulation and other from seismic data) to estimate the agreement between them, and evaluate whether 4D seismic or the simulation models are more precise for each reservoir location, and then use this information to select the most representative simulation models. In both works, the method used to estimate the probability density of the data sets was a critical part of the data analysis, significantly impacting the data interpretations. The work of Gibbons and Chakraborti (2011) presents several methods to estimate the probability density of a certain data set as well as the drawbacks. The selection of the numbers of bins, for instance, is the main limitation of the most used method, the histogram. Another two probability density methods commonly used are Empirical Cumulative Frequency Function (ECDF) and the Kernel Density Estimator (KDE). These methods also have limitations, namely the sensitivity to tail values for ECDF, and choosing the optimum bandwidth value in KDE.

Motivated by studies applied in economics, Takada (2001) compared methods to estimate the probability density of a data set, focusing on the robustness of the methods to estimate heavy tailed data. To evaluate the accuracy of the methods he used Hellinger error. Raykar (2002) developed a qualitative and quantitative comparison between the histogram and kernel based methods using the Kullback-Leibler distance to measure the accuracy of the methods.

As states Shalizi (2009), there are several statistical tools to measure the accuracy of a certain method. In addition to the Hellinger error and the Kullback-Leibler distance, we can

[1] In statistics, the density of a given data refers to the closeness/distance of the elements of this data.

use the mean-squared error (MSE), the total variation (or L_1), or the log-likelihood ratio. However, the choice of the most accurate method strongly depends on the statistical objects of interest, which can be: density functions, regression functions, variance functions, etc. (Hansen, 2009).

From the reservoir engineering perspective, more specifically comparing probabilistic 4D seismic and simulation models, we are interested in a method that generates accurate probability density functions (PDFs) and also generates curves that are easy to analyze, as we need to compare them to define which information, seismic data or simulation is more useful in each reservoir grid block (a reservoir model has thousands of blocks).

Thus, this paper aims to find which of the three methods, histogram, ECDF or KDE, is the most accurate to generate two probability density functions to be compared: one from 4D seismic data and the other from multiple simulation models.

The first part of this work statistically analyzes the parametric curves and the second part evaluates the performance of the three methods when applied to ΔS_w maps from reservoir simulation and 4D seismic data. We briefly summarize the three methods as well as the modified overlapping coefficient (the tool to compare the methods) before presenting the methodology.

2. Literature Review

The probability density function (PDF) describes the probability distribution of a data set. It is a nonnegative function, and the integral over it is equal to one.

To estimate probability density we use information from an observed data set to create a PDF. The several methods to estimate the density of a certain data set are split between two classes:

(1) Parametric: assumes that data are drawn from a known parametric family of distributions, for example the normal distribution with mean μ and standard deviation σ.
(2) Non-parametric: refers to statistical models that do not have data structures or characteristic parameters. These models do not make assumptions about the data distribution. Examples of non-parametric estimators are: histograms, the ECDF, and the KDE.

In this section we present the three non-parametric methods tested in this work: histogram, ECDF and KDE. Other available methods and further explanations about parametric and non-parametric methods are found in Venables and Ripley (2002) and Wasserman (2005).

We also provide more information about the modified overlapping coefficient (OVL), the tool we use to compare the three methods, in Section 2.4.

2.1 Histogram

Histograms are the most used density estimator. It is a graph that shows the probability distribution of data within certain ranges (bins).

The first step in building a histogram is to identify the occurrence of the studied variable, e.g.: the water saturation variation estimates are only defined between zero and one. The second step is to divide the data into intervals (bins), and then count how many estimates are in each interval. The bins cannot overlap, they must be adjacent and with the same size.

Choosing the number of bins is the main limitation when modeling histograms. Different data trends can be observed, for the same data, if different numbers of bins are chosen. Figure 1 presents an example.

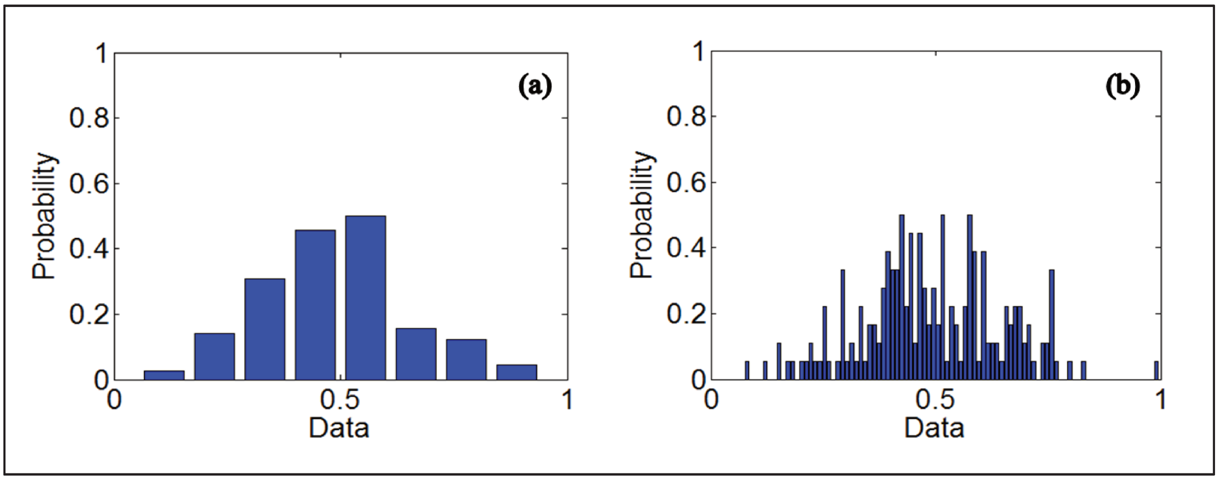

Figure 1–Influence of the numbers of bins in representing the same data using histogram: (a) 10 bins and (b) 100 bins.

To mitigate problems caused by erroneous choice of the number of bins, various methods have been developed to estimate the optimal number of bins. For data with distribution into a certain interval, there are classic theories to estimate the optimal bin size. One of the most commonly used is presented in Equation (1), originally presented by Freedman and Diaconis (1981).

The Freedman-Diaconis rule works well in practice and is immune to outliers (Venables and Ripley, 2002). The rule considered the deviation as well as the size of the sample, as presented in Equation (1).

$$h = 2 * \frac{IQR}{n^{1/3}}, \tag{1}$$

where h and n are bin width and sample size, respectively. IQR is the interquartile range, a measure of dispersion, where 50% of the elements in the sample center are within the IQR (after sorting into ascending order).

The value of the bin width can be used to estimate the number of bins k, as Equation (2) shows.

$$k = \frac{\max(X) - \min(X)}{h}, \tag{2}$$

where $\max(\mathrm{X})$ is the maximum value of the sample X and $\min(\mathrm{X})$ is the minimum value.

In the histogram-based method, Freedman-Diaconis's rule identifies the optimum number of bins. However, the data set is still divided into bins, which is a limitation. Section 2.2 presents the ECDF method, which does not use bins.

2.2 Empirical cumulative frequency curve (ECDF)

The ECDF describes the cumulative probability lower than or equal to a given element of a finite sample. Thus, considering a sample $\mathrm{x} = \{\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{N}\}$, the ECDF is calculated as follows:

$$ECDF(\mathrm{x}) = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{I}(\mathrm{x}_\mathrm{i} \leq \mathrm{p}), \tag{3}$$

where:

$$\mathrm{I}(\mathrm{x_i} \leq \mathrm{h}) = \begin{cases} 1, when\ \mathrm{x_i} \leq \mathrm{p} \\ 0, when\ \mathrm{x_i} > \mathrm{p} \end{cases}.$$

where p is a value defined within the sample x, being a step function which varies from $1/\mathrm{n}$ at each step until the limit n.

In probability theory and statistics, the cumulative distribution function (CDF) is defined as the probability that a real-valued random variable x_i will take a value less than or equal to p. Therefore, instead of using the PDF to analyze the probability of some data, we could use the CDF. Figure 2a presents an example of a CDF and a PDF yielded from the same data.

The CDF is a continuous function, while the ECDF is a discrete curve that shows the cumulative frequency based only on the sample available. Figure 2b illustrates an example in which a specific sample imitates the PDF and CDF from Figure 2a. Likewise, a histogram can be used to represent the PDF of this discrete data set, while the ECDF represents the CDF of this data set (Figure 2b). Remembering that histograms can also be used to generate the CDF of data, however, the main advantage of the ECDF compared to histograms is that it is unnecessary to define the number of bins.

Although modeling parameters are unnecessary, such as the number of bins, sometimes ECDF does not translate the data into a probability density very well. As stated by D'Agostinho and Stephens (1986), ECDF are sensitive to random occurrences in the data. The density of points in the extreme quantiles[2] is especially important to capture the tail behavior. The available information in the extreme quantiles can generate misleading conclusions. Thus, we studied another method, the KDE, explained in detail in Section 2.3.

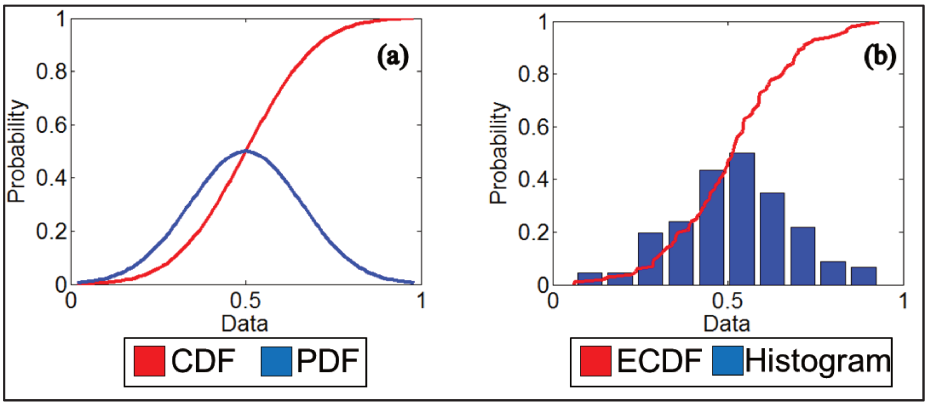

Figure 2–(a) Example of PDF and CDF from a data and (b) difference between the ECDF and the histogram of a sample.

2.3 Kernel Density Estimator (KDE)

The kernel estimator can be viewed as a generalization of the histogram. Also called KDE, abbreviation of Kernel Density Estimator, it is a non-parametric way to estimate a PDF of a random variable. This estimator uses kernel functions to estimate the PDF and was initially proposed by Rosenblatt (1956).

One advantage is a smooth density everywhere, unlike histograms, and we can even use them to estimate the derivatives of the density, if necessary.

[2]Quantiles are cut points dividing the range of a probability distribution into contiguous intervals with equal probabilities, or dividing the observations in a sample in the same way.

For a given sample $x = \{x_1, \dots, x_N\}$, with an unknown probability distribution, the kernel estimator estimates the $PDF(x)$, replacing the kernel function $k(x, X_i; t)$ in every element of the sample. As shown in Equation (4), the $\widehat{PDF}(x)$ from KDE is the sum of each kernel function estimates.

$$\widehat{PDF}(x) = \frac{1}{n} \sum_{i=1}^{n} k(x, X_i; t) \quad x \in \boldsymbol{R}, \tag{4}$$

where $\widehat{PDF}(x)$ is the estimated density function, n the sample size and $k(x, X_i; t)$ is the kernel function. The kernel function is defined by Equation (5)

$$k(x, X_i; t) = \frac{1}{\sqrt{2\pi t}} e^{-(x-Xi)^2/(2t)}, \tag{5}$$

The kernel function uses the local value (X_i) and the bandwidth parameter t to estimate the $\widehat{PDF}(x)$. The parameter t is similar to the bin width (h), in the histogram-based analysis.

Figure 3 shows an example of a density estimate using KDE and Gaussian kernel function. The most commonly used kernel functions are: Gaussian, Epanechnikov and triangular. We chose the Gaussian kernel function for this work due to the characteristics of data used, but this could change for other applications.

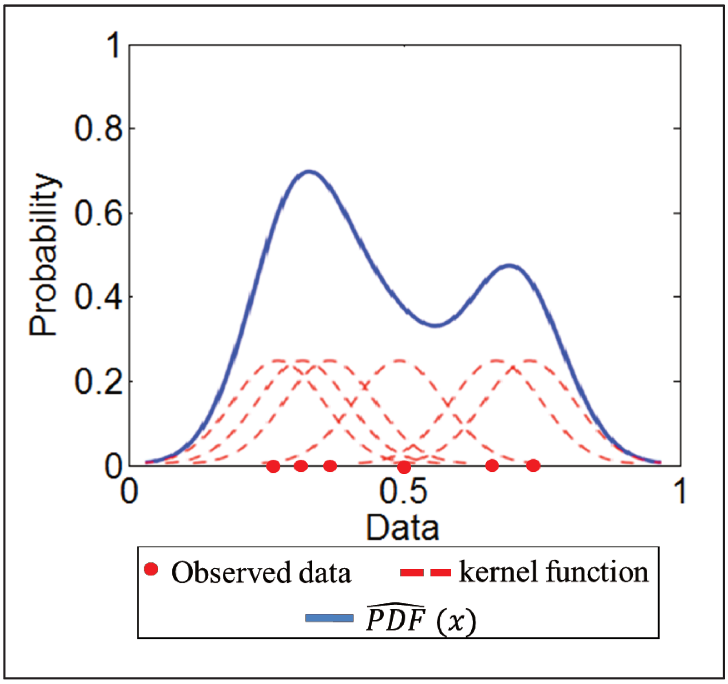

Figure 3–Kernel functions from an observed data and the estimated PDF using these kernel functions.

The main challenge in generating the $\widehat{PDF}(x)$ from kernel functions is to smooth local data, where inferences about the population are made based on neighborhood information. Based on the kernel function attempt, the estimations find the optimal value of t, analogous to the optimum parameter h in histogram analysis.

However, most of the kernel estimators have some border bias that is not always realistic, for example, negative estimates for variables that are known to be strictly positive.

Botev *et al.* (2010) presents an alternative to the usual method of estimating kernel functions, where prior knowledge of the data is taken into account. Thus, for instance, if we know that a data set is nonnegative, this is taken into account, resulting in a more consistent estimate of the data. Equation (6) shows the kernel estimator proposed by Botev *et al.* (2010):

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} k_b(x, X_i\,; t), \tag{6}$$

where the kernel function $k_b(x, X_i; t)$ is defined by Equation (7).

$$k_b\,(x, X_i; t) = \frac{1}{n}\sum_{k=-\infty}^{\infty} k(x, 2k + X_i; t) + k(x, 2k - X_i; t), \tag{7}$$

Sections 2.1, 2.2 and 2.3 show the application of the methods used in this work, as well as the benefits and drawbacks of each one. The next section details the parameter OVL, the tool used in the present work to compare the three methods.

2.4. Modified overlapping coefficient (OVL)

PDF_{SIM} and PDF_{SEIS} denote the probability distribution functions of two vectors, SIM and SEIS. These vectors represent, for instance, saturation variation estimated from multiple simulation models (SIM) and 4D seismic data (SEIS), respectively: SIM = $\{\Delta S_{w1}, \Delta S_{w2}, \ldots \Delta S_{wm}\}$ and SEIS = $\{\Delta S_{w1}, \Delta S_{w2}, \ldots \Delta S_{wn}\}$.

