EDUARDO DE ROSSI BORIN

HISTORY MATCHING ON REAL CASE STUDY USING ENSEMBLE SMOOTHER WITH MULTIPLE DATA ASSIMILATION

Trabalho de Conclusão de Curso apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do diploma de Engenharia de Petróleo.

SANTOS

2022

EDUARDO DE ROSSI BORIN

HISTORY MATCHING ON REAL CASE STUDY USING ENSEMBLE SMOOTHER WITH MULTIPLE DATA ASSIMILATION

Trabalho de Conclusão de Curso apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do diploma de Engenharia de Petróleo.

Área de concentração: Engenharia de Petróleo

Orientador: Prof. Dr. Leonardo Azevedo Guerra Raposo Pereira

Supervisor: Prof. Dr. Marcio Augusto Sampaio Pinto

SANTOS

2022

Catalogação-na-publicação

Borin, Eduardo de Rossi
History matching on real case study using Ensemble Smoother with Multiple Data Assimilation / E. R. Borin -- São Paulo, 2022.
103 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Minas e de Petróleo.

1.History matching 2.Data assimilation 3.Uncertain parameters 4.ES-MDA 5.Real response I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Minas e de Petróleo II.t.

AGRADECIMENTOS

Gostaria de expressar minha mais sincera gratidão a todas as pessoas que tornaram esta oportunidade possível para mim, concedendo uma experiência única e excepcional durante estes 6 longos anos de POLI.

Em primeiro lugar, quero agradecer aos meus pais pelo apoio sem fim nesta jornada em que me envolvi. Sem eles, nada do que fiz, dentro e fora do meu país de origem, teria sido possível. Obrigado por sua paciência e pelo apoio, principalmente durante aqueles tempos em que estávamos a mais de oito mil quilômetros de distância.

Em segundo lugar, agradeço ao meu professor e também coordenador do Mestrado em Engenharia de Petróleo do Instituto Superior Técnico (IST) da Universidade de Lisboa: Leonardo Azevedo, o responsável por me abordar com a oportunidade de estágio durante meu programa de Duplo Diploma. Ele deu todo o apoio necessário durante toda a experiência. Sinto-me honrado em ser reconhecido por você como um estudante capaz de assumir novas tarefas e sempre pronto para os desafios que ainda estão por vir. Gostaria também de mostrar minha gratidão a todos os amigos que fiz durante estes dois anos de estudos no Instituto Superior Técnico. Agradecimentos especiais principalmente a Francisco Graterol, Gonçalo Ferreira, Héllen Mariano, Manuella Campos e Youssef Omar. Foi com eles que compartilhei os bons e também os maus momentos durante este período, onde sempre me apoiaram em todos os aspectos possíveis. Espero que este tenha sido apenas o início de nossas amizades e que ainda possamos estar em contato no futuro.

Em terceiro lugar, gostaria de agradecer ao Sr. Jean-Louis Lesueur que me deu a oportunidade de fazer meu estágio de 6 meses e trabalhar na Equipe de Pesquisa e Desenvolvimento da TotalEnergies no CSTJF, localizado em Pau, França. Além disso, expresso também minha sincera gratidão ao Sr. Philippe Berthet e por todo seu amplo conhecimento compartilhado comigo sobre meu assunto de estágio e por ter confiado em mim durante este tempo. Sua assistência foi fundamental durante estes seis meses de trabalho dentro da TotalEnergies e este Trabalho de Conclusão de Curso não poderia ter sido desenvolvido sem sua orientação e experiência. Além disso, sou grato pelos novos contatos e amigos que fiz durante meu tempo lá: Abbass Sahily, Abderrahmane Yewgat e Eduardo Jose Airoso Barrela. Obrigado pelas grandes discussões que tivemos a oportunidade de ter. Um agradecimento especial a Mathieu Lapassade, que me deu um excelente apoio durante meu trabalho com o código Python utilizado para obter meus resultados para o trabalho.

Por último, mas não menos importante, quero agradecer aos meus dois amigos mais próximos que a universidade me deu: Anderson Castro e Jose Ricardo. Por todos os momentos que passamos juntos: noites inteiras acordadas estudando para as provas desafiadoras que tivemos durante a graduação ou indo para São Paulo durante a noite para fazer a entrevista para o Duplo Diploma e, até mesmo aproveitando a praia de Santos com algumas cervejas... estas são algumas das muitas experiências que vivemos juntos. Sem vocês, a POLI não teria sido “A POLI”. Portanto, muito obrigado, meus amigos!

RESUMO

O ajuste de histórico na engenharia de reservatórios é uma das etapas mais importantes do fluxo de trabalho de geo-modelagem durante a fase de produção de qualquer campo petrolífero. Esta técnica permite ter um bom conhecimento sobre como o campo está produzindo e também ter um modelo de previsão. O maior desafio é ser capaz de gerar modelos com o nível de complexidade do reservatório que representem a realidade da melhor maneira possível. O objetivo deste trabalho é obter uma correspondência sobre as propriedades vetoriais de produção em um estudo de caso real de um reservatório turbidítico. Para tanto, foi utilizado um método baseado em conjuntos para assimilar os dados dinâmicos, o método *Ensemble Smoother with Multiple Data Assimilation*, a fim de encontrar a melhor combinação de parâmetros incertos utilizados como entrada para o simulador do reservatório. O modelo geológico usado é de um conjunto de dados reais e os canais de areia elementares são gerados com ULIKE™: um processo de modelagem baseado em regras que seguem os princípio da caminhada aleatória. Os processos de assimilação de dados produziram conjuntos de modelos históricos que honraram tanto os dados dinâmicos quanto o modelo geológico. Além disso, também foi utilizado um software de correspondência de histórico baseado em desenhos experimentais e modelos proxy usando superfícies/polinômios interpolados e krigagem para inspecionar os impactos dos parâmetros incertos, principalmente os multiplicadores do índice de produtividade. Os resultados mostram que o uso da interpolação para estimar a resposta do reservatório é mais rápido e exige um custo computacional mais baixo, mas requer a confirmação com a resposta numérica de um simulador de reservatório.

Palavras-chave: Ajuste de histórico, Assimilação de Dados, ES-MDA, Parâmetros de incerteza, Resposta Real.

ABSTRACT

History matching in reservoir engineering is one of the most important steps of the geo-modelling workflow during the production phase of any oil field. This technique allows to have a good knowledge about how the field is producing and also have a forecast model. The biggest challenge is to be able to generate models with the reservoir's level of complexity that represents reality in the best way possible. The objective of this work is to get a match on production vector properties in a real case study of a turbiditic reservoir. For that matter, it was used an ensemble-based method to assimilate the dynamic data, the Ensemble Smoother with Multiple Data Assimilation method, in order to find the best combination of uncertain parameters used as input to the reservoir simulator. The geological model used is from a real dataset and the elementary sand channels are generated with ULIKE™: a rule-based modeling process based on the random-walk principle. The data assimilation processes produced ensembles of history matched models that honor both dynamic data and the geological model. Furthermore, it was also used an history matching software based on experimental designs and proxy-models using interpolated surfaces/polynomials and kriging to inspect the impacts from the uncertain parameters, mainly the productivity index multipliers. Results show that using interpolation to estimate the reservoir response is a quicker and demands a lower computational cost but requires the confirmation with the numerical response from a reservoir simulator.

Keywords: History Matching, Data Assimilation, ES-MDA, Uncertain Parameters, Real Response

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

ACRONYMS

2G&R	Geological, geophysical and reservoir
AHM	Assisted history matching
ANN	Artificial Neural Network
BHP	Bottom Hole Pressure
CSTJF	Centre Scientifique et Technique Jean-Féger
DOE	Design of Experiment
ED	Experimental Design
EKF	Extended Kalman Filter
EnHM	Ensemble-based History Matching
EnKF	Ensemble Kalman Filter
EOR	Enhanced Oil Recovery
ES	Ensemble Smoother
ES-MDA	Ensemble Smoother with Multiple Data Assimilation
GOC	Gas-oil contact
HM	History Matching
iES	Iterative Ensemble Smoother
MCMC	Markov Chain Monte Carlo
NTG	Net-to-Gross
OF	Objective Function
OPR	Oil Production Rate
RSM	Response Surface Methodology
SQA	Sequential Data Assimilation
WBHP	Well Bottom Hole Pressure
WCT	Water Cut
WOC	Water-oil contact
WOPR	Well Oil Production Rate

NOMENCLATURE

LATIN SYMBOLS	DEFINITION
C_D	Covariance matrix of measurement errors
C_{DD}	Auto-covariance of predicted data
C_M	Covariance matrix of model parameters
C_{MD}	Cross-covariance matrix between prior model parameters and predicted data
I_N	$N \times N$ identity matrix
K_n	Kalman gain matrix
N_a	Numbers of data assimilation process
N_d	Number of observation data
N_e	Numbers of ensemble (ensemble size)
N_m	Number of uncertain parameters
N_y	Dimension of state vector
$O_d(\cdot)$	Objective function (data mismatch)
$O_m(\cdot)$	Objective function (model mismatch)
c_f	Formation's compressibility
c_t	Total compressibility
d_{obs}	Observation data vector
d_{uc}	Uncertain (perturbed) observation data vector
m_{pr}	Prior uncertain parameters vector
p^n	Dynamical system state
y^n	State vector
z_d	Gaussian sample
ℓ	Iteration step
G	Normalization constant
H	Sensitivity matrix
$L(\cdot)$	Likelihood function
$O(\cdot)$	Total Objective Function
c	Fluid's compressibility
$f(\cdot)$	Probability distribution function
$g(\cdot)$	Forward model operator
k	Permeability in the direction of the flow
m	Uncertainty parameters vector
p	Pressure
t	Time

GREEK SYMBOLS	DEFINITION
α_ℓ	Inflation coefficient
$\mathcal{N}$	Gaussian Distribution
μ	Viscosity
ρ	Density
σ	Standard deviation
ϕ	Rock porosity

1 INTRODUCTION

Along with the large variety of studies that are made to fully understand the area in which a production activity will take place, reservoir studies and, more specific, fluid flow simulations play one of the most important roles in this whole hydrocarbon recovery process. To build a reservoir simulation model, it is necessary to have data about rock and fluid properties (petrophysical attributes), in order to characterize the fluid flow and the entire physical process behind it. Besides, it is also of utmost importance to bear in mind that the data available for characterization is not exact and entirely correct, and therefore is uncertain.

The reservoir simulation problem is one type of numeric simulation method used in Reservoir Engineering to estimate characteristics and predict behavior of an oil reservoir, following the techniques based on material balance, decline curves and Buckley-Leverett theory of displacement of non-miscible fluid. Indirect methods of evaluation are mainly used to acquire information about the subsurface formations' physical properties, with reservoir models being built inside an uncertainty parameters domain, meaning that any prediction made from these models are also considered uncertain.

Even though it may not be the best option for the problem in first sight, it is important to remember that obtaining information directly from the reservoir is an utterly difficult and costly operation, which is why it is needed to relay on dynamic information available from historical field production. Therefore, that is the reason history matching processes are being largely utilized in the industry in the last decades: improve the reliability of reservoir predictions using the models generated by available petrophysical data and statistical methods trying to incorporate field observation data in reservoir simulations models, allowing a better description of the uncertainty both in simulation predictions and reservoir parameters.

1.1 Objective

The main objective of this work is to obtain a model that can predict the future behavior of the reservoir in order to optimize the production process and oil recovery. For this, given the observation data from some of the production vector properties (i.e., bottom hole pressure, water cut and oil production rate) and the simulation results from the model previous created with the uncertain parameters as inputs, a history matching process is proposed using the ensemble smoother with multiple data assimilation (ES-MDA) to find the best combination of these uncertain parameters defined beforehand that gives the best match from the observed data.

This workflow was applied on a real scale turbiditic reservoir with elementary sand channels modeled using a state-of-the art rule-base modelling tool: ULIKE™. The area of study will be designed as the Alpha Field. The analysis and history matching will focus on two of the three panels of the field: the Central panel and the Eastern panel. Moreover, for the Central compartment of the reservoir, it was realized a comparison with EST, an internal TotalEnergies software for assisted history matching that uses kriged responses, with the aim to see some parameters impacts and to check for a possible non-linearity from one of the uncertain parameters with kriged responses surfaces, also trying to find better new ranges for the initial ensembles used for ES-MDA runs.

All the research work done and reported in this work was developed during a 6-month internship inside TotalEnergies' S.A R&D department team on the *Centre Scientifique et Technique Jean-Féger* (CSTJF). The ES-MDA algorithm was developed by the IT department of TotalEnergies and all the reservoir simulations were run inside the supercomputer *PANGEA III* using INTERSECT™ fluid flow simulator.

1.2 Motivation

Data assimilation problems estimate parameters given both vectors of observation and simulated data and try to minimize the difference between both. Thus, for reservoir history matching with uncertain parameters, it is also considered a data assimilation process: one with a high number of variables, few data considered exact and certain, and the multiphase fluid flow of porous media, which is a highly non-linear phenomenon that demands a high computational cost in order to obtain good results for the history matching steps.

Making a series of history-matched models is one method currently being used in the industry to measure the uncertainty in the reservoir model parameters. Decisions based on a single reservoir model may be misleading due to a model variability error. However, the generation of several history-matched models by itself does not provide accurate uncertainty assessment.

The energy industry has since welcomed ensemble-based approaches of history matching that evolved in this situation. They can produce numerous history-matched models and give an indication of how uncertain the results of the reservoir simulation are. Examples of this kind of history matching techniques include the Ensemble Kalman filter (EnKF: EVENSEN, 1994) and the Ensemble Smother with Multiple Data Assimilation (ES-MDA: EMERICK; REYNOLDS, 2013). Despite being reliable methods, both still rely on linear Gaussian assumptions, hypotheses which may generate unrealistic outcomes and produce erroneous results in certain reservoir models with complex non-linear and non-Gaussian geological environments.

1.3 Work organization

This work is divided into six chapters, including the Introduction (Chapter 1) and Conclusions (Chapter 6). Bellow, the chapters are briefly described.

- Chapter 2 provides an overview of the theoretical background that has been used to build this final graduation work. It discusses the ideas of reservoir simulation and history matching with uncertainty quantification utilizing data assimilation techniques and proxy models;

- Chapter 3 provides details on the approach used to create this work and the methodology used;
- Chapter 4 includes a 2G&R (i.e., Geological, Geophysical and Reservoir characterization) synthesis of the study case area and the description of the complete procedure;
- Chapter 5 shows the results obtained from the real case study reservoir, evaluating the ensemble-based data assimilation method used in this work for the history matching. Moreover, it also contains results from another history matching approach using mainly interpolation, making a comparison with the before used method.

2 BIBLIOGRAPHIC REVIEW

2.1 Reservoir Fluid Flow Simulation

Numerical tools are used in different scientific areas in order to replicate, in the best and most accurate way possible, the real case scenarios of physical phenomena. These tools demand a good understanding and description of the physical system under investigation. For that, it is applied mathematical models with complex equations, based on theories and laws that describes the problem which is being investigated. For that matter, numerical simulations are used to solve these models, as they are considered complex problems with a considerate number of unknows and equations. Solving these equations would be unfeasible to do without the aid of a computer. Within the reservoir engineering area, the scope of interest is the petroleum reservoir system, consisting of multiple geologic formations, with heterogeneities and with the possibility of containing hydrocarbons that are economically viable to be extracted.

The reservoir model is a mathematical representation of a specific volume of rock incorporating all the characteristics necessary to analyze the dynamic behavior of fluids moving through the porous medium. This description is usually developed using a complex workflow which involves a great number of data sources that span a large variety of spatial and temporal scales, going from geological history of the basin on the surroundings with seismic and well logs to rock samples extracted from exploration and production wells (PYRCZ; DEUTSCH, 2014) .

Even though simplifying hypothesis can be assumed, such as homogeneous and isotropic porous media with linear/radial one-dimensional flow displacement to describe the reservoir behavior, problems related with the energy industry are often linked to complex models with nonlinear partial differential equations relating pressure and saturation changes throughout time and space, in addition to complex multidimensional, multiphase and multicomponent flow models (ROSA; CARVALHO; XAVIER, 2006).

Given the high complexity of the subsurface geology, it is impossible to successfully run a reservoir simulation in which the model does not contains uncertainty. In fact, the fluid flow inside the pore system occurs in such a detailed level that is unconscionable to model it in perfection, resulting in an immense increase in computational cost to run the simulations and a higher degree of uncertainty related to these simulations' outputs. Moreover, the scarcity of high-cost direct measurements and the big extension of the reservoir also contributes to these uncertainties in the models, and the reservoir behavior may not be precisely simulated as expected. In summary, the uncertainty reflects our lack of understanding about the subsurface geology. For that, the reservoir model has to be created including all available data possible in order to reproduce properly the fluid flow of the system.

Before starting the production of any oil field, it is necessary to know the quality of the reservoir in which the production wells are being drilled into. In that case, in addition to the seismic data obtained during exploration phases and also some core information retrieved from appraisal wells, reservoir models are built into state-of-the-art software's that tries to simulate all the conditions in which the actual reservoir is at the very beginning of production, with pressure data, initial saturation of fluids, WOC (water-oil contact) and GOC (gas-oil contact). With that information, the software proceeds to the fluid simulation part, where numerical methods are applied in order to solve the highly nonlinear set of differential equations that describe the fluid flow in porous media. The basics of reservoir fluid flow are very well described in the literature, such as in Gupta et al. (1991), Heimsund (2005) and Gerritsen & Durlofsky (2005).

2.1.1 Concepts used in reservoir fluid flow simulations

Mass transfer is the main processes that occur in fluid flow. Up to three immiscible phases (water, oil, and gas) can flow at the same time, with mass transfer between them. The flow process is influenced by gravity, capillary, and viscous forces. All of these forces must be accounted for in the model equations, as well as an arbitrary reservoir description in terms of geological heterogeneity and geometry (PEACEMAN, 1977). For each phase, the differential equations are constructed by combining Darcy's law with a simple differential material balance, resulting in the hydraulic diffusivity equation.

By being able to solve this equation, it is possible to identify all changes in pressure in the porous media versus space and time. The first derivation for the hydraulic diffusivity equation is considering a single-phase flow during a certain period of a porous medium which is a box on a tri-dimensional space (x, y and z-axis). In this case, Eq. (2.1) is derived on the form as a diffusivity equation which governs the three-dimensional linear flow though the porous medium fully saturated with only one fluid (ROSA; CARVALHO; XAVIER, 2006):

$$\frac{\partial}{\partial x}\left(\rho\frac{k_x}{\mu}\frac{\partial p}{\partial x}\right)+\frac{\partial}{\partial y}\left(\rho\frac{k_y}{\mu}\frac{\partial p}{\partial y}\right)+\frac{\partial}{\partial z}\left(\rho\frac{k_z}{\mu}\frac{\partial p}{\partial z}\right)=\phi c_t\rho\frac{dp}{dt}, \tag{2.1}$$

where ρ is the fluid's density, k_d is the permeability of the porous medium in the flow direction (x, y or z-direction), μ is the fluid viscosity, p is the pressure, t is the time, ϕ is the rock porosity, c_t is the total compressibility – corresponding to the sum of the fluid compressibility c and effective compressibility of the formation c_f.

To solve the mathematical model, values of independent parameters must be obtained that concurrently check the governing equations and boundary conditions (i.e., external and wells conditions). Because it is implicitly dependent on density, permeability, viscosity, and compressibility, Eq. (2.1) is a nonlinear differential equation. It is impossible to solve this type of equation analytically (PEACEMAN, 1977). Furthermore, depending on the study's aims, the addition of mathematical statements makes the solution of the mathematical model more complex. Multiphase and multicomponent systems, as well as more particular and complex processes like steam injection, polymer injection, and thermal EOR, are examples (ROSA; CARVALHO; XAVIER, 2006).

It is important to solve differential equations subject to the right boundary conditions in order to utilize them to forecast the behavior of a reservoir. The standard methods of mathematical physics can only find solutions for the simplest scenarios with homogenous reservoirs and very regular boundaries (such as a circular barrier around a single well). Numerical methods performed on high-performance computers, on the other hand, are exceedingly broad in their applicability and have proven to be effective in obtaining answers to extremely complex reservoir problems (PEACEMAN, 1977). To solve reservoir simulation models, most commercial simulators use finite difference approaches. The nonlinear partial differential equation is replaced by the finite-difference equivalent, which is derived from a Taylor series expansion of the function at a given location (ROSA; CARVALHO; XAVIER, 2006). Time is discretized into many time steps and the spatial domain is divided into a finite number of cells, commonly known as grid blocks. As a result, the numerical solution calculates solutions for each discrete spatial element within each discrete time interval.

2.2 History Matching

When working with reservoir simulations, it is necessary to infer the behavior of a real reservoir from the performance of a model – which may be physical and/or mathematical - of that reservoir. In this case, a mathematical model of a physical system is a set of partial differential equations, together with a set of boundary conditions that we believe sufficiently represents the significant physical processes occurring in that system.

History matching is considered an inverse problem, because from a set of observed data, one predicts the factors that caused these observations. In other words, from the resulted effects of a given analysis, we use history matching to calculate the causes and what main parameters are associated with these direct observations. A classic example of this application is for weather predictions and atmospheric modeling (CHEN et al., 2021; CHO et al., 2020).

Over the last decades, due to the need of inferring the behavior of a real reservoir through numerical model results and make forecasts from historical data, history matching began to be used in petroleum literature. To build a reservoir simulation model, not only it is required to have data about rock and fluid properties (i.e., petrophysical attributes), in order to characterize the fluid flow and the entire physical process behind it, but also it is of utmost importance always bear in mind that the data available for characterization is inaccurate and have uncertainty.

With the reservoir simulation, it is possible to estimate characteristics and predict behavior of an oil reservoir, following the techniques based on material balance, decline curves and Buckley-Leverett theory of displacement of non-miscible fluid (E BUCKLEY; LEVERETT; AIME, 1942). That said, as much as numerical methods and approximations generate satisfactory results, the lack of knowledge about the geology that composes the reservoir is the main source of uncertainty in this activity. Thus, reservoir models are built inside an uncertainty parameters domain, meaning that any prediction made from these models are also considered uncertain. Even though it may not be the best option for the problem in first sight, it is important to remember that obtaining information directly from the reservoir is an utterly difficult and costly operation.

Knowing the limitations that reservoir models can give and the uncertainties it generates after the incorporation of production data and with prior geostatistical knowledge, it was necessary to come up with a solution that could manage the uncertainty quantification of future reservoir performance. Uncertainty quantification in reservoir performance predictions and descriptions are made mainly to measure results and manage risks (i.e., decision-making), after the generation of multiple history matched models and their assessment of uncertainty (EMERICK, 2012).

Generally speaking, the history matching problem can be thought as an optimization problem, in which the main objective is to find the best production vector of a certain reservoir model with a certain number m of model variables such that:

$$m = F^{-1}(d_{obs} + \epsilon). \tag{2.2}$$

In Eq. (2.2), d_{obs} is the vector of observed data, with ϵ being the uncertainty of the measurements, and $F(\cdot)$ is the forward model that maps the data into the model domain and therefore can be used to predict the reservoir behavior (i.e., the reservoir fluid flow simulator) (OLIVER; CHEN, 2010). The parameters chosen beforehand are determined by the model complexity (e.g., grid size, fluid composition, flow physics), always demanding a deep knowledge of the study area and the reservoir processes. Even though the main idea of a history matching process is to obtain less mismatches from the model response given a certain set of parameters vectors, in other words, a direct goal, there are innumerous algorithms that tackle this problem and try to assess uncertainty. These algorithms are mainly separated into two main categories: manual history matching or assisted history matching.

2.2.1 History matching algorithms

Manual history matching was the first method applied in reservoir engineering with good engineering judgment and workflows that has been developed with years of experience in order to analyze the quality of the reservoir model and its parametrization. The work presented in Williams et al. (1998) is a well-structured approach to history matching. The strategy starts with an effort to match pressures at the large (field) scale by modifying a few important parameters, including the permeability multipliers, aquifer transmissibility, rock compressibility, and the ratio of vertical to horizontal permeability. The properties of individual flow units or layers are altered once the pressure has been properly matched at the field level, and then the properties of wells, or cells nearby wells, are adjusted. Once the pressure has been matched, Williams et al. (1998) suggest making modifications in other variables in order to match water arrival times and water cut, for example, relative permeability curves and vertical transmissibility values.

Despite the convenience of this method, the experience of the engineer and the budget's size have a significant role in how well this form of history matching turns out. Since reservoirs are typically quite diverse, a typical reservoir simulation model has hundreds of thousands of grid blocks to estimate reservoir parameters in high resolution. Computers are used to automatically alter the parameters because manual history matching is frequently unreliable over prolonged periods of time and is always associated with significant uncertainty (RWECHUNGURA; DADASHPOUR; KLEPPE, 2011). Therefore, it is more challenging to create a complex geological model which better depicts the reality, thus, limiting the prediction power of this method.

The assisted history matching methods come to bypass these problems. Those methods also relate in finding the best match of a reservoir model, but now, automatized algorithms are developed with the purpose to minimize the objective function, which calculates basically the quadratic deviation between the simulated data and the observation one. A possible way of solving these types of problems is to use iteratives algorithms that will calculate the unknown parameters by successive approximations, trying to find the locals minima of the complex objective function, or even a global minimum (RWECHUNGURA; DADASHPOUR; KLEPPE, 2011).

An efficient class of optimization methods which have been used for a very long time are gradient-based history matching algorithms. In general, they are based on the calculation of the gradient of the objective function in respect of the model parameters, solving the adjoint state (COURANT; HILBERT, 2007). The history matching problem using adjoint approach is handled as an optimal control problem using the unknown model parameters as the control variables, with the objective function being minimized while being constrained to have the state variables follow the specified reservoir model (KALETA et al., 2010). The gradients methods have been used extensionally in reservoir history matching, with examples of applications in Anterion et al. (1989), Bissel et al. (1992) and Volkov et al.(2018).

The minimization procedure is typically conducted using first-order gradient-based minimization techniques, like the Gauss-Newton and Levenberg-Marquardt algorithms, which avoid the explicit computation of the Hessian (second derivative) (KALETA et al., 2010). However, when using automatic history matching with gradients methods, it is necessary to calculate the derivatives while doing the reservoir fluid flow simulation. These processes are usually difficult to be made inside complex simulators, being necessary to come up with other implementations of the adjoint methods for derivative calculation. An example of an implementation for this is shown at Rodrigues (2005), which can be applied even outside the framework of standard optimization algorithms, for instance, with history matching integrating 4D seismic data attributes, such as: fluid contacts evolution with time; compartmentalization pressure estimates and fault seal; oil bypass locations, all in order to obtain better results and description of the reservoir (EMERICK; MORAES; RODRIGUES, 2007).

Nevertheless, the fact that such methods typically converge to a local minimum in the objective function rather than the global minimum presents a possible issue. Inverse problems are known to be often ill-posed. In other words, these kinds of problems either have no solution in the desired class, or have many (two or more) solutions, or the procedure to obtain them is unstable, meaning that arbitrarily small errors in the measurement data may lead to indefinitely large errors in the solution (KABANIKHIN et al., 2008). It is well recognized that because of how ill-posed the reservoir history matching inverse problem is, there may be numerous local solutions, but only those that create an adequate fit to the data are interesting (GOMEZ; GOSSELIN; BARKER, 2001).The Jacobian matrices of the system are typically available in reservoir models since they are used in the Newton-Raphson iteration during the forward simulation. Still, applying the adjoint equations involves a significant amount of programming and necessitates having access to the numerical simulation code (KALETA et al., 2010).This suggests that there is a need for gradient-based, adjoint-free optimization techniques or another approach on this history matching problem.

Alternatively, there are methods that do not require the calculation of derivatives. Algorithms that search for the global optimum without using derivatives are classified as stochastic methods of assisted history matching. Some of the main algorithms are: Evolutionary Strategy, Genetic Algorithm, Simulated Annealing and Particle Swarm Optimization (RWECHUNGURA; DADASHPOUR; KLEPPE, 2011). The main drawback is that they are expensive and require a significant number of function evaluations. Furthermore, when a reservoir numerical simulator is used in such function assessments, this method could become unfeasible.

In addition to all the methods previously cited, data assimilation processes have also been largely used in the last couple of decades (CARRASSI et al., 2018b; NÆVDAL; MANNSETH; VEFRING, 2002; VAN LEEUWEN; EVENSEN, 1996). The main goal is also to merge the information from observations into a numerical model while introducing uncertainty quantification, by using multiple samples drawn from conditional distributions from parameters available from the observation data. These methods relay on the use of different types of data combining readily with any type of numerical reservoir simulator.

This work focuses on the ensemble methods, where they differ from common assisted history matching algorithms by mainly providing multiple simultaneously history matched models and calibration of both state (pressure, saturations) and model (porosity, log-permeability) variables (OLIVER; CHEN, 2010).

2.2.2 Formulation of history matching problem for petroleum reservoirs

It is possible to use the Bayesian formulation inside the history matching problem, mainly when the model parameters are poorly known and conditioned to data considerate to be inaccurate. With Bayes' theorem, one is allowed to write the conditional probability distribution function (pdf), $f(m|d_{obs})$, of a N_m-dimensional vector of parameters, m, given a N_d-dimensional vector of observations, d_{obs}, as

$$f(m|d_{obs}) = \frac{f(d_{obs}|m)f(m)}{f(d_{obs})} = \frac{f(d_{obs}|m)f(m)}{\int_D f(d_{obs}|m)f(m)dm} = GL(m|d_{obs})f(m). \quad (2.3)$$

In Eq. (2.3), $f(m)$ is the prior pdf of the vector containing the parameters' models and $f(d_{obs})$ is the pdf of the vector of observation data. $f(d_{obs}|m)$ is the conditional pdf of d_{obs} given m. The pdf can also be written as a likelihood function, which is denoted by $L(m|d_{obs})$, and G is a normalizing constant. If the prior's pdf and measurement errors can be assumed to be Gaussian distributed (EMERICK, 2012), the conditional pdf of model parameters given observation data can also be written as:

$$\begin{aligned} f(m|d_{obs}) = Gexp&\left\{-\frac{1}{2}\left(m - m_{pr}\right)^T C_M^{-1}\left(m - m_{pr}\right)\right\} \\ &\times exp\left\{-\frac{1}{2}(g(m) - d_{obs})^T C_D^{-1}(g(m) - d_{obs})\right\} \\ &= Gexp\left\{-\frac{1}{2}\left(m - m_{pr}\right)^T C_M^{-1}\left(m - m_{pr}\right)\right. \\ &\left.-\frac{1}{2}(g(m) - d_{obs})^T C_D^{-1}(g(m) - d_{obs})\right\} = Gexp\{-O(m)\}, \end{aligned} \quad (2.4)$$

where

$$O(m) = O_m(m) + O_d(m), \quad (2.5)$$

being

$$O_m(m) = \frac{1}{2}\left(m - m_{pr}\right)^T C_M^{-1}\left(m - m_{pr}\right), \quad (2.6)$$

and

$$O_d(m) = \frac{1}{2}(g(m) - d_{obs})^T C_D^{-1}(g(m) - d_{obs}). \quad (2.7)$$

In the equations above, C_M is the $N_m \times N_m$ prior covariance matrix of model parameters; C_D is the $N_d \times N_d$ covariance matrix of measurement errors and $g(m)$ is the vector of predicted data using the created model for a certain vector m. In history matching, $g(m)$ are the results obtained with the reservoir flow simulation run. In Eq. (2.4), $O(m)$ is what is known as the objective function, in which the main objective inside a history matching problem is to find a minimum of this objective function in order to maximize the posterior pdf, $f(m|d_{obs})$. It is important to notice that the objective function is made of two parts: the data mismatch part, $O_d(m)$ (Eq. 2.6)), and the model mismatch part, $O_m(m)$ (Eq. (2.7)). Finally, it is common to assume m_{pr} as a mean value of the vector of model parameters.