The classic coefficient of overlapping, in this work called OVLC, is defined according to Weitzman (1970) apud Schmid & Schmidt (2006) in Equation (8) and can be used to identify the similarity between PDF_{SIM} and $PDF_{SEIS.}$

$$\mathrm{OVLC} = \int_0^1 \min\{\mathrm{PDF_{SIM}}; \mathrm{PDF_{SEIS}}\}\mathrm{dx}, \tag{8}$$

The OVLC is a classic tool to compare two data sets, commonly used in reliability analysis (Al-Saleh and Samawi, 2007).

Based on OVLC, Assunção *et al.* (2016) proposed the modified overlapping coefficient (OVL), which generates two parameters, named OVL_{SIM} and OVL_{SEIS}. Instead of only identifying similarities of the data, OVL_{SIM} and OVL_{SEIS} also show which data are more precise[4], SIM or SEIS.

In this case, after calculating the OVLC, we identify the OVLC interval, as presented in Figure 4a. The OVLC interval is the coincidence interval between the two PDFs. From the OVLC interval, we calculate the proportion of each PDF within it.

The proportion of PDF_{SIM} within the OVLC interval is called OVL_{SIM}, and is calculated by Equation (9). Figure 4b illustrates this parameter.

$$\mathrm{OVL_{SIM}} = \frac{\text{Area under } \mathrm{PDF_{SIM}} \text{ within OVL interval}}{\text{Overall area under } \mathrm{PDF_{SIM}}}, \tag{9}$$

OVL_{SEIS} represents the proportion of PDF_{SEIS} within the OVLC interval, as shown in Equation (10). Figure 4c presents a visual example of the OVL_{SEIS}.

[4] Precision and accuracy are two different statistical indicators:

- Precision refers to the closeness of two or more measurements to each other. It is related to the variability of the data: high precision means low variability and vice versa.
- Accuracy refers to the closeness of a measured value to a standard or known value.

$$\mathrm{OVL_{SEIS}} = \frac{\text{Area under PDF}_{\text{SEIS}}\text{ within OVL interval}}{\text{Overall area under PDF}_{\text{SEIS}}}, \tag{10}$$

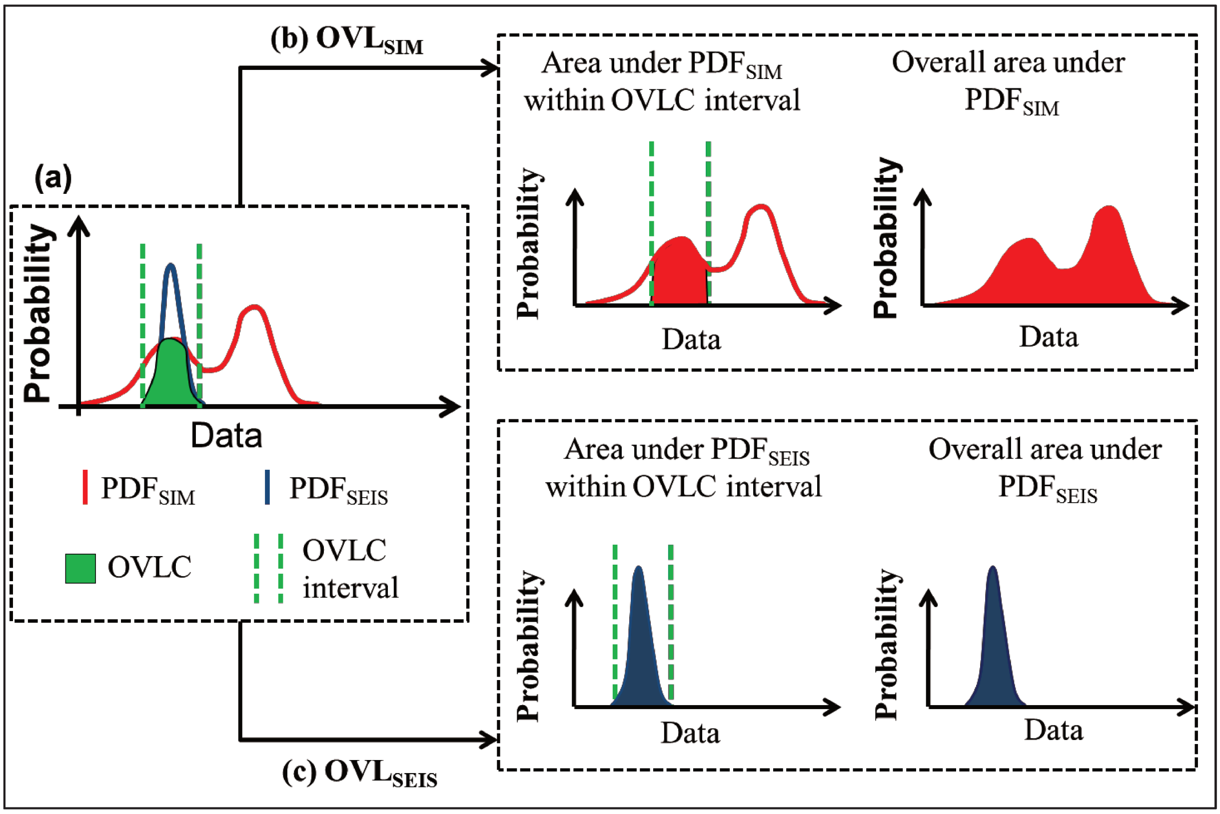

Figure 4–Identification of (a) OVLC interval, (b) $\mathrm{OVL_{SIM}}$ and (c) $\mathrm{OVL_{SEIS}}$.

Calculating $\mathrm{OVL_{SIM}}$ and $\mathrm{OVL_{SEIS}}$, we can identify three possible arrangements: (I) both PDFs have similar distributions, (II) both PDFs show different trends, and (III) one PDF is more precise than the other. Figure 5 shows examples of these arrangements.

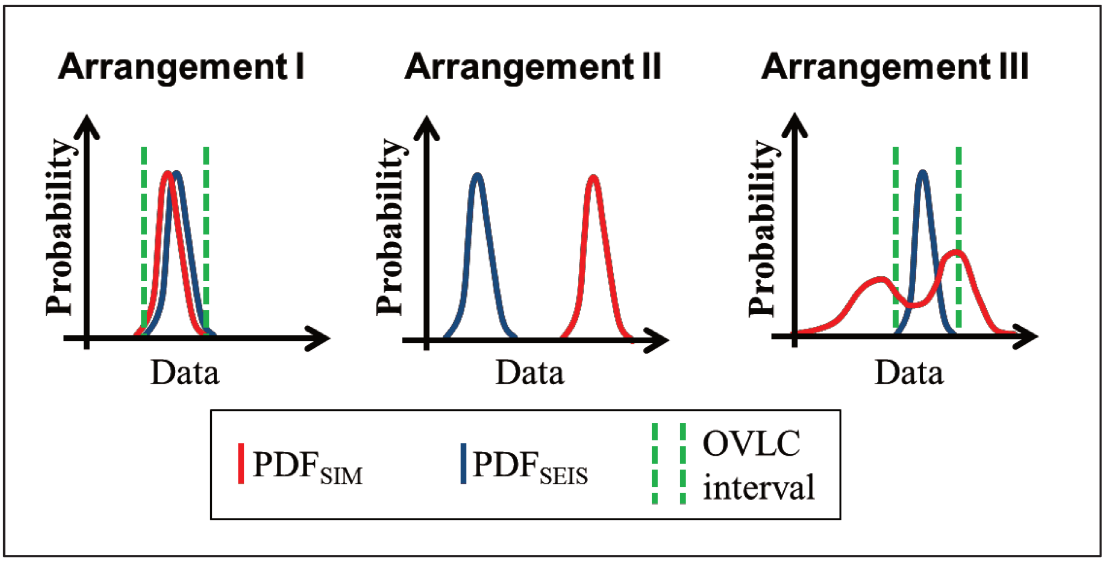

Figure 5–Example of the three possible arrangements obtained from comparison of two PDFs.

Table 1 presents the OVL_S cutoff values to define each arrangement. Thus, if both curves are more than 80% within the OVL interval, they are considered to be in agreement (Arrangement I). If both are less than 80% they are in disagreement (Arrangement II) and if one has an OVL greater than 80% and the other lower than 80%, the PDF with the greater OVL is more precise than the other (Arrangement III).

Table 1–Characteristics of the arrangements

Arrangement	OVL values
I	OVL_{SIM} > 80% and OVL_{SEIS} > 80%
II	OVL_{SIM} < 80% and OVL_{SEIS} < 80%
III	OVL_{SIM} > 80% and OVL_{SEIS} < 80% (or vice versa)

3. Methodology

In section 3.1, the main contribution of this work, we present the methodology to calculate the accuracy of the three methods using the OVL. This first part identifies the most appropriate PDF builder using synthetic parametric curves. In the second part, section 3.2, instead of using synthetic parametric curves, we used data obtained from reservoir simulation models and from a probabilistic synthetic seismic inversion.

3.1 Calculating the accuracy of three methods

- **Step 1: generation of parametric curves and calculation of the true OVLs**

First, we define the arrangement to be studied (Figure 5) and then we generate two parametric curves, one representing PDF_{SIM} and another, PDF_{SEIS}. These PDFs mimic probabilistic data that could be obtained from simulation models or 4D seismic data, more specifically, ΔS_w estimates as seen in Almeida *et al.* (2014), Davolio and Schiozer (2015), and Assunção *et al.* (2016).

We then calculate the true OVL_{SIM} and OVL_{SEIS}, using the parametric curves with well-defined characteristics. The OVLs are obtained using the trapezium rule to calculate the area under the curves and considering the three-sigma of the curve (which guarantees that all the values to be taken lie within three standard deviations of the mean).

- **Step 2: generating the discrete data sets**

Based on Latin hypercube sampling (LHS), we obtain the vectors SIM and SEIS, with the same size. These vectors must have the same behavior as the PDF_{SIM} and PDF_{SEIS} from the previous step and are the samples used to estimate the probability density distributions using the histogram, ECDF and KDE. More details about this sampling method can be found in

Note that several vectors can have the behavior of a specific parametric curve. Therefore, instead of considering only one vector SIM and another vector SEIS, y vectors of SIM and SEIS are used, all with the characteristic of the PDFs defined in the previous step. We took this precaution to avoid interference caused by the Latin hypercube technique itself, since a vector can represent the specific PDF but have values more concentrated around the mean of the data, while others vectors may not.

Therefore, vectors from SIM_1 to SIM_y mimic the PDF_{SIM} while vectors from $SEIS_1$ to $SEIS_y$ represent the PDF_{SEIS}. We highlight that every vector, SIM_y or $SEIS_y$ follows the specific PDF in the study, thus, for instance: SIM_1 replicates the studied PDF_{SIM}, and SIM_y

replicates the studied PDF_{SIM}. Figure 6a and 6b exemplify possible vectors SIM and SEIS displaying PDF_{SIM} and PDF_{SEIS} respectively. Two data sets are presented for SIM and SEIS (with 3 elements): SIM_1 and $SEIS_1$ (Figure 6a) present values close to the mean while SIM_2 and $SEIS_2$ in Figure 6b have values farther from the parametric curve mean. However, as shown in this example, both data sets ($SIM_1/SEIS_1$ and $SIM_2/SEIS_2$) can represent the parametric PDFs from step 1.

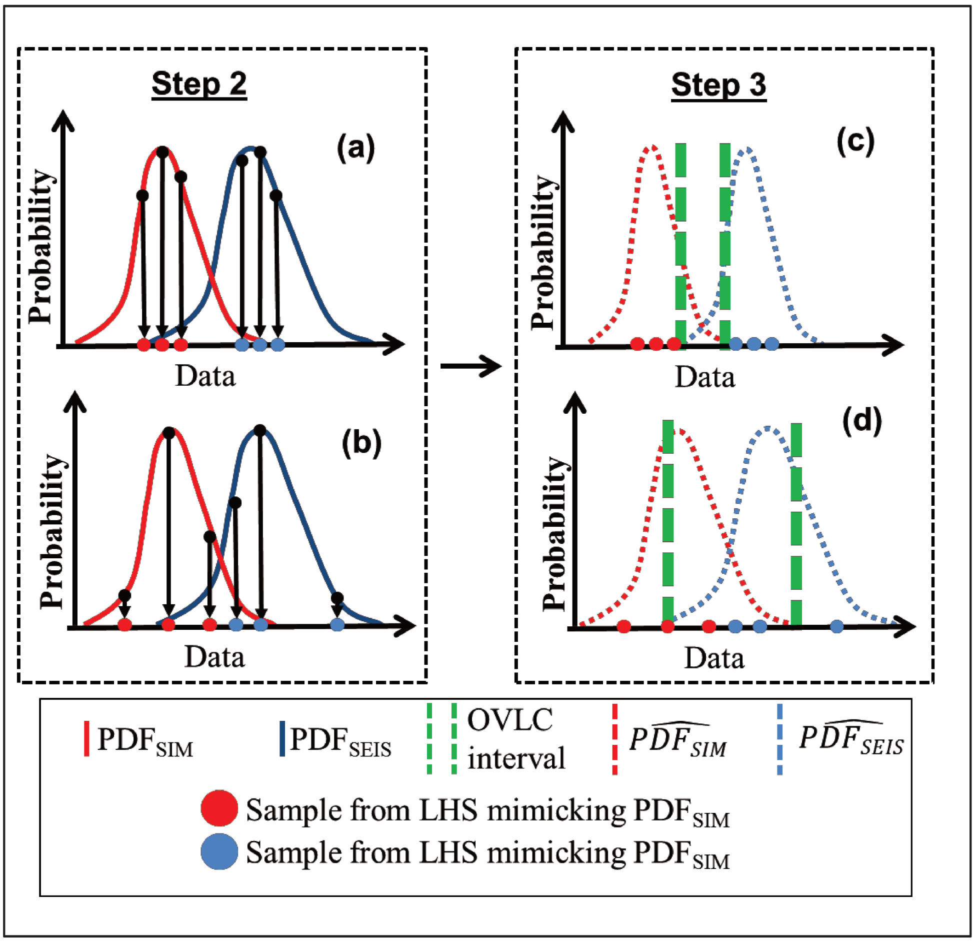

Figure 6–Steps of 2 and 3 of the methodology: (a) and (b) sampling examples; (c) and (d) example of PDFs and OVLC interval obtained from the sample using KDE.

- **Step 3: calculating OVL_{SIM} and OVL_{SEIS} using non-parametric estimators**

Using *y* combinations of SIM and SEIS ($SIM_1/SEIS_1$; $SIM_2/SEIS_2$...$SIM_y/SEIS_y$), we calculate the probability of the data using each of the three methods: histogram, ECDF and KDE. Figure 6c and 6d illustrate the estimates from KDE using the available SIM and SEIS. Then, we calculate the OVLs, according to Equations (9) and (10).

At the end of the procedure, we have OVL_{SIM} and OVL_{SEIS} estimated using each of the three methods for every SIM and SEIS studied, i.e.: $OVL_{SIMy\text{-}HIST}/OVL_{SEISy\text{-}HIST}$; $OVL_{SIMy\text{-}ECDF}/OVL_{SEISy\text{-}ECDF}$; $OVL_{SIMy\text{-}KDE}/OVL_{SEISy\text{-}KDE}$.

- **Step 4: quadratic difference between estimated OVLs and true values**

The estimated parameters OVL_{SIM} and OVL_{SEIS} are then compared with the true values, using the quadratic difference, as presented below in Equation (11) for OVL_{SIM}.

$$\mathrm{QD}_{\mathrm{SIM(METHOD)}} = \sum_{i=1}^{y} \left(OVL_{SIMy(METHOD)} - OVL_{SIM(\mathrm{TRUE})} \right)^2 , \tag{11}$$

where $METHOD$ represents the method studied: Histogram, ECDF or KDE and $TRUE$ is the true answer calculated in step 1, y is the vector used (SIM_1, SIM_2 or SIM_y).

The same procedure is repeated for OVL_{SEIS}, as shown below in Equation (12).

$$\mathrm{QD}_{\mathrm{SEIS(METHOD)}} = \sum_{i=1}^{y} \left(OVL_{SEISy(METHOD)} - OVL_{SEIS(\mathrm{TRUE})} \right)^2 , \tag{12}$$

where y is the vector used ($SEIS_1$, $SEIS_2$ or $SEIS_y$).

To compare estimates of QD_{SIM} yielded from each method, we propose the following normalization.

$$\text{Normalized QD}_{SIM(METHOD)} = \frac{QD_{SIM(METHOD)}}{\max\{QD_{SIM(Hist\,\&\,ECDF\,\&\,KDE)}\}} , \tag{13}$$

where: denominator is the value of the method presenting the greatest QD_{SIM}.

Normalized QD_{SEIS} is calculated as well, using Equation (14).

$$\text{Normalized QD}_{SEIS(METHOD)} = \frac{QD_{SEIS(METHOD)}}{\max\{QD_{SEIS(Hist\,\&\,ECDF\,\&\,KDE)}\}} , \tag{14}$$

At the end of the procedure, we select the method with the lowest Normalized QD_{SIM} and QD_{SEIS}, as the most accurate, to calculate the OVLs.

3.2. Comparing the three methods using realistic reservoir data

In section 3.1 we calculated the true value of the OVLs, and therefore we can now measure the accuracy of the three methods, since we have the real response from step 1. To prove the results from the previous section and apply the OVL methodology in realistic reservoir model, we follow this second section.

Instead of using a parametric curve as the starting point, in this section, the available information are *m* maps of ΔS_w obtained from *m* simulation models and *n* maps of ΔS_w yielded from probabilistic 4D seismic data. Figure 7 presents an example with $m = n = 3$ in a reservoir with 4 grid blocks. Every reservoir grid block provides the information necessary to estimate the probability of water saturation change from simulation and seismic data. Thus, using each of the three methods we can estimate the probability of the data sets. We then calculate the OVLs parameters for every reservoir grid block, to identify whether simulation or seismic data are more precise.

In comparison with the previous section (3.1), there are no steps 1 or 4, because there are no parametric curves to represent the real OVLs. This section evaluates how the three methods can affect the estimation of OVL_{SIM} and OVL_{SEIS} when realistic simulation and seismic data are used (although these data were obtained from a synthetic reservoir model). OVLs parameters are calculated for every grid block using the three methods and the similarities and discrepancies are analyzed.

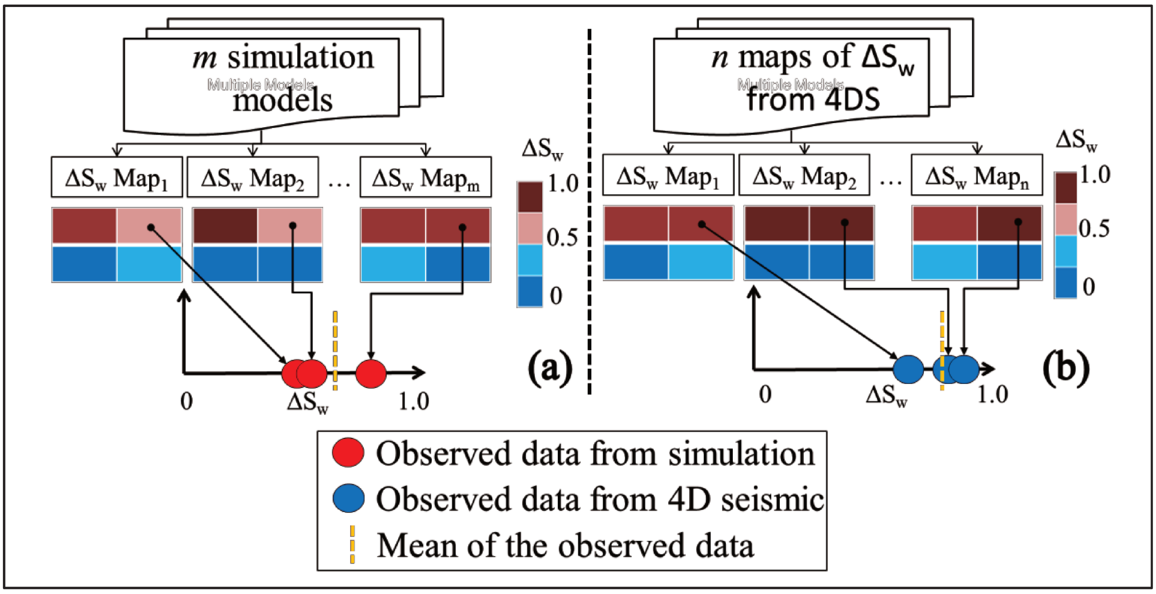

Figure 7–Example of the available information in a simplified reservoir with four blocks (a) multiple simulation models and (b) probabilistic 4D seismic data.

4. Application

4.1 Parametric curves

Tables 2 and 3 present the characteristics of the curves used in this study. We generated 10 data sets for each of the three arrangements (Figure 5). These data sets are the true response used in step 1 from section 3.1. Figure 8 shows examples of the curves obtained from these tables.

Table 2–Parametric curves generated to Arrangements I and II.

	Arrangement I (Normal curves)				Arrangement II (Normal curves)			
Arrangements	PDF_{SIM}		PDF_{SEIS}		PDF_{SIM}		PDF_{SEIS}	
	μ	σ	μ	σ	μ	σ	μ	σ
I.1	0,50	0,07	0,40	0,07	0,60	0,01	0,20	0,10
I.2	0,75	0,15	0,55	0,15	0,70	0,10	0,15	0,10
I.3	0,75	0,05	0,72	0,03	0,70	0,05	0,50	0,05
I.4	0,55	0,30	0,45	0,3	0,70	0,02	0,60	0,02
I.5	0,50	0,08	0,60	0,08	0,80	0,20	0,20	0,20
I.6	0,20	0,10	0,25	0,08	0,40	0,07	0,60	0,07
I.7	0,30	0,07	0,25	0,08	0,40	0,08	0,70	0,08
I.8	0,43	0,07	0,35	0,07	0,75	0,03	0,55	0,05
I.9	0,50	0,10	0,50	0,05	0,10	0,03	0,30	0,05
I.10	0,65	0,08	0,55	0,08	0,60	0,05	0,80	0,05

Table 3–Parametric curves generated to Arrangement III.

	Arrangement III (Normal curves)				Arrangement III (Beta curves)			
Arrangements	PDF_{SIM}		PDF_{SEIS}		PDF_{SIM}		PDF_{SEIS}	
	μ	σ	μ	σ	α	β	α	β
I.1	0,75	0,05	0,70	0,02	2,0	7,0	2	50
I.2	0,80	0,20	0,50	0,05	2,0	7,0	3	70
I.3	0,90	0,30	0,80	0,01	1,0	2,0	4	80
I.4	0,20	0,08	0,10	0,01	1,5	1,5	3	40
I.5	0,36	0,09	0,20	0,05	1,0	2,0	1,50	40
I.6	0,45	0,05	0,40	0,01	1,0	2,0	2,0	30
I.7	0,68	0,09	0,49	0,03	2,0	4,0	3,0	55
I.8	0,72	0,06	0,65	0,03	2,0	4,0	2,0	80
I.9	0,66	0,07	0,55	0,02	3,0	6,0	3,0	85
I.10	0,58	0,10	0,50	0,03	2,0	6,0	2,0	55

For Arrangement III, we studied two types of parametric curves: normal and beta curves. Figure 8c presents Arrangement III using normal curves and Figure 8d, beta curves. Beta curves were also used in this case because they perform well with low variability curves in the interval [0,1]. These low variability curves aim to represent reservoir locations far from injectors, where simulation models can for instance estimate $\Delta Sw \sim 0$ while 4DS estimate ΔSw different from zero (or vice-versa).

To perform steps 2, 3 and 4 from section 3.1 of the methodology, we must generate the discrete data sets. For every arrangement, y=100 vectors SIM and SEIS were generated. Each vector was created with 500 elements, considered a sufficient number for Latin hypercube analysis and also because the available seismic and simulation data generates vectors with this size (see Section 4.2). However, this value can change according to the study.

In this work, therefore, 4000 iterations are performed for each of the three methods studied (4 arrangements x 10 data sets per arrangement x 100 vectors SIM and SEIS), which was considered sufficient for a numerical analysis.

4.2. Synthetic reservoir simulation and 4D seismic data

4.2.1. Reservoir simulation data

The simulation models were based on a reference model that represents a true earth model for this work. The reference model has two facies, four faults (and 13 sub seismic faults), 19 wells (11 injectors and 8 injectors) and is modeled in a fine grid (270x330 blocks and 18 layers). Figure 9a presents the faults of the reservoir and Figures 9b, 9c and 9d show porosity, facies and horizontal permeability distributions, respectively.

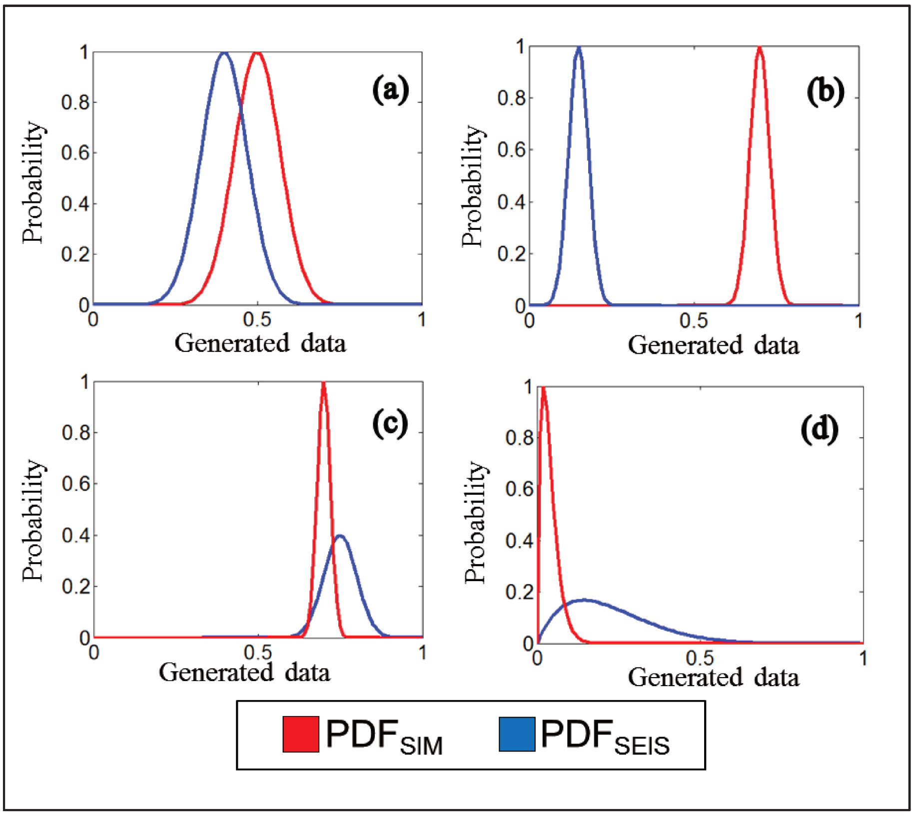

Figure 8–Example of the arrangements used in the present work: (a) Arrangement I, (b) Arrangement II, (c) Arrangement III with normal curves and (d) Arrangement III with beta curves.

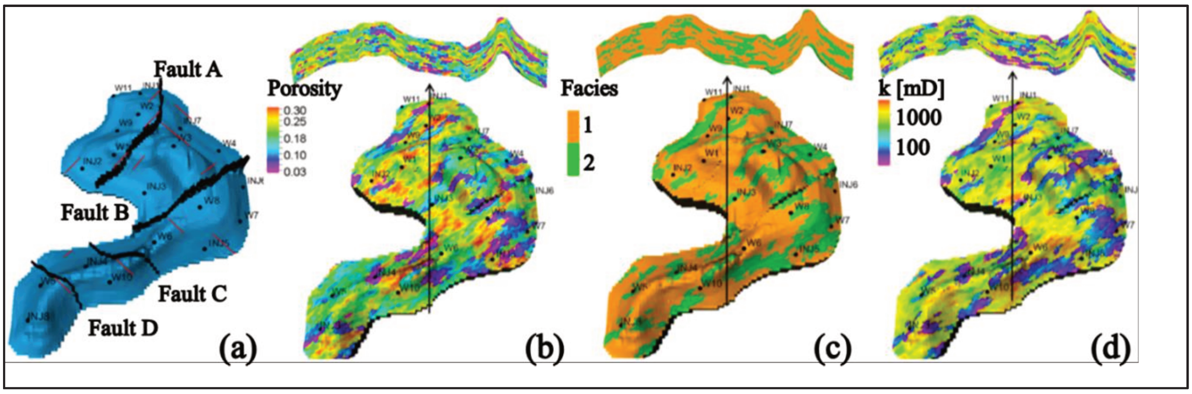

Figure 9–Reference simulation model: (a) main faults and sub seismic faults (red lines), (b) porosity distribution, (c) facies and (d) horizontal permeability distribution.

Information from this reference model, such as production data and well log, is used to build a set of simulation models considering reservoir uncertainties. We consider seven reservoir uncertainties: (1) relative permeability of the two facies, (2) ratio of vertical and horizontal permeability (Kz /Kx), (3) transmissibility of four faults, (4) facies distribution, (5) porosity, (6) absolute permeability, and (7) net-to-gross ratio (NTG). Note that the reference models were used to obtain only data generally measured in real fields.

Considering the reservoir uncertainties described and using the Latin Hypercube technique, we generated *m=500* simulation models in a course grid (90x110 blocks and 9 layers). Almeida *et al.* (2014) presented details about the generation of these models and also about the partial history matching, using well data, performed in these models. Therefore, the simulation data used here are a set of models obtained from an intermediate step of an iterative and probabilistic well history matching process. By simulating these 500 simulation models, we obtained 500 maps of Δp and ΔS_w. These maps are probabilistic information of the simulation models used in this work.

Figure 10a presents the third layer (out of 9) of the average ΔS_w map from the 500 simulation models.