Hence, the history matching problem for reservoirs become an optimization problem, in which the objective is to maximize the posterior probability distribution function of the defined parameters conditioned to field measurements (i.e., observed data). The problem of characterizing the uncertainty in the reservoir model parameters is reduced to the problem of sampling the probability density function from the vector of model parameters (EMERICK, 2012). Those are the main objectives of the statistical and numerical methods applied to solve the optimization method of these history matching problems for oil reservoirs (EVENSEN, 2018; TIERNEY, 1994).

2.2.3 History matching using proxy models – Response-Surface Methodology (RSM), Design of Experiment (DOE) and Artificial Neural Networks (ANN)

Even though with the advance of simulation resources and computational power, it is still desirable, and sometimes necessary, to have a method that solves inverse problems in a quick and efficient way, reproducing as best as possible the uncertainty of the problem, considering all data available and its quality. The use of computationally efficient proxy-models is therefore receiving a lot of attention as researchers continue to search for ways to minimize the computational burden associated with numerical fluid flow simulations.

Proxy-models are referred to as a mathematical or statistical defined function which replicated a certain model output for different input parameters. In several scientific fields, proxy models are used to approximate mathematical modeling. In the scope of history matching, common use cases include: sensitivity analysis of uncertain variables; probabilistic forecasting and risk analysis; field development planning and production optimization and history matching (ZUBAREV, 2009). Proxy-models and strategies for design of experiments are frequently utilized in sensitivity analysis. The conventional one-parameter-at-a-time method for linear sensitivity assessments as well as sophisticated experimental designs that can resolve correlation and higher order effects are examples of application situations.

Yeten et al. (2005) examined sensitivity analysis and probabilistic forecasting as they investigated various experimental designs and proxy-models and their capacity to assess uncertainty in field performance. They claim positive outcomes when using space-filling designs and proxy-models based on polynomial, kriging, and splines. The same kinds of proxy-models produced worse outcomes for conventional designs that sample at the edges of the uncertainty area (e.g., Placket-Burman, central composite and D-optimal designs).

History matching is frequently a challenging process that calls for numerous simulations to investigate the problem space and come up with workable answers. Proxy-models are appealing instruments to utilize as an effective replacement for complete reservoir simulation in the event that they can accurately represent important simulation output parameters. Nevertheless, the high quality of these models is directly related to the input dataset. Highly non-linear output is dealt with in reservoir simulation. Therefore, it may not be possible to design a suitable proxy-model using an input dataset with experiments spread equally across the uncertainty domain (ZUBAREV, 2009).

Eide et al. (1994) introduced the idea of response surfaces and experimental design for automatic reservoir history matching. The main idea was to estimate response surfaces based on a set of reservoir simulations with different combination of reservoir parameters and estimate a simplified relation between reservoir simulator input and output (response). The authors managed to apply the workflow in a synthetic example with four input variables and five response variables. The quality of the response surfaces depends on the experimental design and on the model used for it. Additionally, Eide et al. (1994) concluded that a Bayesian strategy that uses prior knowledge of the distribution and the response surface can be used to focus the search on the most likely area of the input variables after matching and to add uncertainty to prediction estimations.

Amudo et al. (2009) described their practical experience in constructing experimental designs and using response surface models techniques in reservoir simulation studies. Their concepts and results were obtained after studying fifteen reservoirs situated in four different fields and at distinct stages of maturation. The authors conclude that is necessary an iterative and on-going process throughout the experimental design workflow in order to analyze the impacts of different choices of uncertain parameters and how they behave in the interpolation process during the response surface modeling. Besides, even with the best parameter conditioning, the combination of extremes can occasionally lead to non-physical experiments, which can require a lot of computer work and delay any further study.

There have been studies that integrated the Response-Surface Methodology for history matching and probabilistic forecasting of reservoirs. Slotte et al. (2008) discussed the theory for assisted history matching and uncertainty evaluation using a Bayesian framework, conditioning the posterior pdf on the *a priori* geological information through ensembles of reservoir models sampled by a Markov Chain Monte Carlo (MCMC) technique. Slotte et al. (2008) built proxy functions for the flow simulator's output for each measurement that enters a global goal function. The proxy functions are built using multidimensional kriging and polynomials. To raise the caliber of the proxy functions, an iterative loop was undertaken in which ensembles of reservoir models are sampled from the posterior pdf.

Wantawin et al. (2017) used a design of experiment (DOE) to investigate the relationship between those uncertain parameters and the variations known as response parameters. The authors developed a workflow combining history matching solutions with pdf forecasts into an integrated proxy-based approach that searches for both processes, evaluating them simultaneously. The combined workflow is an iterative approach, as the original proxy-model is continuously updated into higher degree of polynomial. Compared to the frequently utilized quadratic form, using higher-degree polynomial equations seems to offer the advantage of providing a broader set of HM solutions inside the uncertain parameter space (WANTAWIN et al., 2017).

Artificial neural networks (ANN) are another proxy model. The ability of ANNs to approach any continuous function with the appropriate precision is what enables them to be applied in a very efficient manner for various models in a variety of industries, such as the aerospace for autonomous control and the defense for facial recognition. In fact, ANNs have taken the place of the reservoir simulator in several history matching and optimization studies to speed up the operations, e.g., Costa et al.(2010), Foroud et al. (2014), and Guerillot et al. (2017; 2016).

Bruyelle et al. (2019) use an artificial neural network-based proxy model and assessed and contrasted with proxy models typically used for history matching, such as polynomials or kriging techniques. The suggested method offers precise prediction outcomes that speed up and enhance the history matching process. An ANN was defined for each observed data that provides the mean and standard deviation of the data for a group of geological realizations. To construct an approximate definition of the goal function, all ANN are concatenated. To minimize the approximation of the objective function, a global optimizer is employed. The methodology was applied in a synthetic case study – the Brugge field – where the findings demonstrate that ANN has significantly superior prediction power than kriging or quadratic polynomial proxies (BRUYELLE; GUÉRILLOT, 2019).

In a recent literature review about the use of proxy modelling in numerical reservoir simulation, Jaber et al. (2019) mention that it is better to understand the main aim of the proxy modeling (e.g., prediction, optimization, uncertainty) in order to choose the right approach for each of these categories. Successful use of proxies is documented in the review for a variety of reservoir simulation modeling applications, including assisted history matching, reservoir performance prediction, uncertainty analysis, and optimization. Traditionally, these applications have relied on reservoir simulation modeling, which is expensive in terms of the data resources used and the time required. While proxy modeling can give a quick assessment of the response with sufficient accuracy and simplify a complex procedure with respect to unclear parameters in the interested region. High nonlinearity, however, makes it more difficult to create a reliable proxy model (JABER; AL-JAWAD; ALHURAISHAWY, 2019). Building the proxy model involves many steps, one of which is evaluating the proxy model's quality. Therefore, it is important to verify the resulting solution obtained from the proxy model checking with the reservoir simulator. If the results from the simulator are away from the results obtained by the proxies, the input data must be amended by new data and the process becomes an iterative one, where finding the best proxy model is the main objective.

2.3 Ensembled-based methods for data assimilation

The estimation of the status of a system, such as the atmosphere, the ocean, or any part of the Earth system or its entirety, at any arbitrary past, present, or future time is the problem that is supposed to be solved using data assimilation methods. There are two complementing sources of information: the observations and predicted data through numerical simulations, but they both have errors. By identifying connections between the model and the observations and utilizing the knowledge contained in each, data assimilation provides the conceptual and methodological tools to address the issue, as it is conveniently formalized as a discrete-model/discrete-observation estimation problem (CARRASSI et al., 2018a).

In the scope of geoscience, innumerous research about data assimilation have contributed for the development of this area, being possible to combine data with geological models in a more precise and efficient way. One category of the different techniques inside data assimilation is statistical schemes, based on the theory of estimation and error covariances calculations of observations and model predictions in order to find the best linear combination between these two (JUNG et al., 2018).

The ensemble-based methods follow a Monte Carlo implementation of a data assimilation scheme following the original Kalman Filter (KF) (KALMAN, 1960) equations and updates steps, having all the correlations between model forecast and observation data estimated from the ensemble of models and their respective data mismatch. They are used as an alternative to the approximate error covariance evolution equations used on data assimilation for non-linear models, such as the extended Kalman Filter (EKF), to compute forecast errors estimates with a significantly lower computational cost. In other words, ensembles methods rely on the stochastic integration of an ensemble of model states followed by observational updates using the forecast error covariance implicitly by the ensemble spread (BRASSEUR, 2011).

For history matching, the first application of an ensemble method was presented by Lorentzen et al. (2001), where Ensemble Kalman Filter (EnKF) was applied in a two-phase flow in a wellbore during drilling operations, to improve predictions of pressure behavior. In terms of petroleum reservoirs application, Nævdal et al. (2002) also used EnKF to update the permeability fields for near-well reservoir models. Nowadays, many ensemble methods are applied even more in a more geological complex environment with multiple different goals, such as:

- maximize the net present value of a hydraulically fractured horizontal well inside a shale gas reservoir using an ensemble-based optimization method (XUE et al., 2021);
- quantify uncertainty and calibrate a reservoir model using dimension reduction techniques to enhance computation time (TADJER; BRATVOLD; HANEA, 2021);
- combine the workflow with a feature-oriented parametrization of geophysical data with ensemble-based history matching in order to characterize a fractured carbonate reservoir (ZHANG et al., 2022).

2.3.1 Kalman Filter (KF) – the solution for the linear and Gaussian case in sequential data assimilation problems

In reservoir engineering it is common to work with data available sequentially in time (e.g., production, pressure and time-lapse seismic data). The model forecast should be combined with the observation data in the process of updating the dynamic system by incorporating these data sequentially in time. This process is called sequential data assimilation (SDA) and falls into the scope of reservoir history matching problems, being described with a Bayesian framework and conveniently changing the parameter estimation problems to a parameter-state estimation one, the latter being discretized in time (EVENSEN, 2009, cap. 4).

Introduced by Kalman (1960), the KF is a powerful recursive filter that accurately determines the state of a linear dynamical system from a collection of measurements. The filter is based on a model equation, in which the system's present state is connected to an uncertainty (expressed by a covariance matrix), and an observation equation, which connects a linear combination of the states to measurements. There is uncertainty surrounding the measurements as well.

The equations used in the KF can be obtained in several ways. One of them mentioned in Stengel (1994) is by solving a weighted least square problem imposing additional constraints that the measurement noise should be independent in time (i.e., there should be no correlation between the model noise and the measurement noise) and the weighting depends both on the uncertainty in the state variables and the measurements. Another derivation, which is most used, is done using Bayesian methodology, as shown in Cohn (1997). For that, the main requirements that needs to be done is that both the model and measurement noise are Gaussian and that the prior state are Gaussian distributed and unbiased. In that way, it is possible to update the prior conditional probability distribution of the system, based on forecasts and historical data. In that case, the pdf of the state vector y^n would follow Eq. (2.4) formulation, noticing the equivalence between simultaneous and sequential data assimilation and under the KF hypothesis of Gaussian prior and linear relationship between state and predicted data.

The pdf of the state vector in a certain time t_n is, thus, given as (EVENSEN, 2009, cap. 4):

$$f(y^{n,a}) = exp\left\{-\left(\frac{1}{2}(y^{n,a} - y^{n,f})^T\left(C_Y^{n,f}\right)^{-1}(y^{n,a} - y^{n,f}) + \frac{1}{2}(H_n y^{n,a} - d_{obs}^n)^T(C_D^n)^{-1}(H_n y^{n,a} - d_{obs}^n)\right)\right\}, \tag{2.8}$$

where, for each t_n: C_Y^f is the error covariance matrix of the model state; d_{obs} is the observation data vector; H is the sensitivity matrix, which describes the linear relation between state vector and predicted data, i.e.,

$$d^{n,f} = H_n y^{n,f}, \tag{2.9}$$

and C_D is the covariance matrix of the measurement errors.

Furthermore, the state vector and its covariance matrix are updated sequentially in time using:

$$y^{n,a} = y^{n,f} + K_n(d_{obs}^n - H_n y^{n,f}), \tag{2.10}$$

and

$$C_Y^{n,a} = \left(I_{N_y} - K_n H_n\right) C_Y^{n,f}, \tag{2.11}$$

where

$$K_n \equiv C_Y^{n,f} H_n^T \left(H_n C_Y^{n,f} H_n^T + C_D^n\right)^{-1}. \tag{2.12}$$

Eqs. (2.10−(2.12) are known as the KF analysis equations. In them, K_n is known as the Kalman gain matrix. I_{N_y} is the $N_y \times N_y$ identity matrix, where N_y denotes the dimension of the state vector y^n .The superscripts a and f denote analysis and forecast, respectively.

2.3.2 The Ensemble Kalman Filter (EnKF)

Every time new data is available, the mean and covariance in the KF are changed. The state's size may be too big, even for linear situations, for Eq. (2.11) to update the state's covariance matrix computationally. Furthermore, the KF equations are no longer relevant if the problem at hands is nonlinear because the posterior pdf is no longer Gaussian. The Kalman filter was extended to work with non-linear models through the EKF. The EKF uses linearizations of the model and observation equations around the estimated mean of the state. These linearizations, however, may result in unconstrained instabilities in the covariance updates for highly nonlinear systems (EVENSEN, 2009, cap. 4). Additionally, updating the entire states covariance matrix is still required by the EKF, which is computationally unfeasible for high-dimensional problems (AANONSEN et al., 2009).

In this context, Evensen (1994) first introduced the ensemble Kalman filter (EnKF) as an alternative for the EKF in high-dimensional nonlinear dynamical systems. The EnKF is a Monte Carlo approach in which the mean and covariance are successively updated over time and are represented by an ensemble of states. It was further implemented by Houtekamer et al. (1998) and Burgers et al. (1998), where it led to the current implementation of the EnKF by introducing the idea of updating each ensemble member with independently perturbed observations. Therefore, the uncertainty is propagated and represented using these ensembles of states.

There is a tendency to lessen the dependence on Gaussian assumptions because the initial ensemble in the EnKF represents a sampling from the prior distribution, because it is not a straightforward resampling of a Gaussian posterior distribution. Only the linear updates are added to the prior non-Gaussian ensemble. As a result, the updated ensemble will inherit many of the forecast ensemble's non-Gaussian properties. (EVENSEN, 2009, cap. 4). Nonetheless, because the analysis stage still relies on the distributions' first and second order moments, the performance of EnKF may suffer if the distributions deviate too much from Gaussian (AANONSEN et al., 2009). Unfortunately, EnKF's main approximation is also the ensemble representation. Limiting the ensemble's size is required to gain computational efficiency. Small ensembles, however, induce sampling errors that produce spurious correlations. Additionally, the space in which the solutions can be expressed is constrained by the ensemble's size. As a result, after data assimilation, it is noticeable an excessive drop in posterior covariances (EMERICK, 2012).

The history matching problem using the EnKF modifies the classical and traditional problem from parameter estimation to a parameter-state estimation one. In that case, on the EnKF, both the reservoir model parameters to be estimated and primary dynamic variables of the reservoir simulator, such as gridblock pressure, and phase saturation. The problem is being viewed as a parameter-state estimation in order to avoid starting the reservoir simulations from scratch after each data assimilation step (EMERICK, 2012).