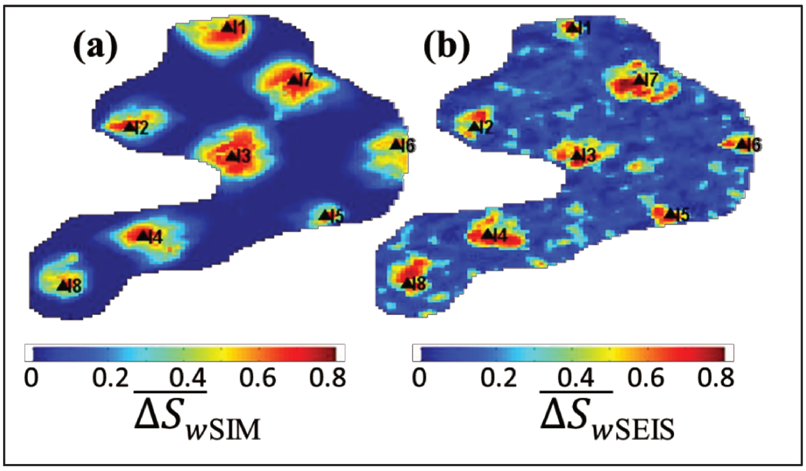

Figure 10–(a) mean of 500 maps of ΔS_w the simulation models and (b) mean of 500 maps of ΔS_w from probabilistic 4D seismic (every map is showing the third layer out of 9).

4.2.2. 4D seismic data

Synthetic 4D seismic data are also obtained from the reference model at two production times (T0 and T1). T1 and T0 are the time of the 3D seismic acquisitions: T0 pre-production (0 days) and T1 after 5 years of production (1800 days). Water saturation estimates are converted to seismic parameters (impedance) by using a petro elastic model (PEM) detailed in Pazetti *et al.* (2015). To produce more realistic data sets a random noise was added to seismic impedances as described in Davolio *et al.* (2014). These impedances with noise are transferred to the simulation model scale (110x90 blocks at 9 layers) and are used as input to a probabilistic inversion procedure, described in Davolio and Schiozer (2015). Consequently, multiple scenarios (*n=500*) of water saturation change are generated from the synthetic seismic impedances. Figure 11 illustrates the overall procedure described above.

This process is repeated for every grid block and gathering the ΔS_w estimates from every block, we obtain the average map shown in Figure 10b, which presents the mean of n ΔS_w estimates for every grid block of layer 3. This map is the average of the 4D seismic information used in the present work.

5. Results and Discussion

5.1. Calculating the accuracy of three methods

- **Step 1: generation the parametric curves and calculation of the real OVLs**

The calculated OVLs are presented in Tables 4 for Arrangements I and II and in Table 5 for Arrangements III.

As explained, we calculate the real values of OVL_{SIM} and OVL_{SEIS} using the PDF_{SIM} and PDF_{SEIS} presented in Tables 2 and 3.

- **Step 2: generating the discrete data sets SIM and SEIS**

Figure 12 exemplifies samples generated in step 2. The black straight lines represent the vectors SIM and SEIS (size 500) created using LHS to sample the PDF_{SIM} and PDF_{SEIS}.

For every PDF, we generated 100 SIM and 100 SEIS vectors with size 500. Figure 13 presents two possible vector SIMs (among the 100 generated) that represent the same PDF_{SIM} with a mean of 0.3 and a standard deviation of 0.07.

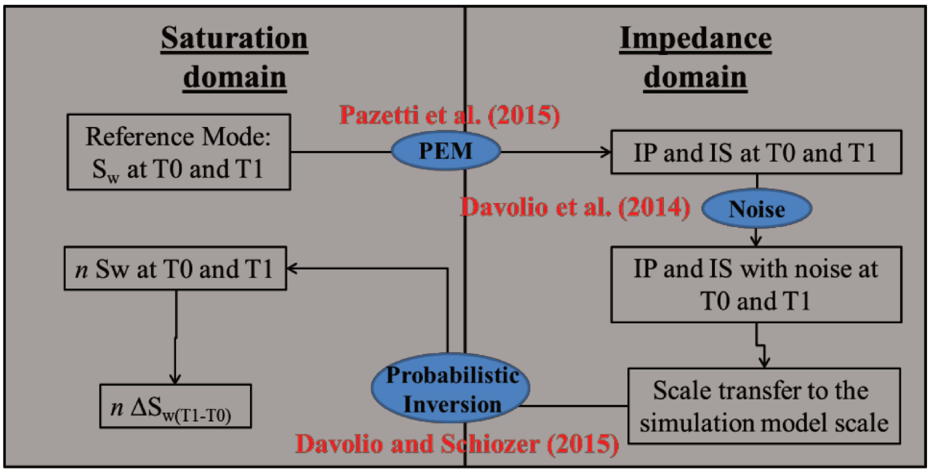

Figure 11–Workflow to obtain 4D seismic data in the present work.

Table 4–Real OVLs calculated in Arrangements I and II.

Arrangement I			Arrangement II		
Arrangements	OVL_{SIM}	OVL_{SEIS}	Arrangements	OVL_{SIM}	OVL_{SEIS}
I.1	91,1	91,1	I.1	24,17	24,17
I.2	89,70	89,70	I.2	4,34	4,58
I.3	95,03	90,62	I.3	24,29	24,15
I.4	82,64	99,18	I.4	7,28	6,84
I.5	100,00	100,00	I.5	51,65	50,10
I.6	91,08	91,08	I.6	55,93	55,13
I.7	96,23	99,21	I.7	30,29	27,04
I.8	98,36	94,42	I.8	23,75	2,47
I.9	91,80	91,80	I.9	25,45	3,62
I.10	80,60	99,04	I.10	21,19	21,19

Table 5–Real OVLs calculated in Arrangement III.

Arrangement III (Normal curves)			Arrangement III (Beta curves)		
Arrangements	OVL_{SIM}	OVL_{SEIS}	Arrangements	OVL_{SIM}	OVL_{SEIS}
III.1	49,26	98,52	IV.1	27,65	91,30
III.2	20,10	99,47	IV.2	21,42	94,38
III.3	5,03	93,69	IV.3	18,80	96,80
III.4	9,07	93,60	IV.4	14,07	98,27
III.5	40,41	96,27	IV.5	20,57	87,93
III.6	23,69	98,00	IV.6	32,40	95,28
III.7	12,25	99,50	IV.7	16,38	96,86
III.8	55,97	99,18	IV.8	6,63	84,06
III.9	22,15	99,21	IV.9	4,88	92,86
III.10	44,51	99,18	IV.10	19,90	90,11

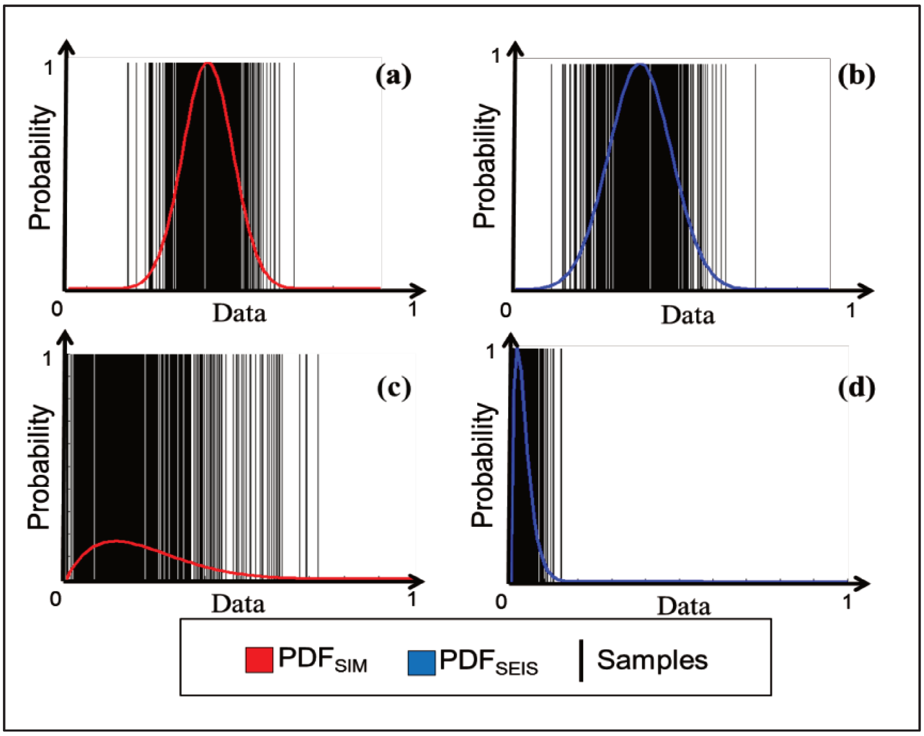

Figure 12–Example of the sample (all with the same size, 500) used in the presented work: (a) SIM_y representing a normal PDF_{SIM}, (b) $SEIS_y$ representing a normal PDF_{SEIS}, (c) SIM_y used to mimic the beta PDF_{SIM} and (d) $SEIS_y$ to mimic the beta PDF_{SEIS}.

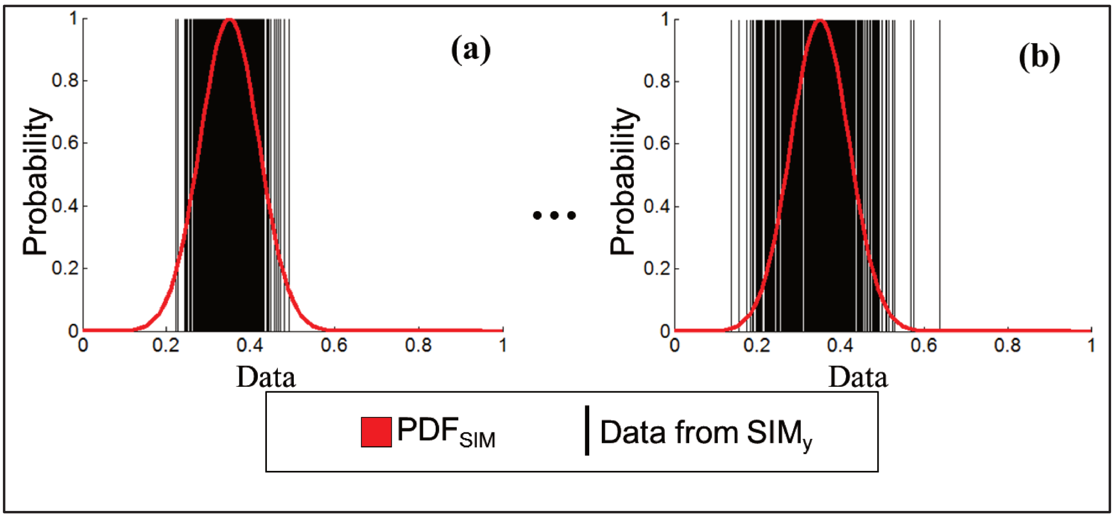

Figure 13–Same parametric curve, but different possible samples.

- **Step 3: calculating OVL_{SIM} and OVL_{SEIS} using non-parametric estimators**

Using SIM and SEIS vectors from the previous step we calculated the probability of the data using each of the three methods. For instance, Figure 14a shows Arrangement III using beta curves. Using SIM_1 and $SEIS_1$ we estimated the probabilities using a histogram, ECDF and KDE presented in Figure 14b, Figure 14c and Figure 14d, respectively.

Dashed green lines represent the OVLCs. Using these estimates, we calculated OVL_{SIM1} and OVL_{SEIS1}. We repeated the same procedure for the other SIM and SEIS vectors (e.g. OVL_{SIM2} and OVL_{SEIS2}…OVL_{SIM100} and $OVL_{SEIS100}$).

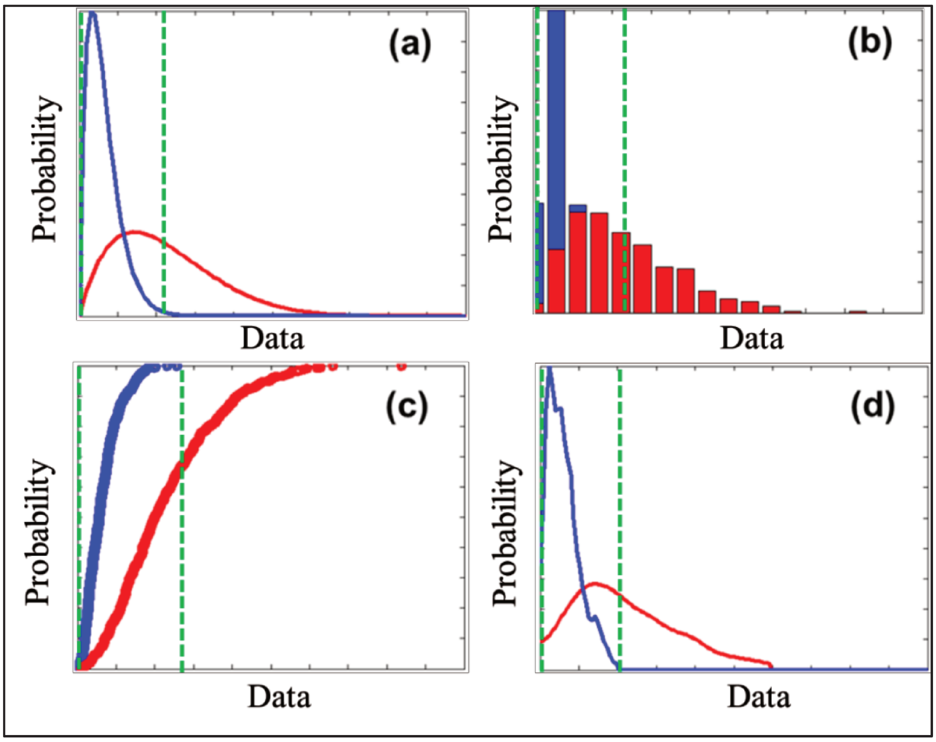

Figure 14–Example of identification of the OVLC interval (in green): (a) parametric curves, (b) histograms, (c) ECDF and (d) KDE.

- **Step 4: quadratic difference between estimated OVLs and true values**

Using OVLs from the previous step, we calculated the QD_{SIM} and QD_{SEIS} for each method (Equation 11 and 12) and then the Normalized QD_{SIM} and QD_{SEIS}, using Equations 13 and 14.

Figure 15a presents the Normalized QD calculated using the 10 Arrangements I (Table 2). The KDE method presented the lowest Normalized QD for this situation, for both OVL_{SIM} and OVL_{SEIS}. ECDF presented the greatest Normalized QD followed by the histogram.

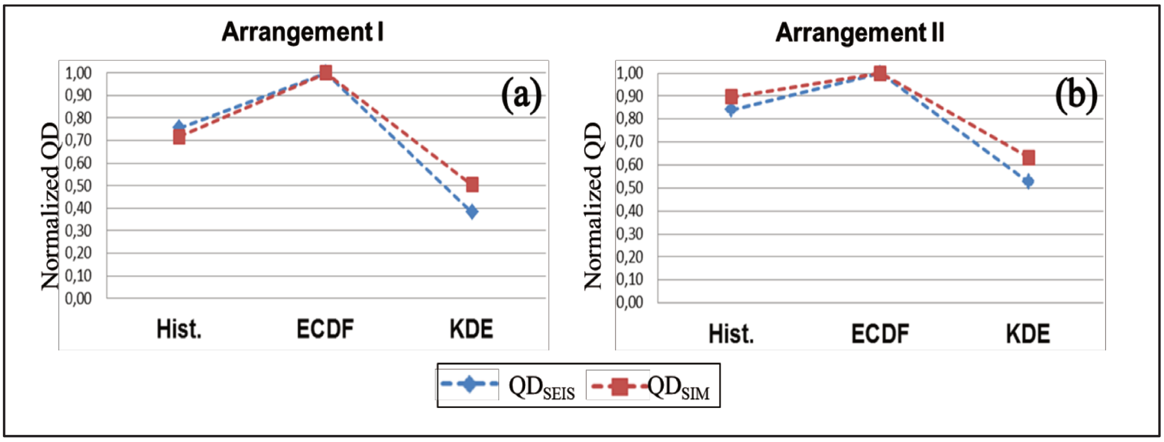

Figure 15–Normalized QD calculated for (a) Arrangement I and (b) Arrangement II.

In Arrangement II, the behavior of the three methods followed the same trends as in Arrangement I, as shown in Figure 15b. The ECDF method had the greatest Normalized QD, followed by the histogram. The kernel estimator had the lowest value of the three methods. However, for Arrangement II, we computed $QD_{SIM\text{-}KDE}$~0,61 and $QD_{SEIS\text{-}KDE}$~0,51, which were greater values than calculated for Figure 15a, $QD_{SIM\text{-}KDE}$~0,50 and $QD_{SEIS\text{-}KDE}$~0,40. The difference in KDE between Figure 15a and 15b is largely due to the type of data analyzed. In Arrangement II, we compare two PDFs with distinct trends; thus, the OVLC interval is identified in the tail region of the PDFs (far from the data average). Data density is low in the tail region; signifying little information to estimate the PDFs. The lack of information in these regions increased the KDE error in Arrangement II. Figure 16 shows an example of the PDFs yielded from KDE for Arrangements I and II. In the dashed rectangle (Figure 16c and 16d), we can see the influence of the lack of information in the tails when estimating the OVLs. For Arrangement I (Figure 16a and 16b), the error due to this lack of information at the tails of the curves is unimportant, since the OVLs in this situation do not heavily depend on the tail values. The lack of information in tail regions also affects the estimates from the histogram and ECDF, mainly the ECDF, however, these errors were less visible because the histogram and ECDF had great errors in both Arrangements I and II.

Despite the differences between the Normalized QD in Arrangements I and II, the KDE presented the lowest values for both arrangements which proves this method to be the most accurate among the three.

For Arrangement III, considering normal parametric curves, we can see a difference between OVL_{SIM} and OVL_{SEIS} (Figure 17a). In this situation, when estimating OVL_{SEIS}, the ECDF showed the greatest QD. However, to estimate OVL_{SIM}, the histogram was less accurate.

We observed that to estimate parametric curves with low standard deviation, the histogram performed well, capturing the trend of the data. However, to estimate parametric curves with a high standard deviation, the histogram had the highest error. As the dispersion of data increased, the error between the parametric curve and estimation also increased.

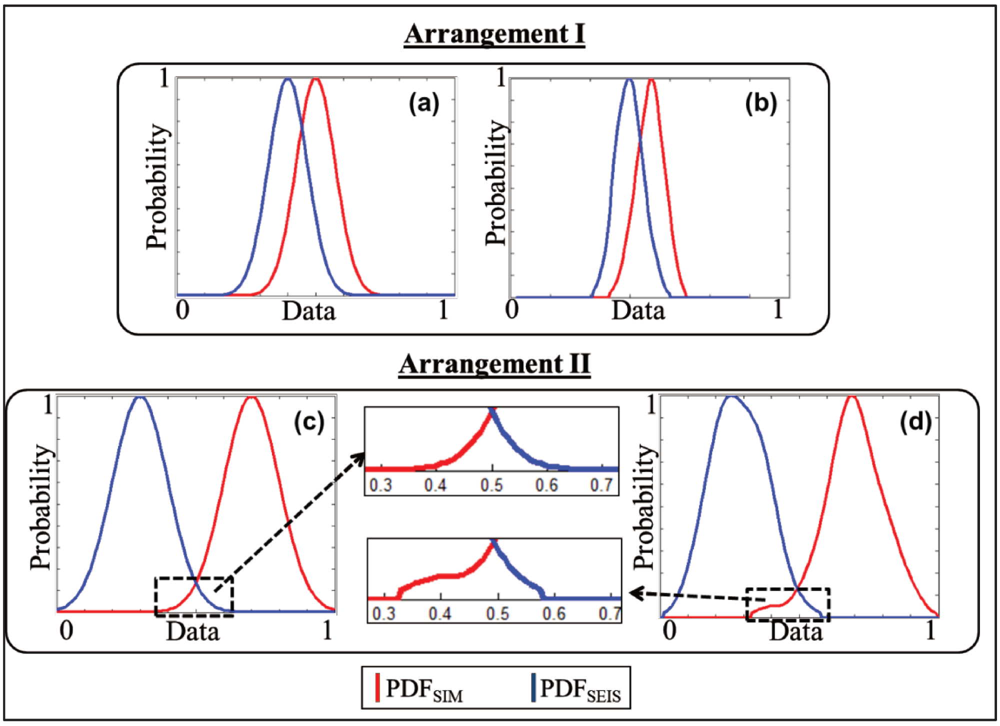

Figure 16–Influence of the tail in the OVLs parameters when Arrangements I and II are studied: (a) parametric curve Arrangement I, (b) KDE for the parametric curve, (c) parametric curve Arrangement II and (d) KDE for estimating parametric curve Arrangement II.

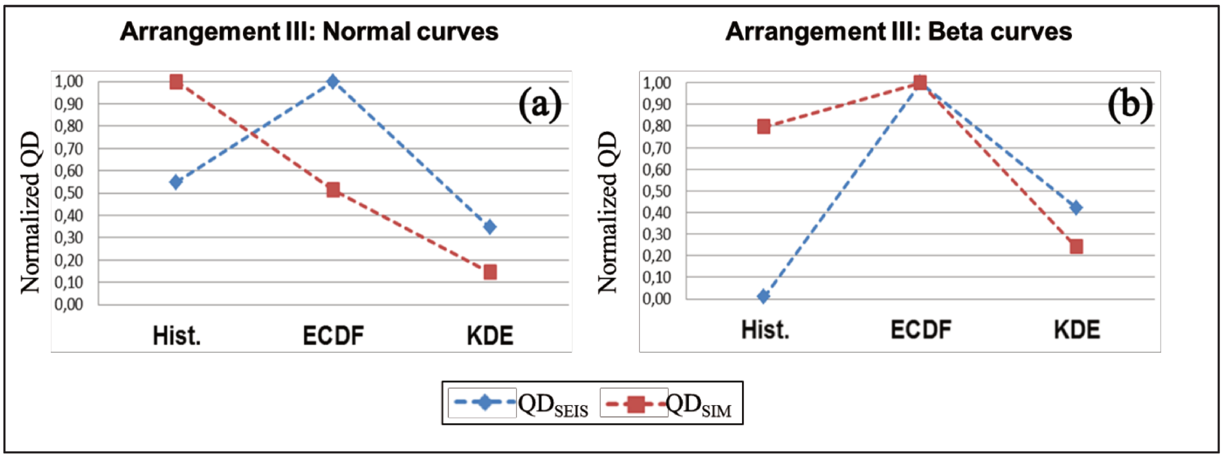

Figure 17–Normalized QD calculated for (a) Arrangement III using normal curves and (b) Arrangement III using beta curves.

Figure 18 shows an example of how the dispersion of the data affected the density estimation yielded from the histogram. For data with high precision ($\sigma = 0.01$), the histogram estimated the normal curve accurately (Figure 18a). When the dispersion increased, the accuracy of the histogram dropped visible in Figure 18b, 18c and 18d: the increase of gaps (whitespaces) between the normal curve and the histogram, highlighted by black arrows.

Observing Figure 17a, we also noted that the ECDF estimated OVL_{SIM} with a lower error than OVL_{SEIS}, even lower than OVL_{SIM} from the histogram estimation. In Figure 19, only the tail values from $ECDF_{SEIS}$ directly affected the OVL interval. The tail values (black circles

Figure 19) of the $ECDF_{SIM}$ did not impact the identification of the OVL interval, therefore, the tail values from $ECDF_{SIM}$ do not change the value of the OVLC interval. The tail values are a critical region in data analysis, because of the low density of points and for this reason, the OVL_{SIM} yielded from ECDF in this situation had a lower error than OVL_{SEIS}.

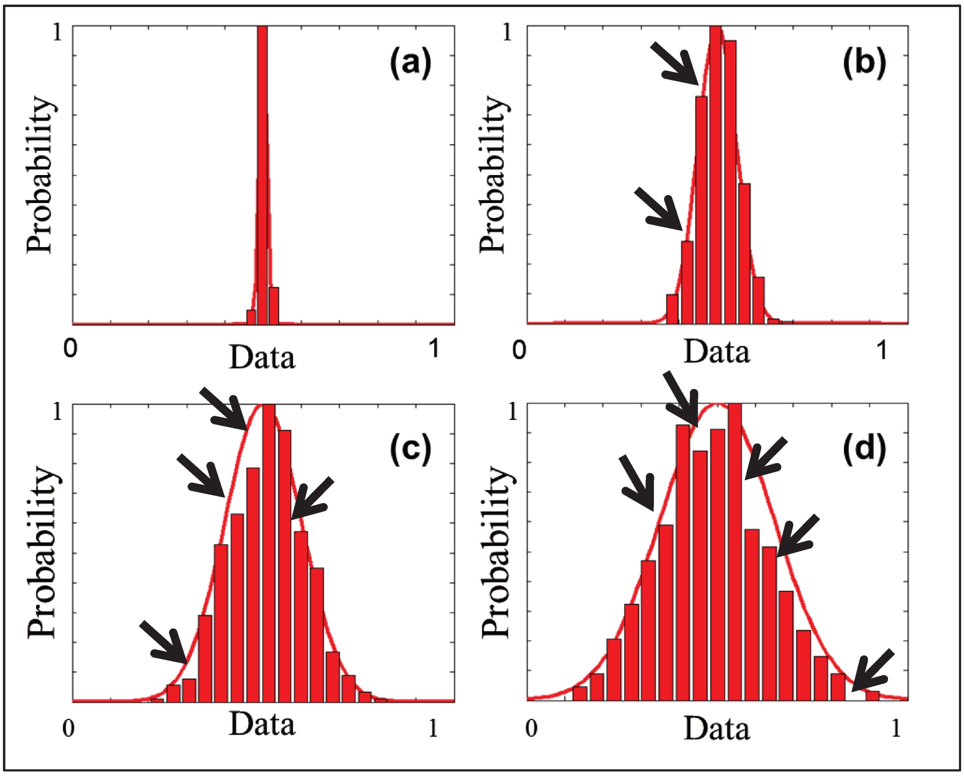

Figure 18–Example of histogram estimations using data with the same mean but different standard deviations: (a) $\mu = 0,5$ e $\sigma = 0,01$; (b) $\mu = 0,5$ e $\sigma = 0,05$; (c) $\mu = 0,5$ e $\sigma = 0,10$ e (d) $\mu = 0,5$ e $\sigma = 0,20$.

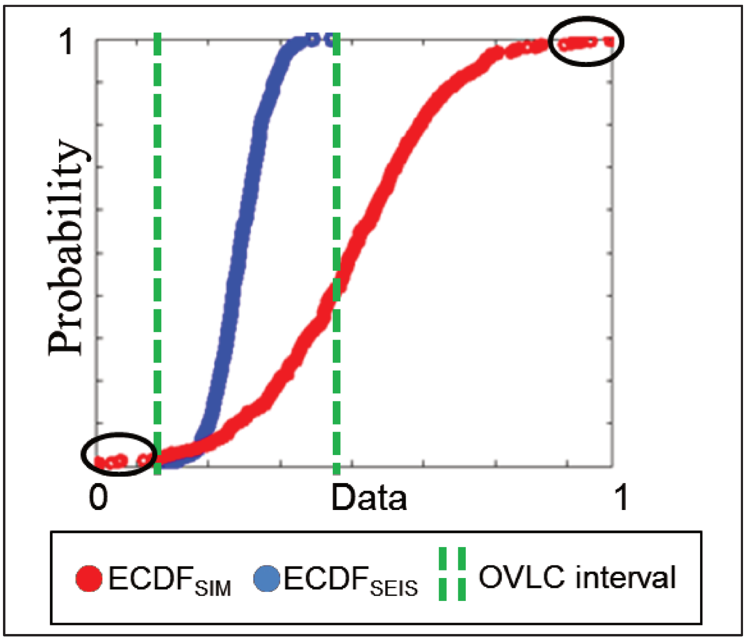

Figure 19–Influence of the tail values in ECDF estimates.

Figure 17b presented the normalized errors for Arrangement III using beta curves. The ECDF presented the greatest error and inaccurately estimated non-Gaussian curves.

The estimated OVL_{SIM} using the histogram presented high Normalized QD_{SIM}, close to ECDF, however, the error of the histogram was practically zero in the OVL_{SEIS} estimation. The characteristic of the beta curves in Figure 19a show high precision of PDF_{SEIS} . As discussed in Arrangement III using normal curves, the histogram was accurate with precise curves. The PDF_{SEIS} in Arrangement III using beta curves was precise, and so the histogram estimated OVL_{SEIS} well. The kernel estimator followed the trends of the previous studied arrangements, presenting low overall error (Normalized QD_{SIM} + Normalized QD_{SEIS}).

Analyzing the overall results (Arrangements I, II and III), we found the KDE to be most accurate and most stable. In the next section we discuss the accuracy of the methods when probabilistic data from simulation and seismic are used.

5.2. Comparing the three methods applied in a synthetic reservoir model

We calculated OVL_{SIM} using the three methods based on $m=n= 500$ maps of ΔS_w from simulation models and 4D seismic data. Figure 20 presents the results. As explained in section 3.2, every reservoir grid block provides the information necessary to estimate the probability of water saturation change from simulation and seismic data; thus, in every reservoir grid block OVL_{SIM} and OVL_{SEI} are computed allowing identifying whether simulation or seismic data are more precise.

Comparing Figure 20a with 20c we observe similar behavior although $OVL_{SIM\text{-}KDE}$ presented slightly lower values than $OVL_{SIM\text{-}HIST}$ (a difference lower than 5%). OVL_{SIM} estimated from ECDF presented significant differences when compared with the histogram and KDE (Figure 20b). The main difference is seen in the yellow rectangle, where $OVL_{SIM\text{-}ECDF}$ ~ 50%, while $OVL_{SIM\text{-}HIST}$ and $OVL_{SIM\text{-}KDE}$ ~ 100 %. This difference reflects the influence of the tail values from the ECDF estimates.

Taking a grid block in this area, ECDF and KDE seem alike (Figure 21a and 21d). However, when we focus on the PDF_{SIM} from ECDF (Figure 21b), the influence of the tail values from ECDF is evident.

The influence of the neighboring values does not account for ECDF analysis, different to KDE. In other words, the ECDF weighted every observed datum uniformly, while KDE considers the density of neighbors. Initially, we considered this to be an advantage of the ECDF method, since we do not assume any information of the data ("raw data"), however, the results from section 5.1 and 5.2 show that it is a drawback of the method, a significant cause of instability seen in ECDF: Figure 21b shows a difference lower than $5*10^{-3}$ in ΔS_w reducing $OVL_{SIM\text{-}ECDF}$ by approximately 50%, while the KDE (more accurate according to results from section 5.1) show OVL_{SIM} ~ 100%.