The EnKF equations are typically introduced by defining the augmented state vector, y^n, in which the predicted data vector, d^n, is also included, i.e.,

$$y_j^n = \begin{bmatrix} m_j^n \\ p_j^n \\ d_j^n \end{bmatrix}, \tag{2.13}$$

where the subscript j refers to the jth ensemble member. m_j^n and p_j^n are, respectively, the vector of model parameters and states of the dynamical system. For reservoir applications, m includes all model parameters required in the history matching, generally gridblock porosities, permeabilities, end points of relative permeability, etc.; p includes the primary variables of the reservoir simulator, typically reservoir gridblock pressure, fluid saturations, and bubble point pressure in a standard black-oil reservoir (EMERICK, 2012). This “trick” of augmenting the state vector allows the derivation of the EnKF equations following the same ideas to the standard KF equations without removing the effect of the nonlinearity, if, and only if, at every data assimilation time-step, the predicted data vector is a linear function of the combined (un-augmented) state vector (LI; REYNOLDS, 2009).

The EnKF analysis equation can be written as

$$y_j^{n,a} = y_j^{n,f} + \tilde{C}_{YD}^{n,f}\left(\tilde{C}_{DD}^{n,f} + C_D^n\right)^{-1}\left(d_{uc,j}^n - d_j^{n,f}\right),, \qquad for\ j = 1, 2, \cdots, N_e, \tag{2.14}$$

in which N_e is the number of state vectors in the ensemble, in other words, the ensemble size; $d_{uc,j}^n$ is a sample from the Gaussian distribution $\mathcal{N}(d_{obs}^n, C_D^n)$, normally called as a perturbed (or randomized) observed data vector. The tilde (~) introduced in the matrices notations is to emphasize that these matrices are estimated from an ensemble. Thus, $\tilde{C}_{YD}^{n,f}$ and $\tilde{C}_{DD}^{n,f}$ are the cross-covariance matrix between the augmented state vector parameters and observation data and auto-covariance matrix of predicted data, respectively, calculated by the following approximations:

$$\tilde{C}_{YD}^{n,f} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left(y_j^{n,f} - \bar{y}_j^{n,f}\right)\left(d_j^{n,f} - \bar{d}_j^{n,f}\right)^{\mathrm{T}}, \tag{2.15}$$

and

$$\tilde{C}_{DD}^{n,f} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left(d_j^{n,f} - \bar{d}_j^{n,f}\right)\left(d_j^{n,f} - \bar{d}_j^{n,f}\right)^{\mathrm{T}}, \tag{2.16}$$

where

$$\bar{y}_j^{n,f} = \frac{1}{N_e}\sum_{j=1}^{N_e} y_j^{n,f}, \tag{2.17}$$

and

$$\bar{d}_j^{n,f} = \frac{1}{N_e}\sum_{j=1}^{N_e} d_j^{n,f}. \tag{2.18}$$

The error in the estimate of the covariance decreases in a proportional rate of $1/\sqrt{N_e}$, meaning that a large ensemble may be necessary for an accurate estimate, and, if the ensemble size chosen is small, the error in the estimate may be large (AANONSEN et al., 2009). Also, in order to keep a reasonable computational cost for each data assimilation step, small ensembles are necessary, which introduce more sampling errors and limit the degrees of freedom to assimilate data and, thus, the ensemble variance obtained after each data assimilation can be underestimated, being this an important limitation of the EnKF (AANONSEN et al., 2009).

To compensate for the underestimation of posterior variances in the EnKF, there have been literature about covariance inflation, method in which inflation factor are used to increase the covariance in the forecast ensemble without changing the mean. Some applications can be seen in Anderson (2016), Liang et al. (2012), Emerick (2019) and Silva et al. (2021), the last two being more focused on the application of covariance inflation in petroleum history matching.

2.3.3 Ensemble Smoothers

Unlike EnKF, in which data is sequentially assimilated in time, the ensemble smoother (ES) was proposed by van Leeuwen and Evensen (1996) and is a method that computes a global update by simultaneously assimilating all data available and realizing only one global change at the end of the operations. van Leeuwen and Evensen (1996) evaluated ES and EnKF with Lorenz equations and concluded that EnKF performed better than ES because the EnKF's recursive updates keep the ensemble of states on track and closer to the actual solution.

Recently, Skjervheim et al. (2011) evaluated ES with EnKF and found that both approaches produced similar outcomes for the reservoir history-matching issues they were focusing on. The fundamental benefit of ES is that it prevents the reservoir simulator from having to restart, as is required by the EnKF sequential data assimilation method. When used to solve reservoir history-matching problems, this makes ES significantly faster and simpler to use than EnKF. ES is a desirable alternative for data assimilation workflows that combine diverse aspects of the reservoir modeling process, such as seismic, structural, and geological modeling with flow simulation. The reduction of simulation restarts is another benefit of ES; see, e.g., Liu and Grana (2020) and Zachariassen et al. (2011).

When implementing ES, we only need to consider the parameter-estimation issue if we overlook model uncertainty, which is a typical assumption in reservoir history-matching situations. In this instance, using ES eliminates the parameter-state consistency problem that was seen while using EnKF to assimilate sequential data (EVENSEN, 2009, cap. 6). ES formulation is similar to EnKF, writing the analyzed vector of model parameters, m_j^a, as:

$$m_j^a = m_j^f + \tilde{C}_{MD}^f \left(\tilde{C}_{DD}^f + C_D\right)^{-1} \left(d_{uc,j} - d_j^f\right). \tag{2.19}$$

Even though the great reduction of computational cost of the ES is its main benefits, the single global data assimilation done by only one Gauss-Newton correction may not be able to provide feasible data matches results to reservoir history-matching problems. In such case, using the EnKF’s sequential application of corrections between data assimilation or, in other words, the accumulation of several Gauss-Newton corrections, would give a better result for the reservoir history-matching problem, as production data is conditioned to each and several assimilation time-steps and the ensemble is conditioned to the production history (EMERICK, 2012).

2.3.4 Ensemble Smoother with Multiple Data Assimilation (ES-MDA)

To overcome the possible mismatch between observed and predicted data and the only global update of the parameters using ES, iterative ensemble smoothers (iES) have been developed. The iES can be used for reservoir models with a moderately large number of wells, a variety of data types and a relatively long production history. Chen and Oliver (2014) used this method in a North Sea field, the Norne field, and with the Levenberg-Marquardt iterative ensemble smoother were able to achieve improved data after three iterations of a model with an approximate total number of parameters of 150.000. Ma and Bi (2019) derived a novel and adaptive iES method, based on the

Levenberg-Marquardt algorithm for nonlinear least squares optimization, and demonstrated that it is possible to use the ensemble smoother as an approximate linear least squares solver and thus avoid expensive adjoint calculations. This transforms the history-matching problem in a nonlinear least squares problem and helps to mitigate the consequences of three of the main assumptions required by the EnKF and ES (MA; BI, 2019): a strictly linear relationship between the model and the parameters, Gaussian prior distribution for the model parameters and Gaussian measurement noise.

To further improve the results and to optimize the data assimilation process in the highly nonlinear problem that is the reservoir modeling and simulation for petroleum engineering, Emerick and Reynolds (2012) first proved that there is an equivalence between single and multiple data assimilation (MDA) for the linear-Gaussian case when the same data is assimilated multiple times with the covariance matrix of the measurement errors multiplied by the number of data assimilations, presenting evidence that multiple data assimilations can improve EnKF estimates for the nonlinear case. Linking with the need to develop efficient iES and already with the knowledge of this method's benefits, Emerick and Reynolds (2013) proposed the Ensemble Smoother with Multiple Data Assimilation method (ES-MDA), which is the use of MDA in conjunction with ES to assimilate the data. ES-MDA can be interpreted as an iterative form of ES, where the number of iterations (e.g., assimilations made) must be selected a priori.

In order to use ES-MDA, it is necessary that the multiplication factors used to inflate the covariance matrix of the measurement errors for a correct *a posteriori* parameter's distribution, satisfy the following expression (EMERICK; REYNOLDS, 2013):

$$\sum_{\ell=1}^{N_a} \frac{1}{\alpha_\ell} = 1, \tag{2.20}$$

where N_a is the number of data assimilation defined and α_ℓ is the inflation coefficient. Therefore, for every ES-MDA problem, it is needed to choose the number of assimilations that will be done and the inflations coefficients before each assimilation step. The parameters updates steps follow the EnKF equations, taking into consideration the inflation coefficients. The full ES-MDA algorithm construction can be visually interpreted in Figure 1, with more details in Section *3.1*.

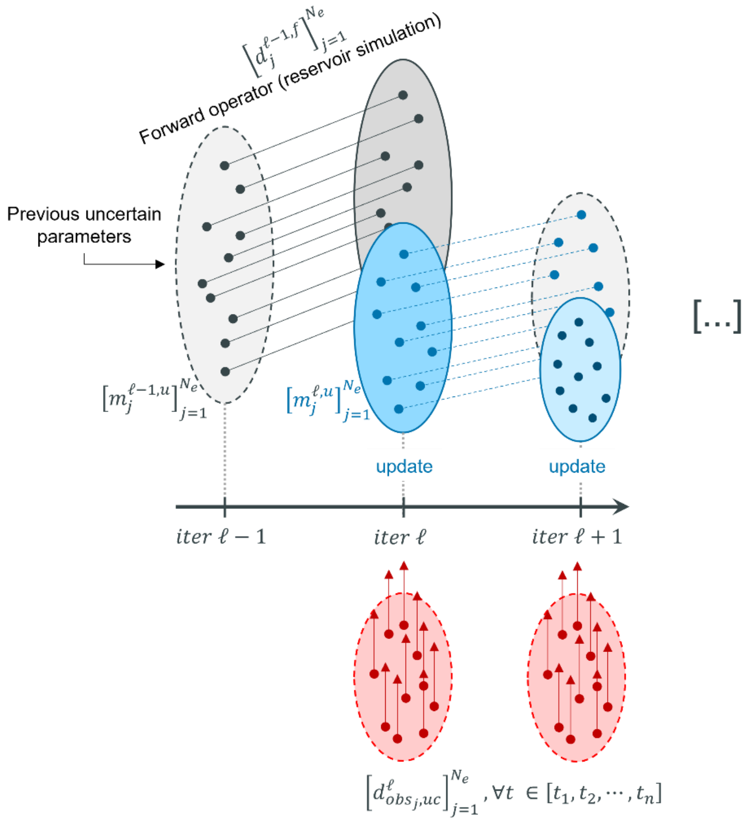


Figure 1 - ES-MDA process

From Eq. (2.20), it is noticeable that a simple choice for α is $\alpha_\ell = N_a$ for all ℓ. Nevertheless, it makes intuitive sense that selecting alpha in descending order can enhance the method's effectiveness (EMERICK; REYNOLDS, 2013). therefore, on initial steps data is assimilated with a high α value, which corresponds to reducing the amplitude of the initial updates, and then we gradually decrease α.

The need to define the number of iterations and the inflation factors before the assimilation process is one of the main drawbacks of the ES-MDA, in fact, being necessary to restart the entire process if the results quality is not desirable after the end of the algorithm (RANAZZI; SAMPAIO, 2019b). Nonetheless, Emerick (2012) shows us that changing the inflation coefficients in a decreasing order resulted in only small improvements compared to using them constant and equal to the number of data assimilations.

Maucec, et al. (2016) used an ensemble-based computer Assisted History Matching (AHM) workflow integrating probabilistic Bayesian interference using ES-MDA of a real-life carbonate oil field, with a sector model of, roughly, 17 million active grid cells with no application with simulation grid upscaling. The authors managed to obtain matches on field-level reservoir pressures spanning a long production history, also reporting a detailed uncertainty matrix with different ensembles of sensitivity scenarios.

Emerick (2016) provided an assessment of the effectiveness of the ES-MDA used in the same field scenario as Emerick and Reynolds (2013) to absorb production and seismic data. According to this author, localization techniques are crucial for dealing with spurious correlations and variability loss. Additionally, an adaptive approach was created, where the iterations and inflation coefficients are chosen in accordance with how the data mismatch is progressing. Localization and inflation techniques for ensemble methods can be found in the literature (e.g., Bjarkason et al.(2021), Soares et al. (2018), Ranazzi et al. (2022; 2019a) and Emerick (2019)).

Rosa et al. (2022) also employed the ES-MDA technique to assimilate 4D seismic data and production forecast in a deep-water oil field, located in offshore Southwest Brazil, in order to compare the impact of using grids with different resolutions, exploring options to incorporate 4D seismic maps in data assimilation and production forecast. The authors proposed the use of two types of grids: a coarser grid and a refined grid. Results shows that both grids can be used depending on the project's objectives. Since it can accurately reflect forecasting with considerably less expensive models, the coarser grid is a useful choice for short-term investigations. Due to its safer risk analyses and improved representation of model uncertainty, the refined grid may be appropriate for long-term decisions.

The development of the ensemble-based history matching (EnHM) software in the TotalEnergies' geoscience platform known as Sismage-CIG is described in Abadpour et al. (2018). The ES-MDA approach forms the basis of the software. A mature large carbonate oil field, a deep-offshore turbidite field, and a carbonate gas field were the three complex genuine fields that were used to demonstrate the performance of the methods.

3 METHODOLOGY

The methodology presented in this work aims to apply ES-MDA (EMERICK; REYNOLDS, 2013) to a typical history matching problem of a reservoir model using real data.

3.1 History matching using ES-MDA

Following the definition of the uncertain parameters (Section *4.3*) and their input on the model of the Alpha-system, it is necessary to sample them given the type of pdf predefined as one of the inputs. Every set of the different combinations of parameters are called ensembles, and they are essential on the history matching process with ensembled-method data assimilation. Each member j of the ensemble with size N_e which contains the uncertain parameters of the model are defined as:

$$m_j^u = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{N_m} \end{bmatrix}, \tag{3.1}$$

where m^u is the vector containing all N_m uncertain parameters that will be updated, such as permeability and productivity multipliers indexes data. The superscript u stands for *uncertain*.

After the sampling of the parameters for the N_e ensembles, it is also generated N_e models which are taken to the forward operator, from time zero to the end of the historical period, in order to compute the vector of predicted data:

$$d_j^{\ell,f} = F\left(m_j^{\ell-1,u}\right), \qquad for\ j = 1, 2, \cdots, N_e, \tag{3.2}$$

where $F(\cdot)$ is the forward operator (i.e., the non-linear model used to obtain the forecast observations); $d_j^{\ell,f}$ is the *forecast* (f) model response with size N_d, the total number of measurements in the historical period, resulted from the combination $m_j^{\ell-1,u}$ of uncertain parameters.

For this work, the INTERSECT™ Reservoir Simulator was employed to compute the non-linear forward problem, hence simulating the dynamic behavior of the reservoir response to the different combination of uncertainty parameters previously defined. INTERSECT™ is a high-resolution reservoir simulator that, when coupled with a high-performance computer, enables this work to fully parallelize the execution of dynamic simulations for every ensemble member.