Analyzing OVL_{SEIS}, we see that the three methods presented similar behavior close to injectors. However, in regions far from injectors, we observed significant differences, highlighted by the yellow arrows in Figure 22.

Although KDE shows that 4D seismic data can be useful to calibrate the water front (locations where $OVL_{SEIS}>80$), in reservoir locations far from injectors, KDE shows that 4D seismic data cannot be used ($OVL_{SEIS} < 80$). Figure 10a shows that this result is accurate, since ΔS_w is close to zero in regions far from injectors. KDE shows that in regions far from well injectors, 4D seismic data are not useful because the seismic estimation of water saturation in these locations are affected by noise, agreeing with Figure 10b.

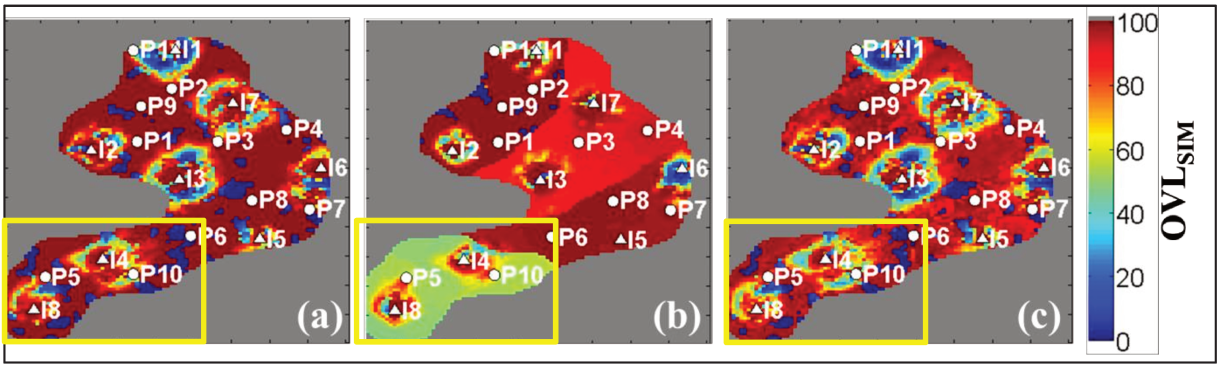

Figure 20–OVL_{SIM} parameters: (a) Histogram, (b) ECDF and (c) KDE.

The histogram and ECDF follow the same pattern as KDE; however, we can see reservoir locations where both methods indicate that seismic information can be used (remarkably ECDF). Note that in the region around well P6, OVL_{SEIS} from the histogram and ECDF are greater than $OVL_{SEIS\text{-}KDE}$. In this region, observed seismic data have considerable noise (Figure 10b), therefore, we would expect this location not to contribute further information to the simulation models; nonetheless, the histogram and ECDF suggest that there are blocks in this region that could be used.

Therefore, OVL_{SEIS} and OVL_{SIM} maps show that KDE presented the most reliable results although the histogram based method also generated good results in section 5.2. Following the same trends showed in section 5.1, the ECDF was the most inaccurate and unstable method.

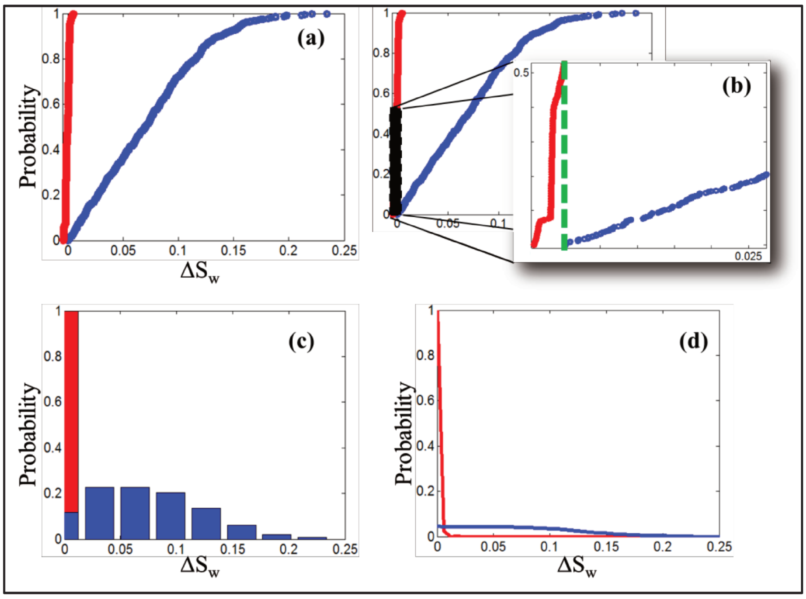

Figure 21–Comparing the tail of the three methods: (a) ECDF, (b) Zoom in PDFs from ECDF, (c) histogram and (d) KDE.

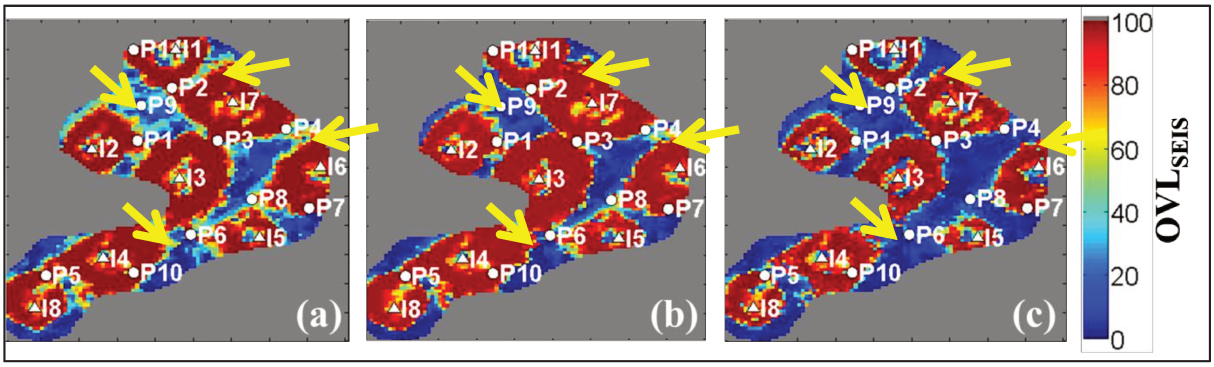

Figure 22–OVL_{SEIS} parameters: (a) Histogram, (b) ECDF and (c) KDE.

6. Conclusions

We proposed a methodology to calculate the accuracy of three non-parametric methods (histogram, ECDF and KDE) using the modified overlapping coefficient. We made two studies: (1) using parametric curves to calculate the real OVLs and then calculate the estimated OVLs and (2) comparing the estimated OVLs using data from probabilistic simulation models and 4D seismic.

In the first study, the kernel density estimator presented the lowest error among the three estimators, validating its stability and accuracy to calculate the OVLs. In this situation the histogram presented regular results and ECDF, the worst errors. In the second study, comparing the available data and trends shown by OVL_{SIM} and OVL_{SEIS}, we again found the KDE to be the most accurate, the histogram presented good estimates while ECDF had several incorrect results.

In conclusion, the KDE is the best alternative to estimate the probability density distribution for data representing estimates of saturation change from reservoir simulation models and 4D seismic data.

Acknowledgments

This work was carried out in association with the ongoing Project registered as "BG-07 Redução de incertezas através da incorporação de sísmica 4D na modelagem do reservatório" (UNICAMP/BG Brasil/ANP) funded by BG Brasil under the ANP R&D levy as "Compromisso de Investimentos com Pesquisa e Desenvolvimento". The authors thank also UNISIM, DE-FEM-UNICAMP, CEPETRO, and CAPES for supporting this work and CMG and Schlumberger for software licenses.

References

Almeida, F. L., Davolio, A., & Schiozer, D. J. (2014). A New Approach to Perform a Probabilistic and Multi-objective History Matching. *SPE Annual Technical Conference and Exhibition.* Amsterdan, The Netherlands.

Al-Saleh, M. F., & Samawi, H. M. (2007). Inference on Overlapping Coefficients in Two Exponential Populations. *Journal of Modern Applied Statistical Methods*, 503- 516.

Assunção, G. S., Davolio, A., & Schiozer, D. J. (2016). A Methodology to Integrate Multiple Simulation Models and 4D Seismic Data Considering Their Uncertainties. *SPE Annual Technical Conference and Exhibition* . Dubai, U.A.E.

Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). KERNEL DENSITY ESTIMATION VIA DIFFUSION. *The Annals of Statistics, 38*(5), 2916-2957.

Bradley, E. L. (2006). Overlapping Coefficient . *Encyclopedia of Statistical Sciences*.

Correia, G. (2013). *Beta model: reservoir characterisation.* Internal report - BG UNISIM, Campinas, Brazil.

D'Agostinho, R. B., & Stephens, M. (1986). *Goodness-of-fit-tecnhiques (Statistics: a Series of Textbooks and Monographs)* (Vol. 68). Dekker.

Davolio, A., & Schiozer, D. (2015). Joining Two Probabilistic Approaches - 4D Seismic Data and Simulation Models. *77th EAGE Conference & Exhibition.* Madrid, Spain.

Davolio, A., & Schiozer, D. J. (2012). Pressure and Saturation Estimation from P ans S Impedance: A Theorical Study. *Journal of Geophysics and Engineering, 9*, 447-460.

Davolio, A., Maschio, C., & Schiozer, D. J. (2013). A methodology to constrain pressure and saturation estimation from 4D seismic using multiple simulation models and observed data. *Journal of Petroleum Science and Engineering*, 51-61.

Davolio, A., Maschio, C., & Schiozer, D. J. (2013). Local History Matching Using 4D Seismic Data and Multiple Models Combination. *EAGE Annual Conference & Exhibition.* London, United Kingdom.

Davolio, A., Maschio, C., & Schiozer, D. J. (2014). A Methodology To Calibrate Water Saturation Estimated From 4D Seismic Data. *Journal Of Geophysics And Engineering, 11*(5).

Emerick, A. A., & Reynolds, A. C. (2012). History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 639-659.

Freedman, D., & Diaconis, P. (1981). On this histogram as a density estimator: L2 theory. *Zeit. Wahr. ver. Geb, 57*, 453–476.

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric Statistical Inference.* Springer.

Landa, J. L., & Kumar, D. (2011). Joint Inversion of 4D Seismic and Production Data. *SPE Annual Technical Conference and Exhibition* . Denver, Colorado, U.S.A.

Maschio, C., Vidal, A. C., & Schiozer, D. J. (2008). A framework to integrate history matching and geostatistical modeling using genetic algorithm and direct search methods. *Journal of Petroleum Science & Engineering*, pp. 34-42.

Mizuno, S., Yamaguichi, T., Fukushima, A., & Ohashi, Y. (2005). Overlap coefficient for assessing the similarity of pharmacokinetic data between ethnically different populations. *Clinical Trials*.

Pazetti, B., Davolio, A., & Schiozer, D. J. (2014). *The impact of roche and fluid uncertainties in the estimation of saturation and pressure from 4D seismic.* Internal Report - BG UNISIM, Campinas, Brazil.

Raykar, V. C. (2002). *Probability Density Function Estimation by different Methods Vikas Chandrakant Raykar.* Report written for ENEE 739Q SPRING 2002.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Annals of Mathematical Statistics, 27*, 832–837.

Sarma, P., Durlofsky, L. J., Aziz, K., & Chen, W. H. (2007). A New Approach to Automatic History Matching Using Kernel PCA. Society of Petroleum Engineers. doi:10.2118/106176-MS.

Schmid, F., & Schmidt, A. (2006). Nonparametric Estimation of the Coefficient of Overlapping - Theory and Empirical Application. *Computational Statistics & Data Analysis*, 1583-1596.

Shalizi, C. R. (2016). *Advanced Data Analysis from an Elementary Point of View.*

Sheather, S. J. (2004). Density Estimation. *Statistical Science, 19.*

Shi, P., Lin, Y.-A., Pastuska, M., Cui, H., & Mackay, J. (2014). Triggered Sorting and Co-Assembly of Generically Enfineered Protein Microdomains in Cytoplasm. *Advanced material*, 449-454.

Silverman, B. W. (1998). *Density Estimation for Statistics and Data Analysis.* (1. ed, 1. CRC reprint. ed.). London: Monographs on statistics and applied probability, 26 - Chapman & Hall.

Stephen, K. D., MacBeth, C., & Christie, M. (2005). Multiple model seismic and production history matching: a case study. *SPE EUROPEC/EAGE Ann. Conf. Exhibition.* Madrid, Spain.

Takada, T. (2001). Nonparametric density estimation: A comparative study. *Economics Bulletin*, 1-10.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S.* Springer.

Wasserman, L. (2005). *All of nonparametric statistics.* Springer.

Weitzman, M. (1970). *Measure of the overlap of income distribution of while and Negro families in the United States.* U. S. Department of Commerce, Bureau of the Census,, Washington, DC.

Yeh, T. -H., Jimenez, E., Van Essen, G., Chen, C., Jin, L., Girardi, A., & Conn, A. R. (2014). Reservoir Uncertainty Quantification Using Probabilistic History Matching Workflow. *SPE Annual Technical Conference and Exhibition.* Amsterdam, The Netherlands.

APPENDIX B

Complementary results of paper "A Methodology to Integrate Multiple Simulation Models and 4D Seismic Data Considering Their Uncertainties"

1. Introduction

Paper 1, in Chapter 2, introduces the OVL methodology and shows its application with a specific set of data. The results of this appendix represent a complementary study performed to evaluate the robustness of the proposed methodology when different quality of simulation and seismic data are available. Quality here refers to the presence of uncertainties and the corresponding accuracy of the data sets.

The OVL methodology is applied and depending of the data sets studied, it is used region 2 or 4 to select the most precise Δp maps from simulation or seismic. Figure B.1 summarizes the steps followed.

Remembering that in region 4 seismic data is more precise than simulation data and region 2 is the opposite. Therefore, it is possible to use the grid blocks in region 4 to select Δp maps from simulation within the OVLC interval, reducing the variability of simulation data (Figure B.1d). This procedure is repeated for every grid block in (and only in) region 4. In the end of the procedure the 10% most frequently selected models considering all grid blocks of region 4 are chosen. The procedure is the same for region 2, however, instead of selection of Δp maps from simulation, it is selected Δp maps from seismic (Figure B.1b), since in this region pressure change estimates from simulation is more precise than seismic pressure changes estimates (Figure B.1a).

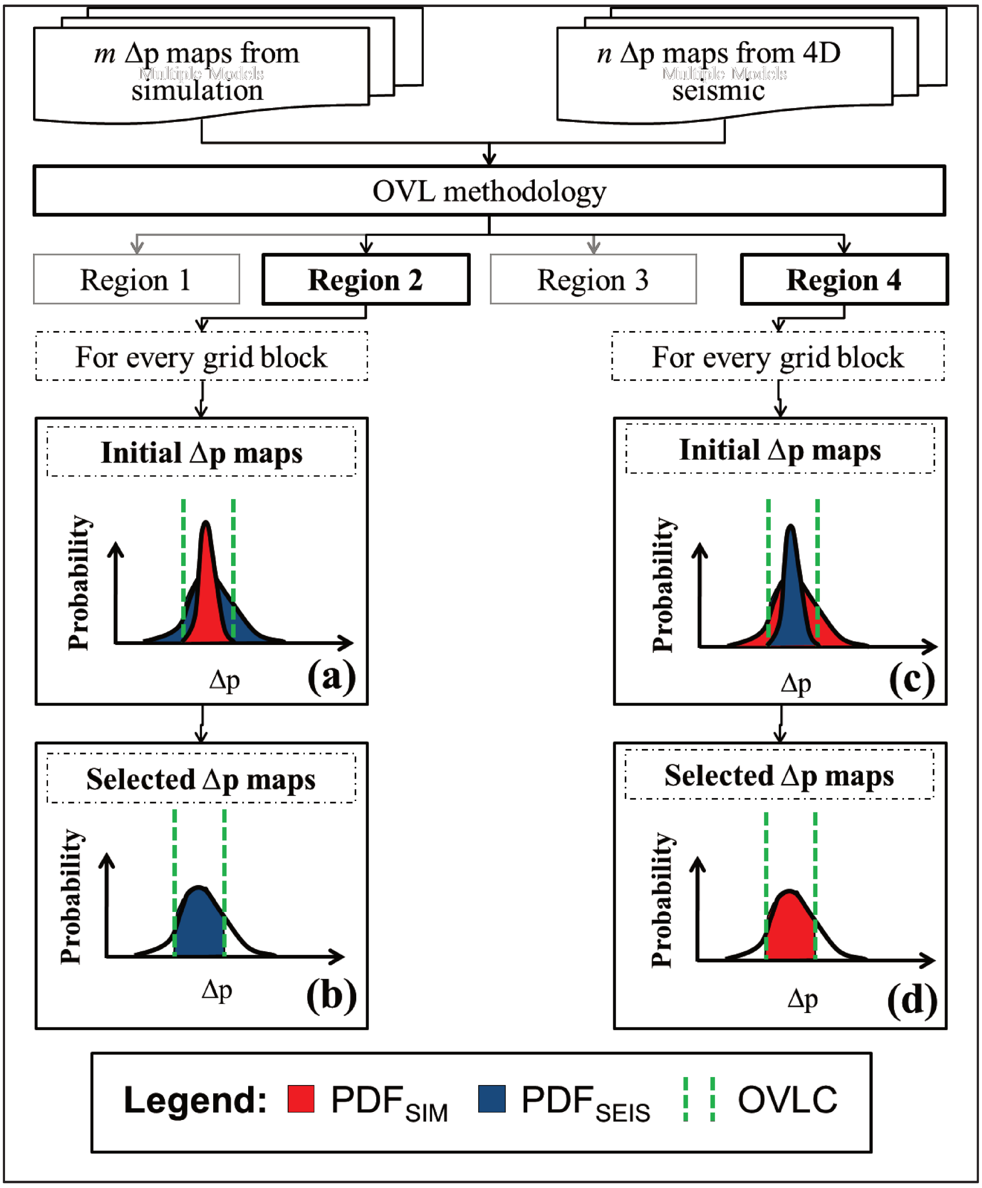

Figure B.1–OVL methodology: (a) example of PDF_{SIM} and PDF_{SEIS} in region 2, (b) example of selected data from seismic using simulation, (c) example of PDF_{SIM} and PDF_{SEIS} in region 4 and (d) example of selected data from simulation using seismic.

2. Application

For this study two sets of simulation data are used: (1) a set of 500 models generated at the beginning of a probabilistic well history-matching process and (2) the set of 97 history-matched models considering only well data. The simulation models used were obtained from the history matching presented in Almeida *et al.* (2014): the 500 models were generated in step 2 (out of 4) and the 97 were generated in step 4. Figure B.2a and B.2b presents the average of the simulation data used in the present section. Figure B.2c illustrates the expected value of Δp, obtained from the reference model (Assunção *et al.*, 2016a).

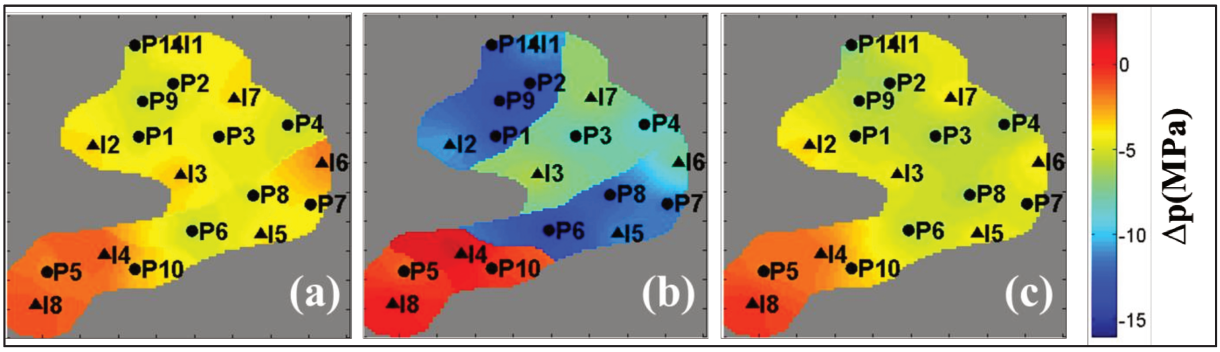

Figure B.2– Maps of pressure changes from simulation: (a) mean of the 97 Δp maps from history matched simulation models, (b) mean of the 500 Δp maps from partially history matched simulation models and (c) reference map (expected values). Every map is showing the 3rd layer out of 9.

Two sets of 4D seismic data are considered, which are represented by multiple maps of Δp estimated from a probabilistic synthetic seismic inversion presented in Davolio and Schiozer (2015). The first set, 125 Δp maps were obtained without noise added and the second set, 500 Δp maps, were generated from the probabilistic inversion with random noise added, as described in Davolio *et al.* (2014). The mean of the 125 Δp maps from seismic without noise and the 500 Δp maps from seismic with noise is shown in Figures B.3a and B.3b, respectively. Again, Figure B.3c illustrates the expected value of Δp from the reference model (the same shown in Figure B.2c, but with a different color scale).

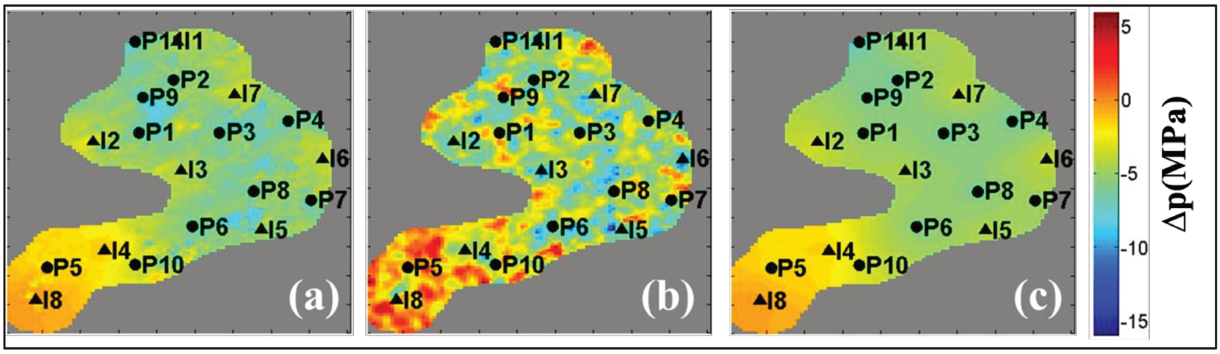

Figure B.3–Maps of pressure changes from seismic: (a) mean of the 125 Δp maps from probabilistic seismic inversion without noise, (b) mean of the 500 Δp maps from probabilistic seismic inversion with noise (c) reference map (expected values). Every map is showing the 3rd layer out of 9.

In this study four cases are built using different combinations of the datasets previously described. Table B.1 presents the cases studied. The reasons to investigate these cases are explained in detail bellow:

- **Case I:** uses the same set of data presented in Paper 1. They are shown again in order to compare the results from this case with the others cases analyzed.
- **Case II:** aims to study if the methodology correctly identifies which information is the most precise, when seismic information are (intentionally) more precise than simulation. Additionally, it is tested whether the selected maps from simulation after the integration of seismic data are more accurate than the initial ones.
- **Case III:** in this case, it was tested if simulation data could be used to constrain and better interpret seismic data. Thus, based on pressure changes estimates from history-matched models, the most precise seismic maps are selected.
- **Case IV:** the idea of this case is the same of the previous case, however when seismic data with and without noise are jointly studied, it is assured that there are "good" Δp maps from seismic available, thus it is possible to observe if the methodology select those maps.

Table B.1– Cases studied

		Simulation Data	
		Partially history-matched models	History-matched models
Seismic Data	With Noise	**Case I**	**Case III**
	Without Noise	**Case II**	——
	With and Without Noise	——	**Case IV**

3. Results and discussions

- **Case I:**

By comparing Δp maps from 4D seismic with noise and Δp maps from simulation data partially history-matched, it was identified regions where seismic data is more precise than simulation, however it was not observed regions where simulation is more precise than seismic data (Figure B.4b). The mean of pressure changes estimates from the 500 initial simulation maps presented significant differences relative to the reference map (Figure B.4a).

Observing Figure B.4c, the 50 (10% of 500) selected Δp maps after 4D seismic incorporation presented better results than the initial ones. The 4 regions map, in Figure B.4d shows that more reservoir locations are in agreement (seismic and simulation data in agreement), since the number of grid blocks in region 1 increased considerably.

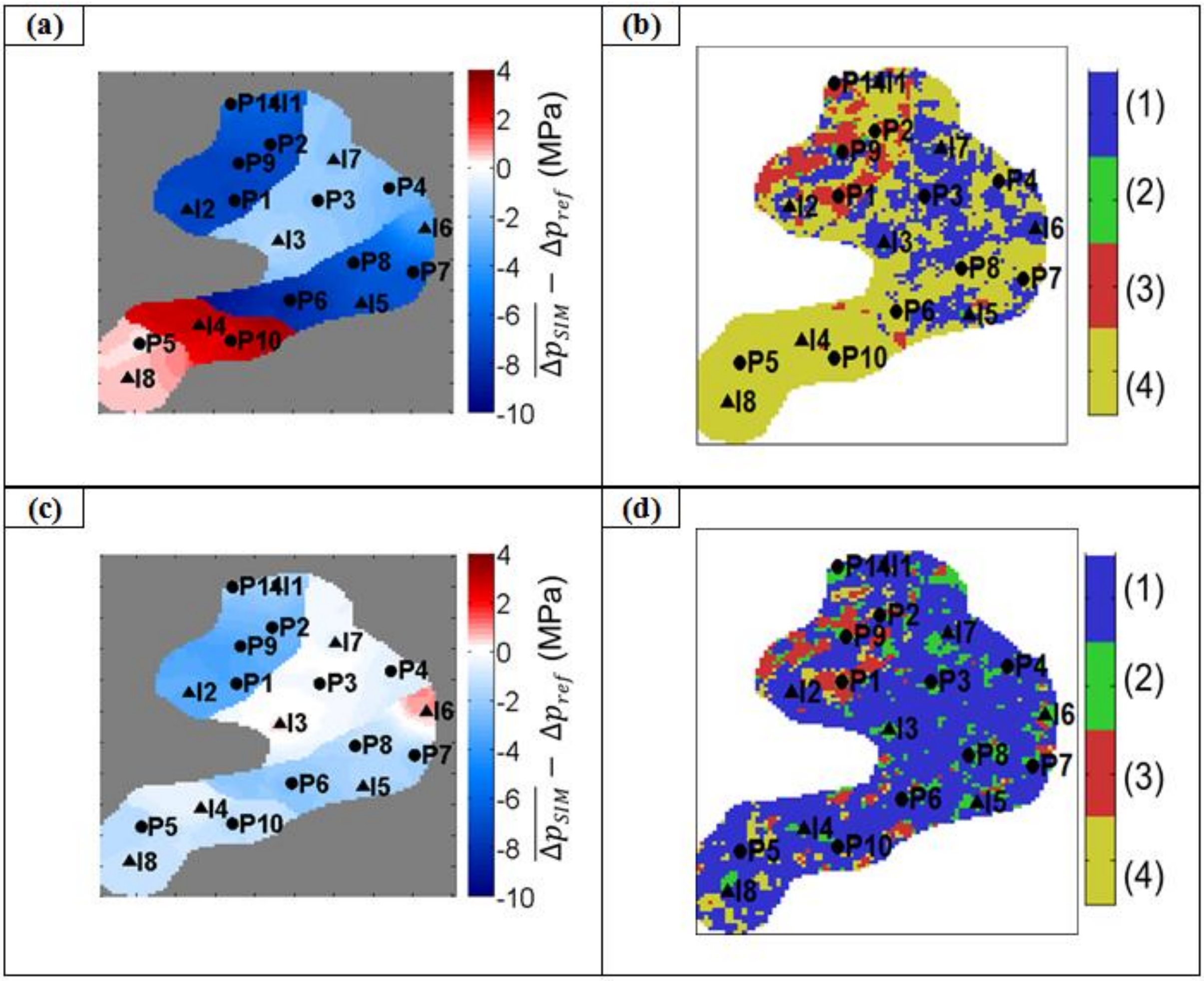

Figure B.4–Case I: before (top) and after (bottom) models selection. (a) Difference between the mean of 500 Δp maps from partially history-matched simulation models and the reference map, (b) 4 regions map comparing the 500 Δp maps from seismic and simulation, (c) Difference between the mean of 50 Δp maps selected and the reference map and (d) 4 regions map comparing the 500 Δp maps from seismic and the 50 Δp maps from simulation. Every map is showing the 3rd layer out of 9.

- **Case II:**

In Figure B.5, Δp maps from 4D seismic without noise are compared with Δp maps from partially history-matched models. As expected, all grid blocks in Figure B.5b are in region 4, showing that the methodology properly identified regions where seismic data is more precise than simulation data. Observing Figure B.5c, the 50 (10% of 500) selected Δp maps from simulation presented results closer to the reference model than the 500 initial maps (Figure B.5a).

Figure B.5d demonstrates that after seismic incorporation, there are more grid blocks in agreement (region 1) than the initial simulation data. Comparing Figure B.5d with Figure B.4d, the number of grid blocks in region 1 was smaller for the first case, which was also expected, because 4D seismic used in case II has less variability than 4D seismic data from case I.