ES-MDA requires the N_d measurements to be perturbed, i.e., apply Gaussian noise in order to be treated as random variables. The mean and standard deviation values are calculated from previous research of the study area during screening methods and single parameter research:

$$d^{\ell}_{obs_j,uc} = d_{obs} + \sqrt{\alpha_\ell}C_D^{1/2} z_d, \qquad for\ j = 1, 2, \cdots, N_e, \tag{3.3}$$

where $d^{\ell}_{obs_j,uc}$ is the perturbed vector of observations; z_d is sampled with $z_d \sim \mathcal{N}(0, I_{N_d})$; C_D is the $N_d \times N_d$ covariance matrix of observed data measurement errors:

$$C_D = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{N_d}^2 \end{bmatrix}. \tag{3.4}$$

After that, the simulations results are compared with the observation data. As every process is based on statistical analysis and from observed data, there will certainly be uncertainty on the model, giving mismatches from the simulation results and the observation data. Hence, these discrepancies will be numerically used inside the data assimilation based on correlation computation, and, for each iteration ℓ, the previous sampled parameters will be updated to a new set of uncertain parameters, in order to reduce the variance of the discrepancy, following:

$$m_j^{\ell,u} = m_j^{\ell-1,u} + \tilde{C}_{MD}^{\ell-1,f}\left(\tilde{C}_{DD}^{\ell-1,f} + \alpha_\ell C_D\right)^{-1}\left(d^{\ell}_{obs_j,uc} - d_j^{\ell,f}\right), \quad for\ j = 1, 2, \cdots, N_e, \tag{3.5}$$

where, $\tilde{C}_{MD}^{\ell-1,f}$ represents the cross-covariance matrix between the prior vector of model parameters, $m_j^{\ell-1,u}$, and the vector of predicted data, $d_j^{\ell,f}$; $\tilde{C}_{DD}^{\ell-1,f}$ is the $N_d \times N_d$

auto-covariance matrix of predicted data. As before, the tilde (~) notation indicates that these covariances matrices are estimated around the ensemble mean:

$$\tilde{C}_{MD}^{\ell-1,f} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left(m_j^{\ell-1,u} - \overline{m}^{\ell-1}\right)\left(d_j^{\ell,f} - \bar{d}^{\ell,f}\right)^T, \tag{3.6}$$

$$\tilde{C}_{DD}^{\ell-1,f} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left(d_j^{\ell,f} - \bar{d}^{\ell,f}\right)\left(d_j^{\ell,f} - \bar{d}^{\ell,f}\right)^T, \tag{3.7}$$

whereby, $\overline{m}^{\ell-1}$ and $\bar{d}^{\ell-1}$ represent the average vector of the prior model parameters and the average vector of predicted data, respectively:

$$\overline{m}^{\ell-1} = \frac{1}{N_e}\sum_{j=1}^{N_e} m_j^{\ell-1}; \qquad \bar{d}^{\ell-1} = \frac{1}{N_e}\sum_{j=1}^{N_e} d_j^{\ell-1}. \tag{3.8}$$

As mentioned, the ES-MDA processes will go on until the pre-defined number of assimilations (N_a) is reached. It is important to mention that the assimilation process is made after the end of the forward operator and that the observation data are randomly sampled according to the data error covariance. Furthermore, for the update part, ES-MDA uses an inflation factor α_ℓ for each assimilation step, and their choice must respect the condition stated in Eq. (2.20) for a correct *a posteriori* parameter's distribution.

In this work, for all study cases using ES-MDA process, the inflation factor used were constant and equals to the number of iterations, i.e., $\alpha_\ell = N_a$. Figure 2 illustrates the workflow containing the steps of applying ES-MDA.

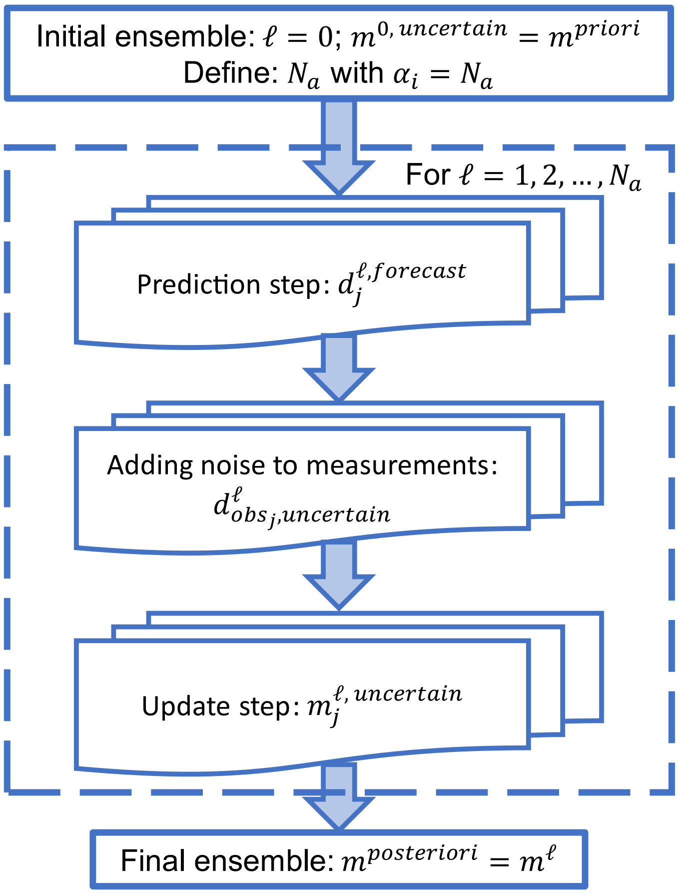


Figure 2 – ES-MDA workflow

Source: Adapted from Ranazzi & Sampaio (2018)

3.2 Objective function

The Objective Function (OF) for each data assimilation at iteration ℓ used in this work is through the quadratic deviation between simulated model forecast and the observation data, normalized by the inverse of the covariance matrix of observed data measurement errors, C_D and with the total number of observations N_d, as shown in Eq. (3.9).

$$OF^{\ell} = \sqrt{\frac{\sum_{j=1}^{N_e}\left(d_j^{\ell,f} - d_{obs_j}^{\ell}\right) C_D^{-2} \left(d_j^{\ell,f} - d_{obs_j}^{\ell}\right)^T}{N_D}}. \tag{3.9}$$

4 STUDY CASE

The study area of the work focuses on the analysis of the Central and Eastern panel of the Alpha-Field. Thus, this section is divided between the initial model used for the reservoir system, containing explanation about the link between the geology and how to transfer the heterogeneities of the model to the fluid flow simulator in terms of numeric parameters.

4.1 2G&R synthesis of the case study

The field under study is located in an offshore basin, with the reservoir about 1000 meters from the seafloor surface. The reservoir is of very good quality and consists of turbiditic sediments composed of a loosely consolidated sand with a very good permeability of up to 10 darcys (D). The oil is also of superior quality.

The depositional environment is turbiditic, corresponding to the upper part of a turbiditic lobe. The channels are erosive-constructive. They form amalgamated channel complexes alternating between channel and levee facies. A fluid barrier is formed when the system comes into contact with diapirs to the east. As a result, the field is a reservoir bounded by the complex's erosive margins, the clay roof, and the diapirs.

The studied field is called Alpha Field, a multi-story erosive constructive channel complex. The reservoir is placed at the level of a turtle-back salt anticlinal created after turbidite reservoir deposition. This structure causes a large number of Keystone faults, which are normal faults that run perpendicular to the system's elongation. The studied area is divided according to the direction of the channels from East to West into 3 zones − Alpha-East, Alpha-Central and Alpha-West (Figure 3). Within the system, structural and stratigraphic heterogeneities will exist. There is little transmissibility between regions because the degree of depletion varies by region. The synthesis will also give us an indication of the uncertain parameters determined later.

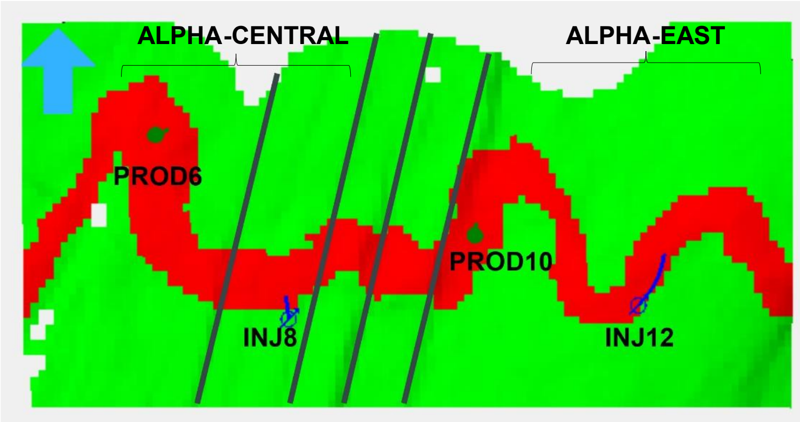


Figure 3 - Top view of the erosive channel complex containing an explanatory diagram of the studied field with the highlighted wells for future analysis

4.2 Initial model

The optimization of parameters is a critical step in the reservoirs and management's future. During the history matching step, dynamic parameters are updated. The fluid flow simulator used in this work − INTERSECT™ − allows us to change them based on the regions using three software's keywords that correspond to three degrees of heterogeneity (SCHLUMBERGER, 2014). These functions are:

- **MULTNUM:** defines regions for applying inter-region transmissibility multipliers, instead of applying them to flux regions defined by FLUXNUM;
- **FLUXNUM:** identifies extent of each flux region, using with the flux boundary option to define regions that can be run as separate models with boundary fluxes defined in a previous full field run;
- **OPERNUM:** defines regions for performing arithmetic operations on property array.

The flow units in our Alpha-system that are cut by the faults within the center block are represented by MULTNUM. The faults run perpendicular to the system's orientation. In our model, FLUXNUM represents the three vertical levels (Alpha-Lower, Alpha-Middle, Alpha-Upper). Clay barriers can be found throughout the complex and are designated as maximum flooding surfaces (MFS), which translate geologically as clays in the vast majority of situations. The last function to represent our turbiditic system geology's is OPERNUM , which represents the elementary channels. In the model, the sedimentary features used were sand channels, proximal levee and distal levee.

According to regional seismics, fluid circulation will be mostly within these formations due to their high permeability.

For the modelling of the elementary channels, ULIKE™ was used. This process is a rule-based modelling tool that simulates the sediment transports by water flow following the random path that a particle would take along any designated fairway, which occurs in most of clastic systems. To build sedimentary systems with a specific domain flow, this approach employs the Random Walk principle, which is a form of stochastic process in which each move step is preset by a particular probability. Then, according to the user's defined geometrical parameters and limitations, the random walk of a particle path is dressed in sedimentary features such as channels, levees, lobes, and so on. To keep the grid consistent, the process is applied to each layer defined on the model. ULIKE™ has the advantage of respecting seismic and well constraints at the level of turbiditic reservoirs (MASSONNAT, 2019). This method is comparable to object-based modeling; however, it yields better results than purely Boolean object-based modeling.

Table 1 summarizes the values given to the regions according to their families. These values are used to identify the different areas of the reservoir, and to easily manage transmissibility between these regions, as multipliers will be applied during the fluid flow simulations and will also be considered as uncertain parameters during the data assimilation process. A reduction in transmissibility may be the effect of faults or system edges for MULTNUMS (Figure 4), MFS or clay barriers for FLUXNUM (Figure 5) and channel skins for OPERNUMS (Figure 6). In the figures below, the reservoir complex is shown in a 3-dimensional space, with each heterogeneity displayed.

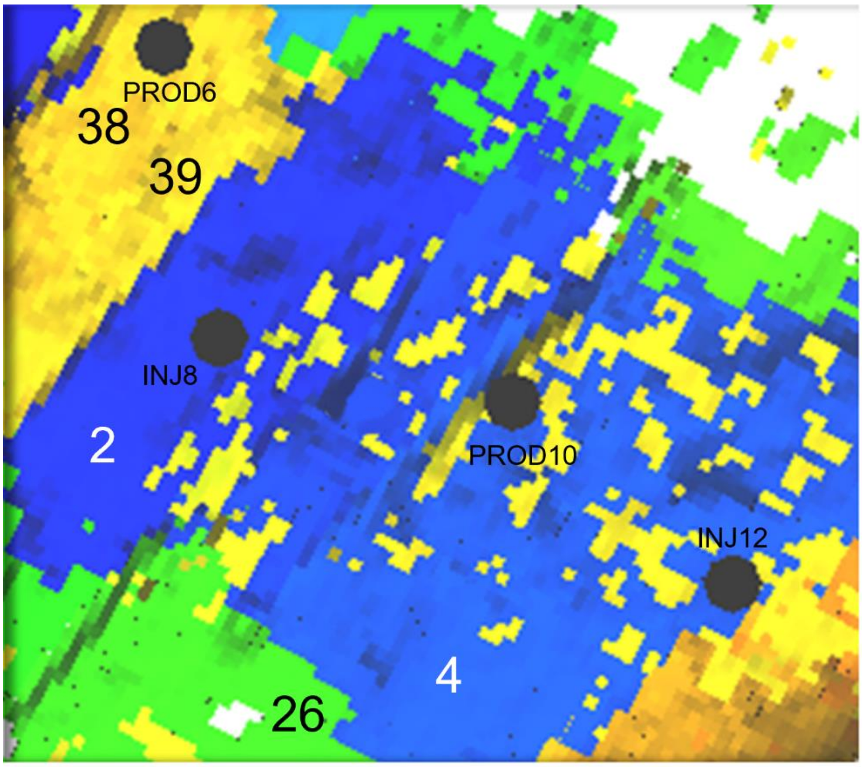


Figure 4 - MULTNUM regions and their specific numeric inside the reservoir model

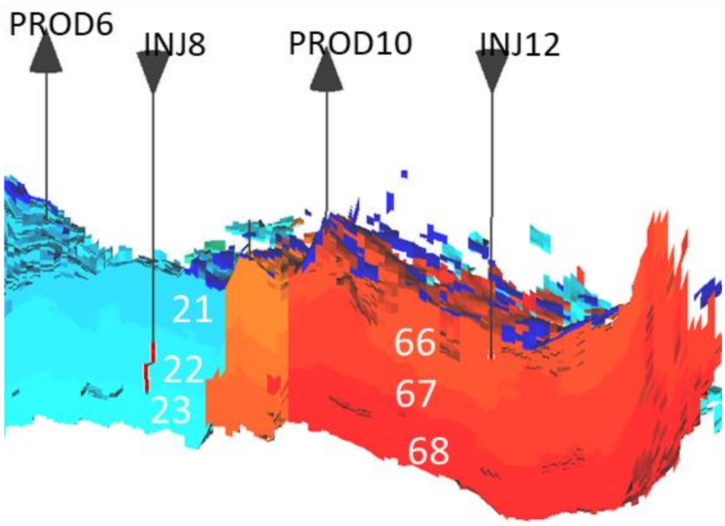


Figure 5 - FLUXNUM regions and their specific number inside the reservoir model

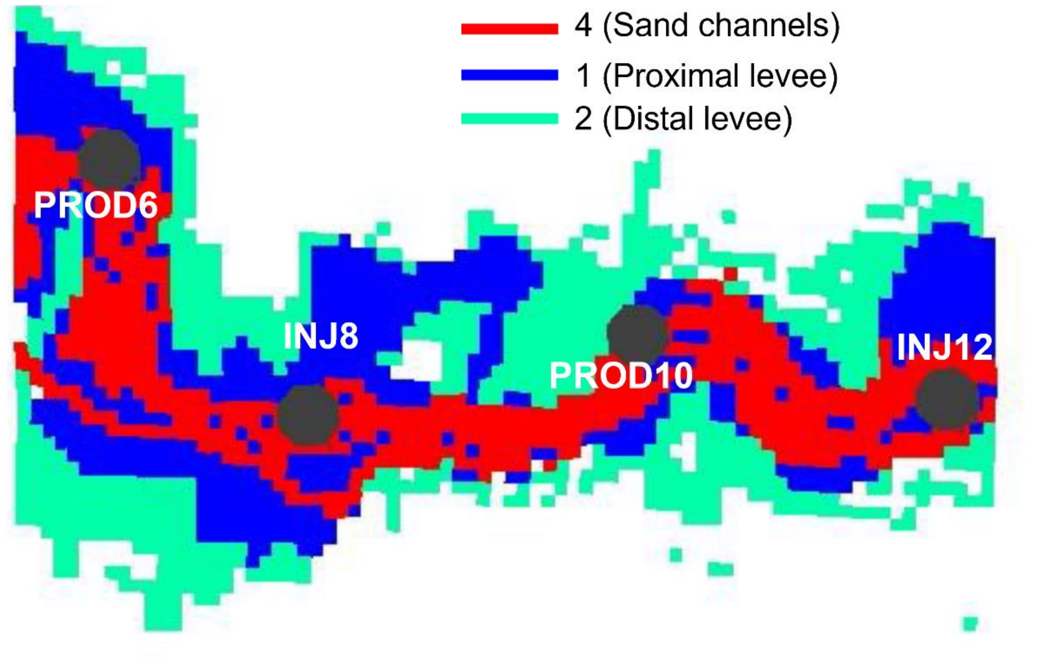


Figure 6 - OPERNUM regions and their specific number and sedimentary representation inside the reservoir model

Table 1 - Figure association with regions by location

	MULTNUM	FLUXNUM	OPERNUM
Regions	Flow units and fault blocks	Alpha-lower, Alpha-middle, Alpha-upper	Elementary channels
Alpha-Central	2, 38 and 39	21, 22 and 23	1, 2 and 4
Alpha-Eastern	4 and 26	66, 67 and 68	

4.3 Influencing parameters on the history matching process of the Alpha field

After the definition of the heterogeneity zones mentioned in Section *4.2*, the uncertain parameters, such as transmissibility multipliers, elementary channels' characteristics (i.e., porosity and permeability) and wells' productivity indexes multipliers need to be chosen in order to assess their impact on the fluid flow. The choice of the parameters is directly related to the studied region and will be further detailed in Sections *5.1.1, 5.1.2* and *5.2* , with their respective definitions in Tables 2, 3 and 4.

This is considered to be one of the most challenging processes on reservoir simulation, as the number of parameters that can be considered and have directly and indirectly response on the flow simulator results is very large, sometimes can pass the 100,000 parameters depending on the cell grid size and the heterogeneity of the reservoir. For the Alpha-system, previous studies have already been carried out on the last couple of years in order to have a better geological understanding of the area and how to improve more and more this in a numerical representation. This work uses mostly multipliers uncertain parameters on flux/fault regions (MULTNUM) and on the vertical layers (FLUXNUM), as well as the elementary channel properties generated by ULIKE™ – permeabilities, porosity, net-to-gross ratio, and transmissibility multiplier applied in the boundaries of sand channel and proximal levee (OPERNUM)

The importance of turbiditic elementary channels in matching water arrivals cannot be overstated. Indeed, the fluids will primarily rush into these places with extremely high permeability; but digitation will occur, leaving other areas un-swept. The ‘skin’ of the channel refers to how the channels communicate with the background.

The seismic structure of the specific field reveals the presence of faults, which might be fluid conductive or fluid impermeable. The first key element in this field is the transmissibility between faults. It has a significant impact on the field’s exploitation. The transmissibility of faults has a big impact on pressure studies (BHP: Bottom Hole Pressure), but it also has a big impact on production mechanisms, especially if the fault restricts access to an aquifer or a connection with an injector.

The analysis of the BHP is one of the most important of the reservoir. It shows two types of information: the flowing pressure and the average reservoir pressure. The first is the pressure when the well is active and producing. The pressure’s profile of a reservoir is a declining curve with a minimum pressure value in the well. The flowing pressure depends on the conditions of the well’s condition (i.e., the skin and the permeability on the well’s vicinity). In the model, the skin of the well and its vicinity’s condition are approximated with a productivity index (PI), as this property is linked with the pressure variation during the flowing part of the production. Therefore, this Productivity Index can be adjusted using a PI multiplier parameter.

Besides, the average reservoir pressure is the one obtained away from the well. When production ceases, the pressure at the well will revert to the reservoir’s average pressure after a sufficient period of time; this is known as a static pressure, which can be calculated by extrapolating the build-up of pressure to a long period of time to achieve the constant pressure. This average reservoir pressure is linked to the connected investigated volume; the lower the volume connected, the faster the pressure will drop. Nonetheless, aquifer and water injection lead to a pressure contribution. Therefore, the aquifer volume can also be considered as an uncertain parameter on the model and the connectivity to the injector.

To keep the parameters consistent, the variation of the parameters is based on the 2G&R synthesis, general reservoir geology understanding, and common sense. The single-parameter research previous made by studies on the Alpha-system in examining the reservoir's behavior also aid on that perspective.

5 RESULTS

The following sections contains the analysis of the two sections chosen of the Alpha field to be matched: Alpha-East and Alpha-Central.

5.1 Alpha-East: PROD10 – INJ12

The study of the Alpha-East compartment of the reservoir is divided into two ES-MDA analysis:

1. The first ES-MDA: data assimilation from a set of runs in the reservoir fluid flow simulator with fewer uncertain parameters in order to check the behavior and stability of the method and if there are any corrections, adjustments and conclusion that could be proposed to obtain a better match of the production vector observation data;
2. The second ES-MDA: data assimilation from the same reservoir compartment with more uncertain parameters defined mainly to obtain a better match on the flowing part of the fluid flow process and the reservoir pressure after the long shut-in of the producer well.

5.1.1 First ES-MDA run on Alpha-East

ES-MDA is used to update the uncertain parameters in order to find a model that behaved similarly to the real reservoir. It is possible to present a few indicators that led to the correct behavior of the interest response. There are three vertical FLUXNUM compartments in the East Block, which is itself a MULTNUM region, which correspond to three erosive channel steps. There are two wells within this compartment which will be analyzed, including one producer (PROD10) and one water injector (INJ12) (Figure 3). The pressure study suggests that this block has strong communication. Water injection is the production mechanism, but it is insufficient to maintain reservoir pressure. Additional pressure support is provided by an aquifer.

The uncertain parameters definition for this panel and their respective definition are shown in Table 2. For this analysis, the number of ensembles used was $N_e = 25$ and with 3 iterations ($N_a = 3$) and the observations to match were: PROD10 water cut and bottom-hole pressure; INJ12 bottom-hole pressure.

Table 2 – Parameter's designation and their explanations on the analysis of Alpha-East

Parameters	Explanations
m_4_26	Transmissibility multiplier between distal levee and channel fairways
multregt_transm	Transmissibility multiplier between channel and proximal levee
op4_perm_i	Elementary channel permeability in i-direction (assumed isotropy and permeability in k-direction is 10% of i-direction)
op4_poro	Elementary channel porosity
op4_ntg_ratio	Elementary channel net-to-gross ratio

One of the simplest ways to verify the efficiency of an ES-MDA process is to analyze the simulator response before and after the adjustments. In this context, the producer PROD10 and injector INJ12 are the selected wells in order to assess the efficiency of the method over each iteration. On Figures 7, 8, 9 and 10, the red lines represent the initial ensemble spread, the green lines represent the last iterations of the assimilation process, and the black squares are the available observation data. It is possible to observe that the response of the model after the adjustments are able to better represent the observation data available compared with the initial ensembles.

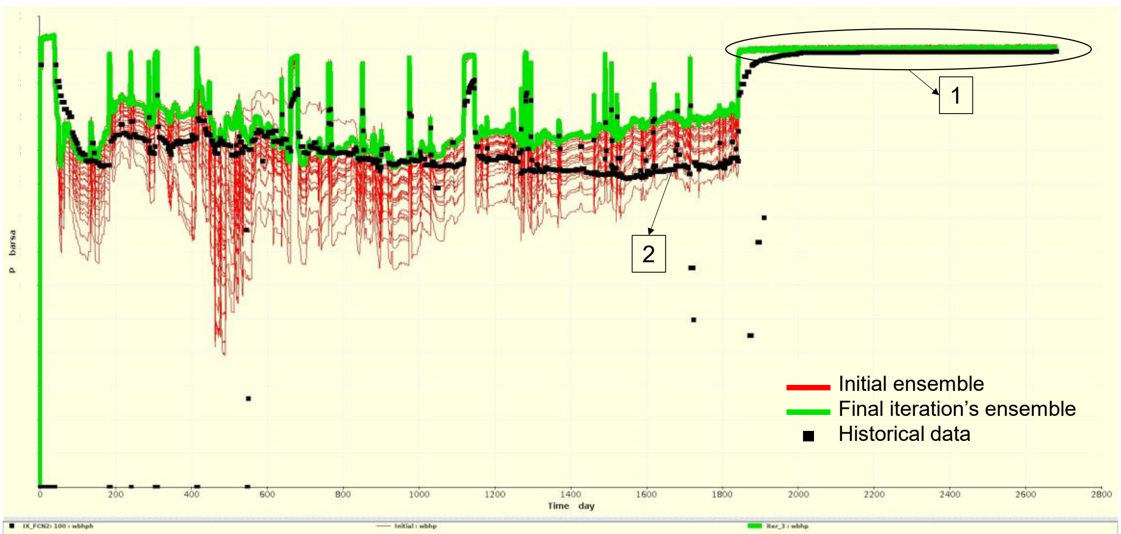


Figure 7 – Well bottom hole pressure for the producer PROD10. Box 1 shows the shut-in pressure and Box 2, the flowing pressure. The red lines indicate the initial ensemble. The green lines indicate the final ensemble. The black squares are the historical data.

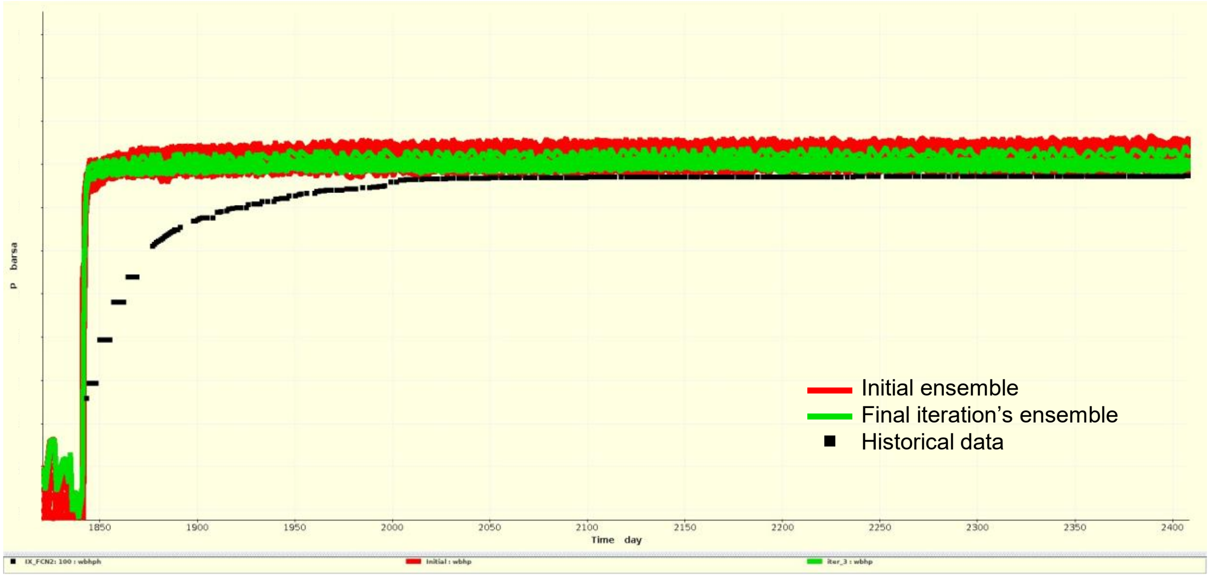


Figure 8 – Local analysis at shut-in pressure of producer PROD10. The red lines indicate the initial ensemble. The green lines indicate the final ensemble. The black squares are the historical data.

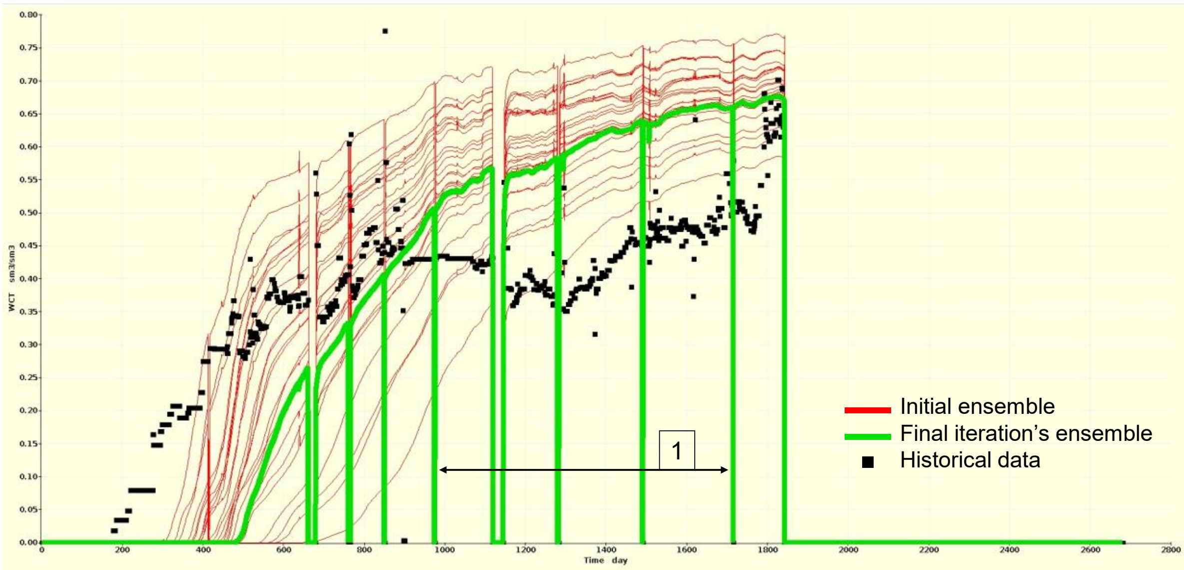


Figure 9 - Water cut for the producer PROD10. The red lines indicate the initial ensemble. Box 1 shows the interval with higher simulated water cut values. The green lines indicate the final ensemble. The black squares are the historical data.

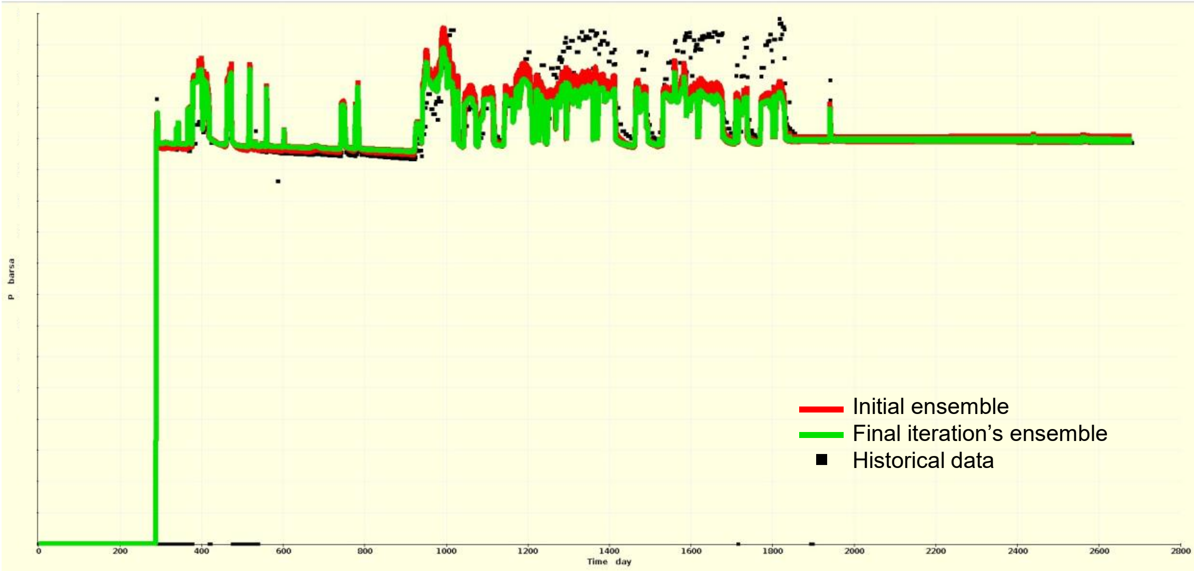


Figure 10 - Well bottom hole pressure for the injector INJ12. The red lines indicate the initial ensemble. The green lines indicate the final ensemble. The black squares are the historical data.