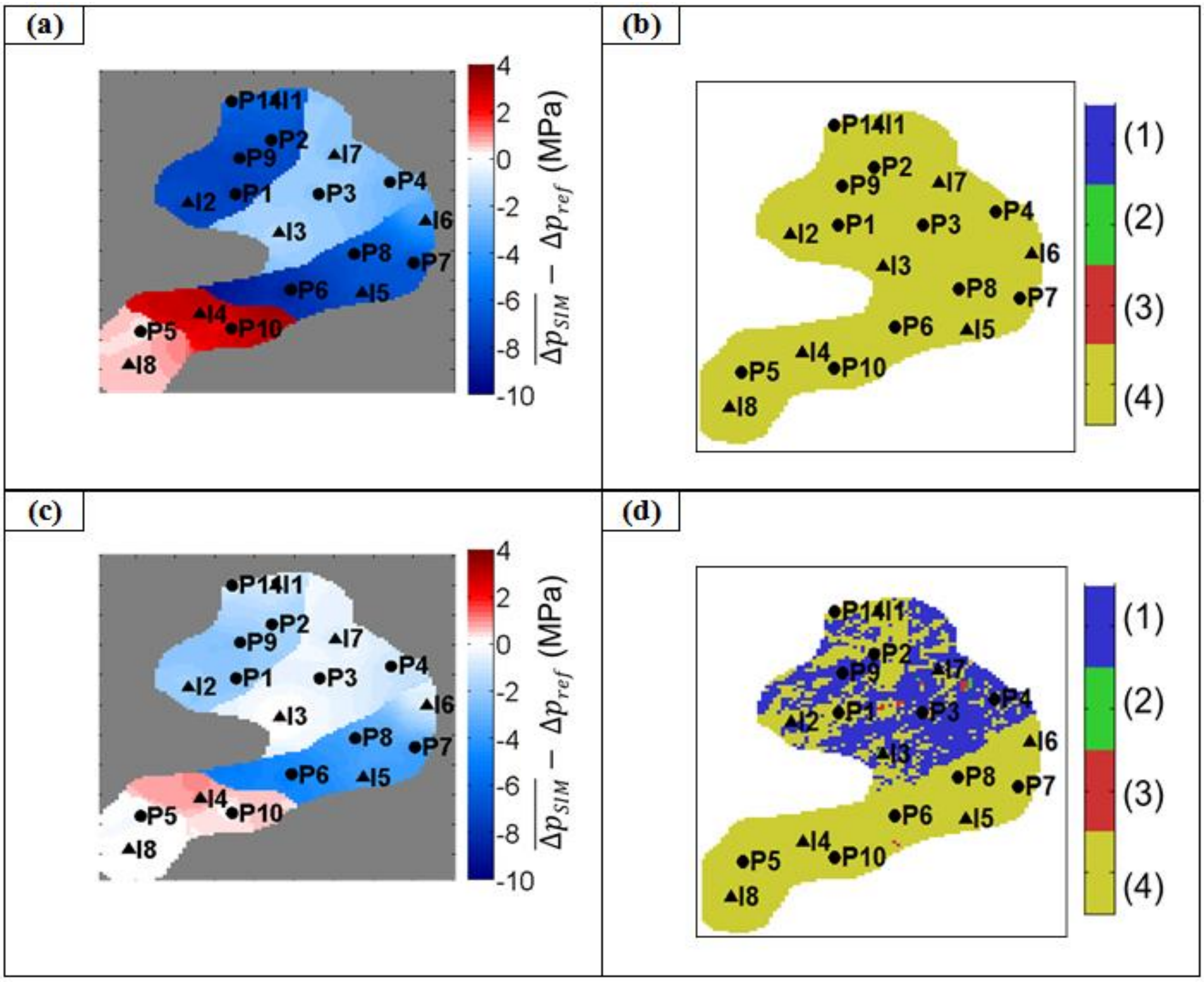

Figure B.5– Case II: before (top) and after (bottom) models selection. (a) Difference between the mean of 500 Δp maps from partially history-matched simulation models and the reference map, (b) 4 regions map comparing the 125 Δp maps from seismic without noise and the 500 Δp maps from simulation models, (c) Difference between the mean of 50 Δp maps selected from the methodology and the reference map and (d) 4 regions map comparing the 125 Δp maps from seismic and the 50 Δp maps from simulation. Every map is showing the 3rd layer out of 9.

- **Case III:**

Note that, in cases I and II the most precise simulation maps were selected, but in cases III and IV, it is selected the most precise seismic maps using history-matched models. Figure B.6a illustrates difference between mean of pressure change estimates from the 500 maps from seismic and the reference map. Observing Figure B.6b, the methodology showed that most of the grid blocks are in region 2, so, simulation is more precise than seismic data.

Using simulation data to constrain seismic data and choosing the most frequently selected seismic maps, it was obtained the mean of pressure change estimate of Figure B.6c. Although the maps of Figures B.6a and B.6c seem to be similar, they contain different information, as presents the maps of Figures B.7.

Differently of the cases I and II, after the selection Δp maps from seismic, simulation and seismic data do not agreed in many regions, as presented in Figure B.6d, although the number of grid blocks in region 1 has increased.

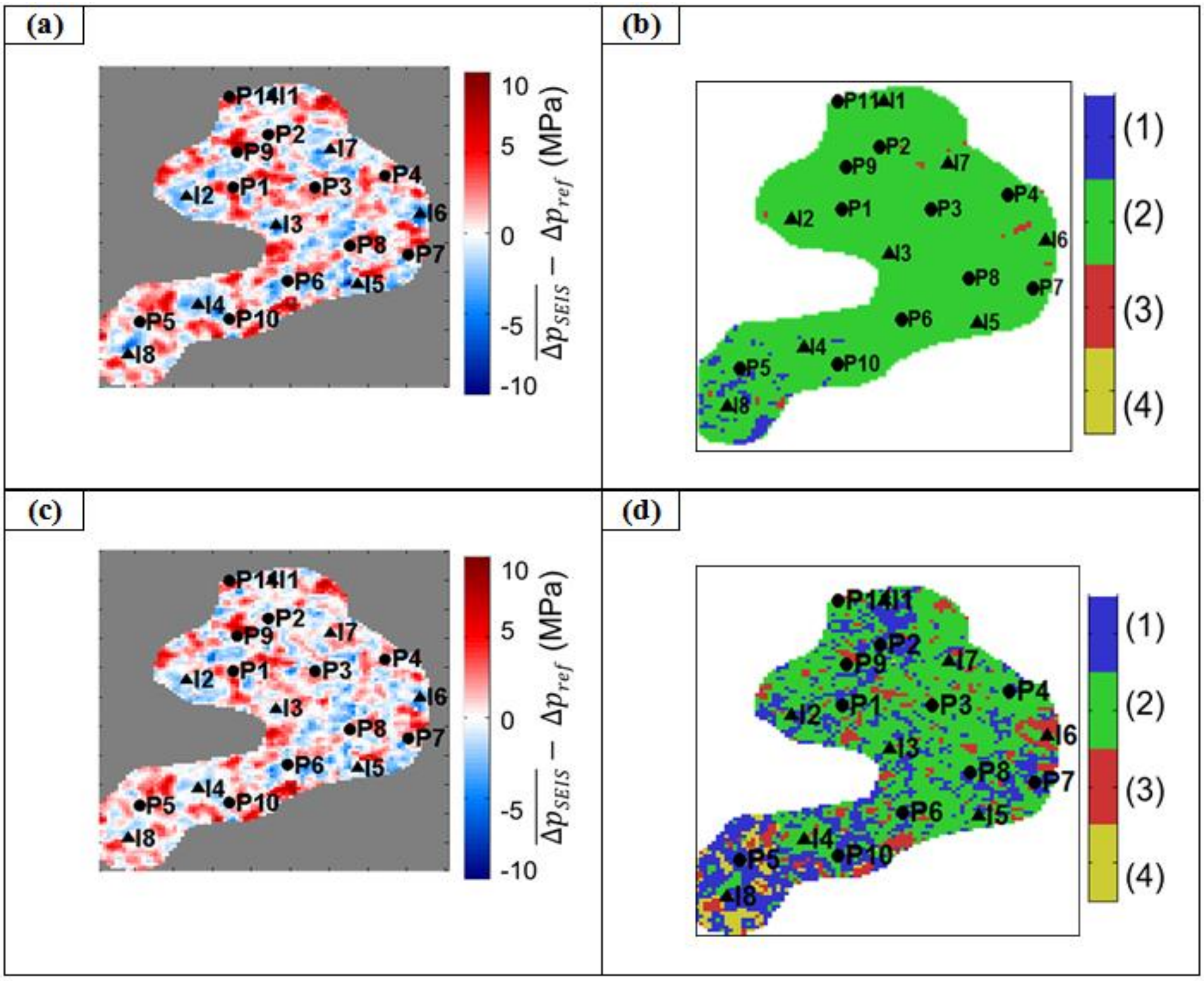

Figure B.6–Case III: before (top) and after (bottom) Δp maps from seismic selection. (a) Difference between the mean of 500 Δp maps from seismic with noise and the reference map, (b) 4 regions map comparing the 500 Δp maps from seismic with noise and the 97 Δp maps from history-matched simulation models, (c) Difference between the mean of 50 Δp maps from seismic selected and the reference map and (d) 4 regions map comparing the 50 Δp maps from seismic and the 97 Δp maps from simulation. Every map is showing the 3rd layer out of 9.

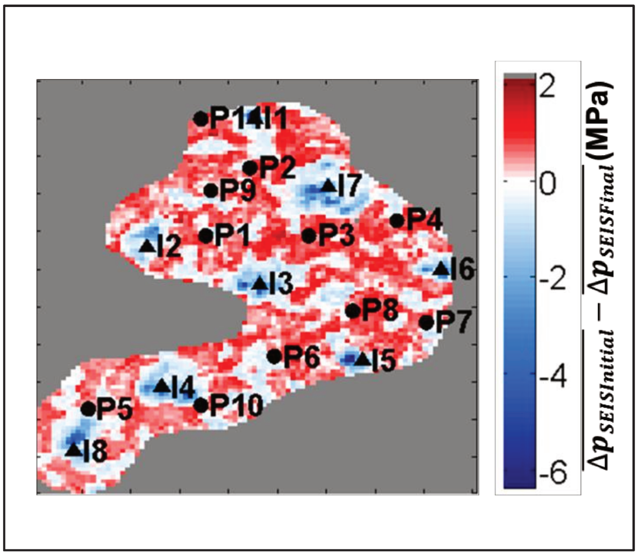

Figure B.7: Difference between the initial 500 Δp maps from seismic (Figure B.6a) and the 50 Δp maps from seismic selected by the methodology (Figure B.6c).

- **Case IV:**

A hypothesis for this result is that, if there are not accurate estimates from seismic or whether several maps from seismic are similar, there is not accurate information to be selected. Thus, to test this hypothesis, a fourth case was studied. In this last case, it is jointly applied the Δp maps with and without noise, therefore, 625 Δp maps from seismic is the initial information. Although these maps (500 maps with noise and 125 maps without noise) had been obtained from different procedures, this study is well-founded, as a validation test. The idea here is to observe if the methodology select Δp maps from seismic without noise.

As shown in Figure B.8c, pressure change estimates is more accurate than initial estimates (Figure B.8a). From Figure B.8c it can be observed that the majority of the Δp maps without noise were selected, which proved the hypothesis. Note that after the selection of the most accurate Δp maps the 4 regions map in Figure B.8.d presents different trends of the initial 4 regions map (Figure B.8.b) as, for instance, the majority of the blocks are now in region 1 and the drastic reduction of grid blocks in region 2.

4. Final remarks

The method studied showed to be promising for calibrating reservoir models, as well as to improve 4D seismic interpretation. Besides being a powerful diagnostic tool to guide the data integration, the methodology shown also to be useful to select the most accurate and precise maps, mitigating uncertainties.

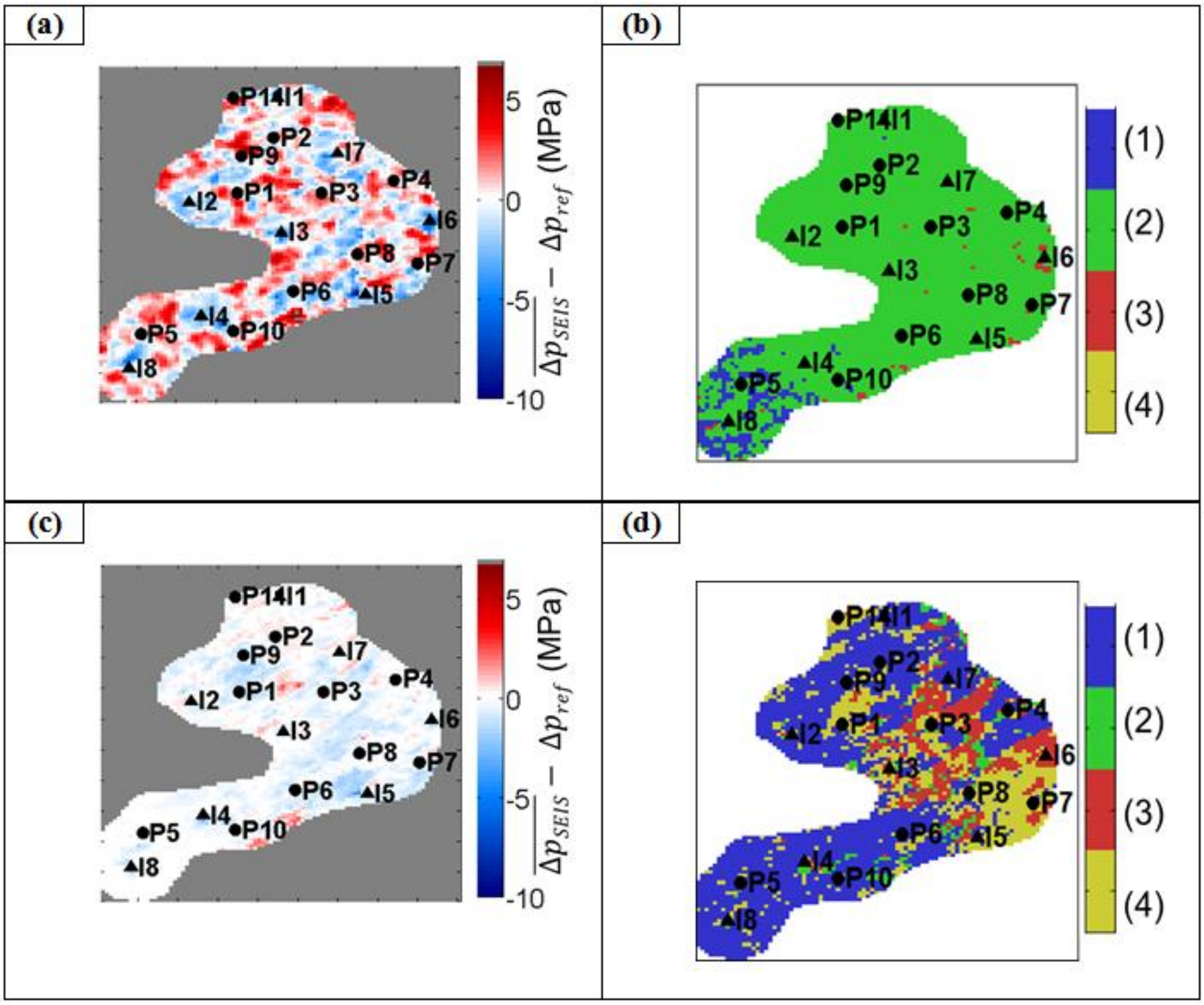

Figure B.8– Case IV: before (top) and after (bottom) ∆p maps from seismic selection. (a) Difference between the mean of 625 ∆p maps from seismic (with and without noise) and the reference map, (b) 4 regions map comparing the 625 ∆p maps from seismic and the 97 ∆p maps from history-matched simulation models, (c) Difference between the mean of 63 ∆p maps from seismic selected by the methodology and the reference map and (d) 4 regions map comparing the 63 ∆p maps from seismic and the 97 ∆p maps from simulation models. Every map is showing the 3rd layer out of 9.