These figures show the response adjustments of the models that minimize the variance from of the ensembles from each iteration. Furthermore, in Figure 8, the simulated response matches well the static pressure behavior, staying few barsas above the observation data.

On the static process of the simulation, the well is shut and the pressure slowly builds-up in order to maintain the equilibrium of the reservoir. The ΔP between the initial pressure (i.e., before production) and the lower long term shut-in pressure is related to the change of the component (e.g., water, gas, oil) volume (ΔV) in the reservoir from the cumulative produced volume of components and the cumulative injected volume of water. Secondly, the initial component connected volume in place impacts also this pressure difference. Finally, this ΔP is also dependent of the total compressibility of the system (i.e., resulting from the compressibility of the components − weighted by the fluid saturation − and the one of the rock). Then, bigger is the initial connected volume, less is the ΔP. Since the aquifer volume is often uncertain, it is considered as a key parameter to match this ΔP difference. The inclusion of the aquifer volume and the analysis of ES-MDA results will be made on Section *5.1.2*.

Still on Figure 7, Box 2, gives us the flowing pressure. This simulation steps gives us more local comprehension of the system compared with the static pressure. In the vicinity of the well PROD10, the effect of the skin and permeability in the field are affecting the flowing pressure. That said, from ES-MDA's final iteration curves, it is necessary to adjust the well productivity index multipliers in order to obtain better matches on the flowing part. The productivity index of a well is defined as being the ratio of the flow rate by the pressure differential and is also very affected by the skin. In other sections (*5.1.2* and *5.2*), the PIs multipliers will be considered in some of the simulations and their strong non-linear response will also be discussed.

From PROD10 water cut profile, Figure 9 shows that the initial ensemble manages to represent the early water breakthrough given the transmissibility parameters of the elementary channels predefined in the uncertain parameter list. However, the last data assimilation results give very tight ensembles, with a relatively late water breakthrough and an excess of water production which is represented in the interval in Box 1 of Figure 9. The final average value of the water cut was matched – roughly 0.6.

In addition to the volume of the eastern aquifer, this mismatch of the water cut from the previous mentioned period of days can also be linked with the injector productivity index values (Figure 10). From INJ12 bottom hole pressure graphic, we can observe good match on the static parts but an underestimation on the flowing pressure.

Figures 11 and 12 show the evolution of the objective function for each property of the production vector we want match. As expected from ES-MDA, the initial twenty-five simulations have very sparse values OF values, as the uncertain parameters are fist randomly sampled given the pdf and the ranges. With the first data assimilation iteration, the OF is reduced for all production vector properties, minimizing the best way possible given the observation data. Figure 12 shows an unstable evolution of OF over the iterations, but on the last ones it is possible to see a certain stabilization on the interval ES-MDA found for all the production vector properties. Finally, in Figure 13 is the evolution of the average value of the OF, showing the same behavior as described earlier.

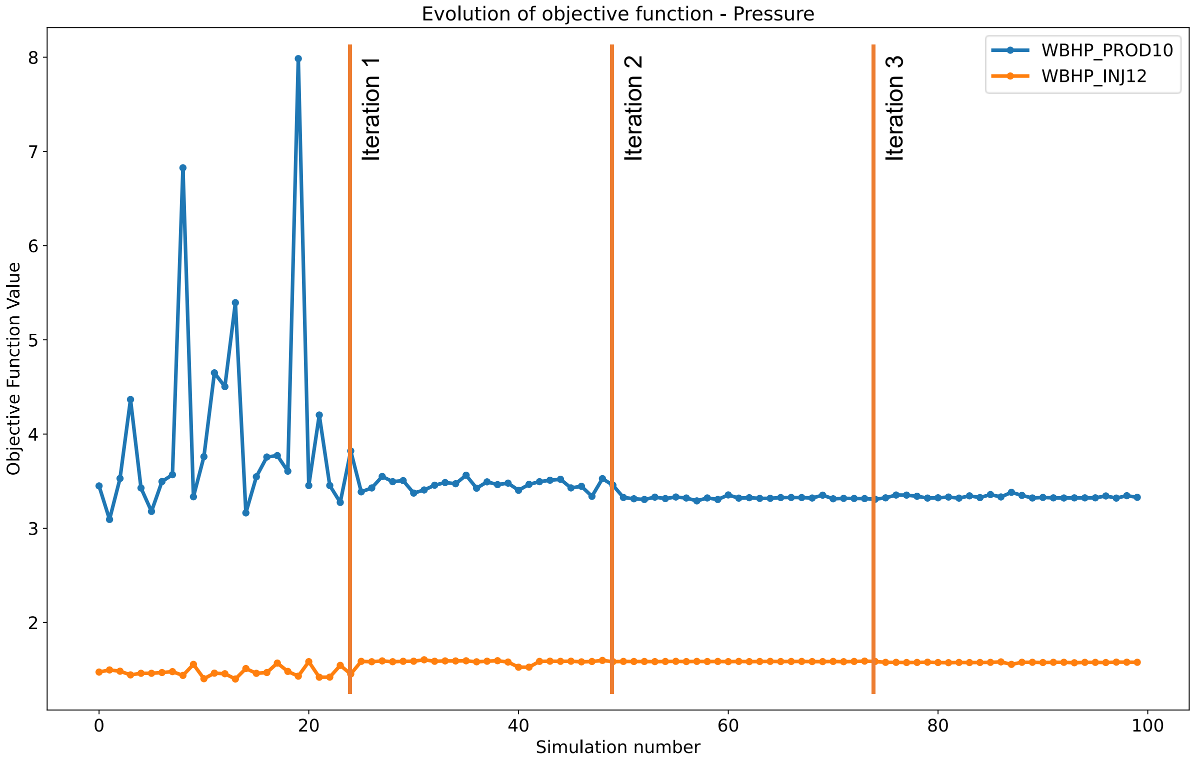


Figure 11 - Evolution of the pressure's portion of the ES-MDA OF

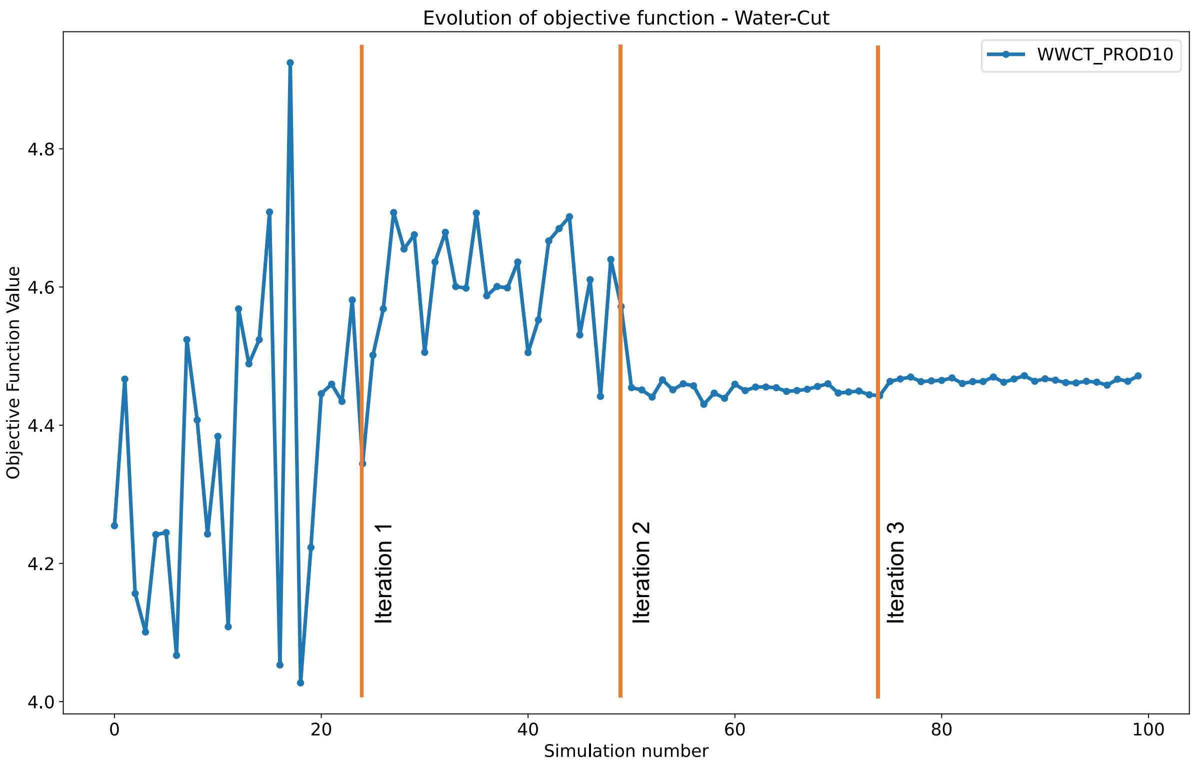


Figure 12 - Evolution of the water cut's portion of the ES-MDA OF

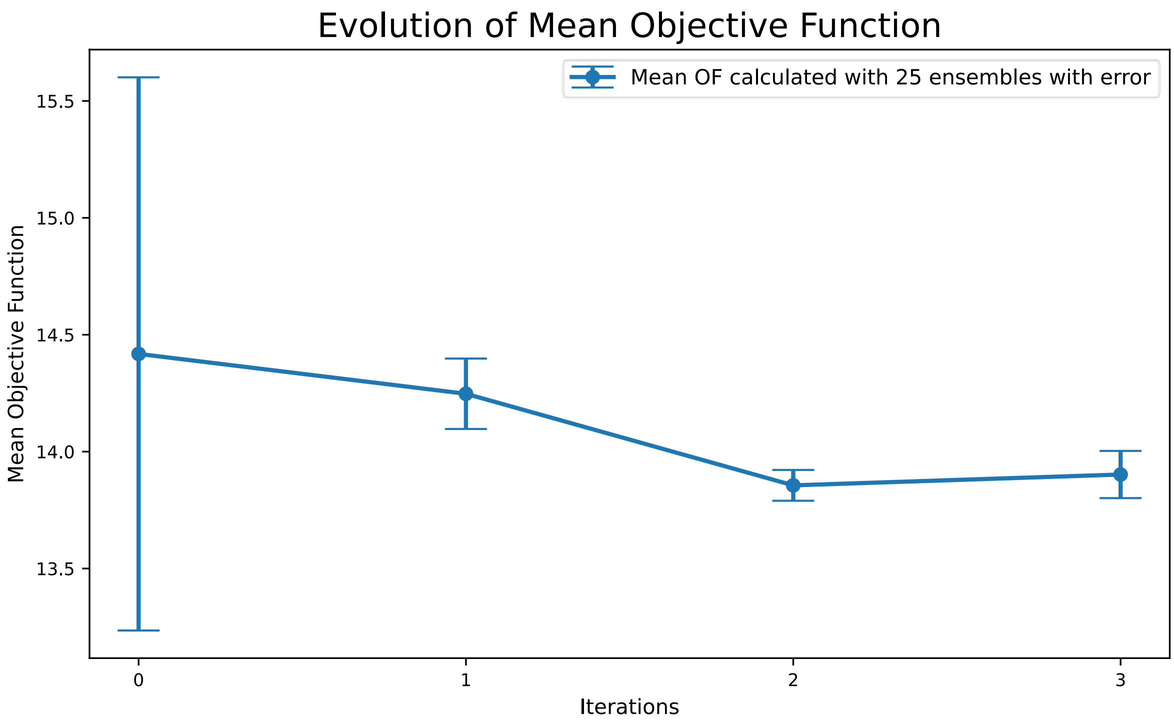


Figure 13 - Mean OF value evolution of first ES-MDA run of Alpha-East

The uncertain parameters values over the iterations are shown in Figures 14 and 15. The first figure shows the mean value calculated based on the numbers of ensembles and the error bar based on the standard deviation calculated from each iteration, giving a good estimation of the variance of each iteration of ES-MDA. The second figure depicts the spread of all the uncertain parameters with the related ensemble size Ne and the OF value for each ensemble.

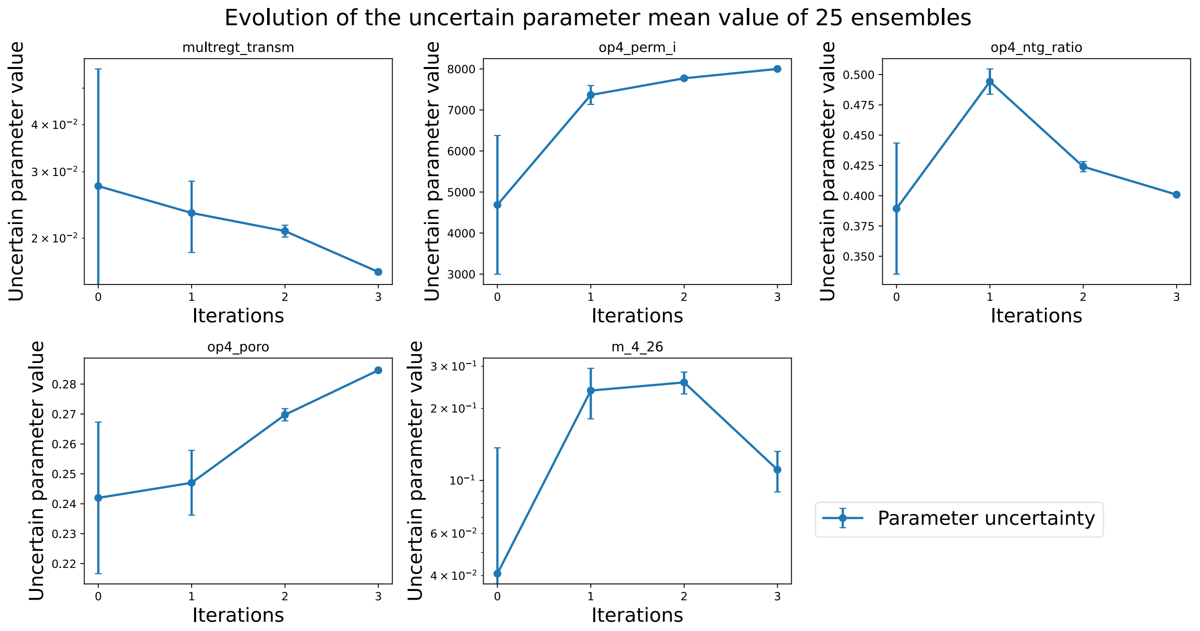


Figure 14 - Uncertain parameter evolution of first ES-MDA run on Alpha-East

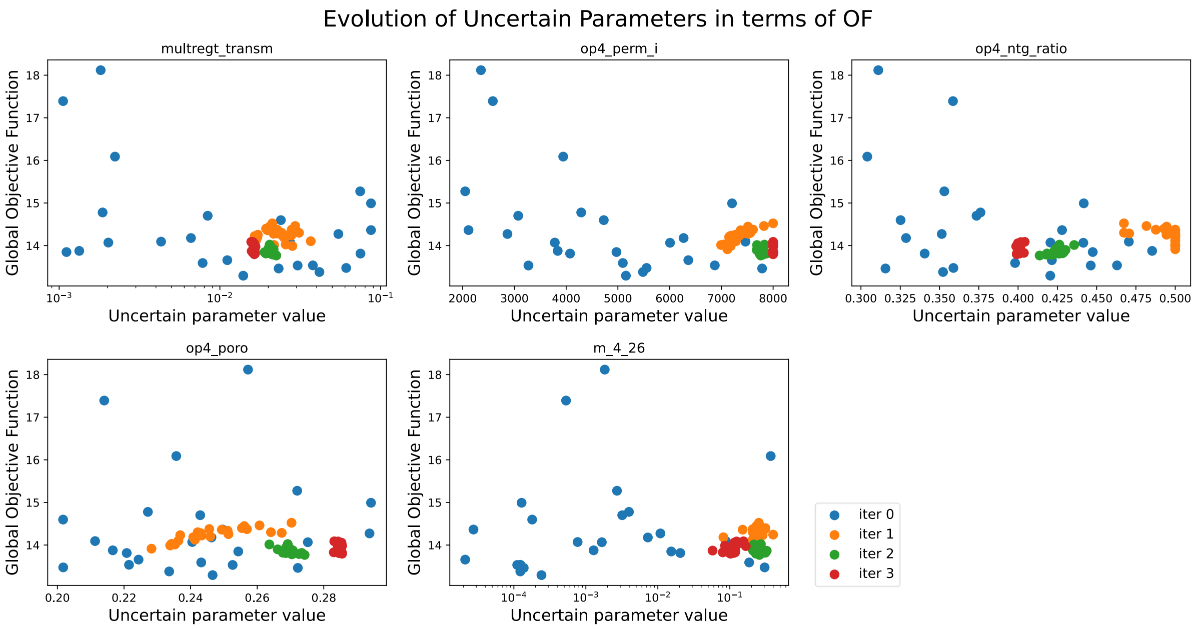


Figure 15 - Uncertain parameters scatter over the iterations of first ES-MDA on Alpha-East

Figure 14 shows that the errors are reduced for most of the parameters over the iterations, being the biggest variance related to the initial ensemble, as it is generated randomly from the *a priori* distributions. In Figure 15, it is possible to identify that, from an initial large spread of the five uncertain parameters values associated with a certain OF, ES-MDA reduces all spreads' ranges into the best interval of parameters for all properties to minimize the OF. For all uncertain parameters, it is shown this behavior, as the uncertain parameters begin to agglomerate in regions where the OF have smaller values. From iteration 0 to iteration 1, all parameters but the elementary channel's porosity (*op4_poro*) reached the regions where they remain gathered until the end of all ES-MDA procedure.

As the problem of reservoir fluid flow simulation is highly non-linear, it may exist the possibility of the OF to not have a global minimum, but several local minima. In this case, ES-MDA can converge to a certain range of uncertain parameters but give bad matches of the production vector properties. This happens because we are using a data assimilation method that depends on the number of ensembles employed; the smaller it is, the bigger are the uncertainties from each iteration, and higher is the probability of not obtaining a good match. For that reason, it is reasonable to pick a large number of ensembles for the next ES-MDA processes to be able to get more accuracy on the covariances calculated and better match results.

5.1.2 Second ES-MDA run on Alpha-East

In order to reduce the average simulated reservoir pressure after the long shut-in period of well PROD10, the aquifer volume was put into the new uncertain parameter list to see the behavior on the BHP in that simulation part. The aquifer is analytical and follows the Fetkovitch model. On the previous ES-MDA, the initial volume of the aquifer was 3.0 x 10^8 m^3. It is also worth mentioning that the multiplier transmissibility between MULTNUM regions 26 and 4 is an important parameter in order to obtain a correct average pressure of the reservoir.

For this new ES-MDA run, the PIs multipliers were considered as new uncertain parameters for both analyzed wells in Alpha-East. The two wells have a total of three completions, each one with its respective productivity index. The PI multiplier will try to simulate the skin factor in the vicinity in the wells. Furthermore, for the PROD10, after its closure in day 1190 (approximately), a second PI multiplier for each completion of the well will be applied, totalizing 9 PIs multipliers as new uncertain parameters.

Lastly, to better represent the geology complexity of the reality, transmissibility multipliers for the three fluid regions (66-68) will also be included in this run. The new uncertain parameter list and their explanation in Table 3. For this analysis, the number of ensembles used was $N_e = 100$ and with 5 iterations ($N_a = 5$) and, in addition to the previous production vectors properties used to get a match, the oil production rate of PROD10 was also considered in this run.

Table 3 - Uncertain parameter list for second ES-MDA run on Alpha-East

Parameters	Explanations
m_4_26	Transmissibility multiplier between distal levee and channel fairways
multregt_transm	Transmissibility multiplier between channel and proximal levee
op4_perm_i	Elementary channel permeability in i-direction (assumed isotropy and permeability in k-direction is 10% of i-direction)
op4_poro	Elementary channel porosity
op4_ntg_ratio	Elementary channel net-to-gross ratio
f_66_67	Transmissibility multiplier between flux region 66 and 67
f_67_68	Transmissibility multiplier between flux region 67 and 68
aq2_vol	Analytical aquifer initial volume
w12_pi_1	PI multiplier of INJ12 first completion
w12_pi_2	PI multiplier of INJ12 second completion
w12_pi_3	PI multiplier of INJ12 third completion
w10_pi_1_a	PI multiplier of PROD10 first completion at start of production
w10_pi_2_a	PI multiplier of PROD10 second completion at start of production
w10_pi_3_a	PI multiplier of PROD10 third completion at start of production
w10_pi_1_b	PI multiplier of PROD10 first completion at day 1092
w10_pi_2_b	PI multiplier of PROD10 second completion at day 1092
w10_pi_3_b	PI multiplier of PROD10 third completion at day 1092

The results obtained are shown in on Figures 16, 17, 18 and 19 shown below. In these figures, the initial ensemble spread is represented in grey scale, and the last ensemble is on red. The black curve is the best ES-MDA run given the global OF. The observation data is in blue, and its uncertain region is also represented given the observation data variance.

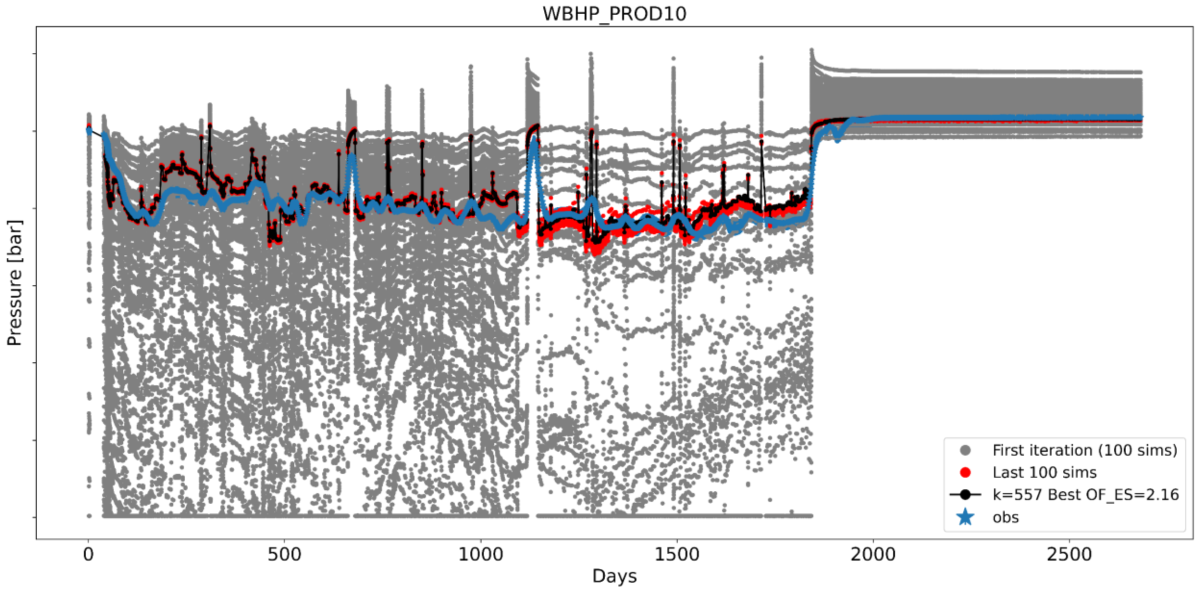


Figure 16 - Well bottom hole pressure simulation results from second ES-MDA run for the PROD10. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.

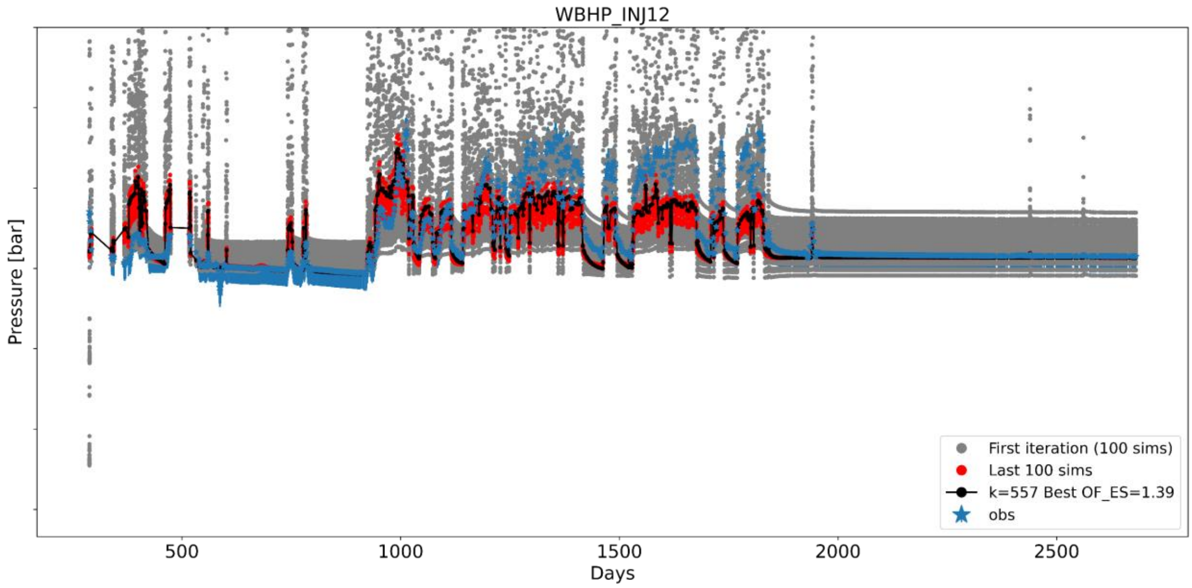


Figure 17 - Well bottom hole pressure simulation results from second ES-MDA run for the INJ12. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.

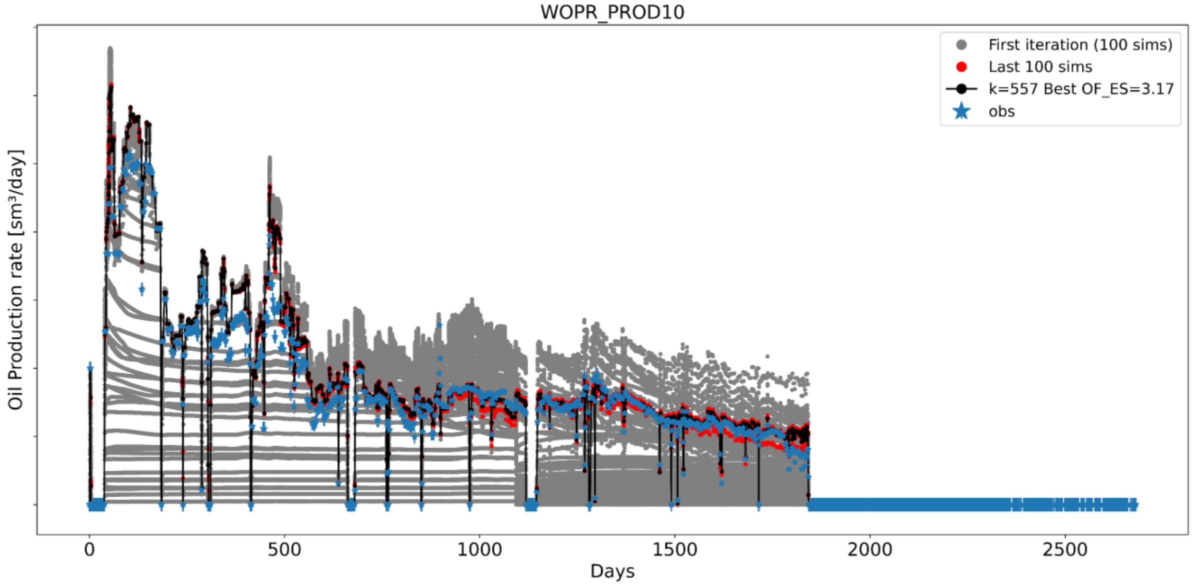


Figure 18 - Oil production rate simulation results from second ES-MDA for the PROD10. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.

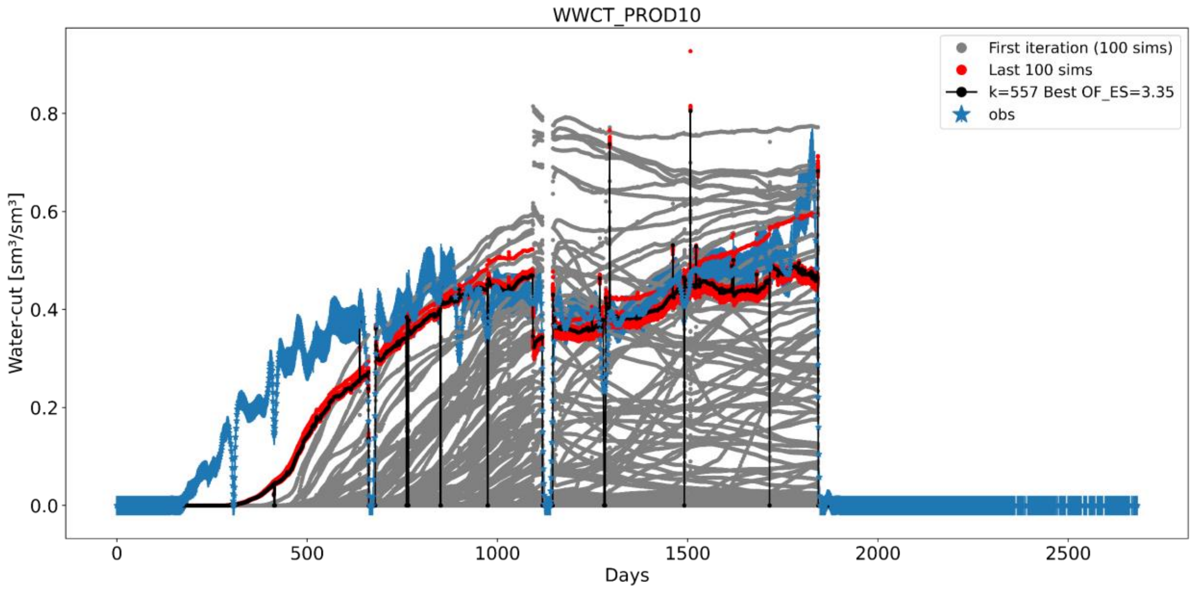


Figure 19 - Water cut simulation results from second ES-MDA run for the PROD10. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.

Analyzing the BHP of PROD10 in Figure 16, it is possible to observe that a bigger ΔP from the initial pressure and the final average reservoir pressure was obtained in comparison with Figure 8. Moreover, from the evolution of the mean uncertain parameters used in this ES-MDA run (Figure 21), the initial aquifer volume converged to a smaller value than the one used on the first run, confirming the hypothesis listed on the previous section about using a smaller aquifer volume in order to obtain a better match on the average reservoir pressure after the long pressure build-up.

Furthermore, it was obtained a good match on the flowing pressure of PROD10 in comparison to Box 2 of Figure 7. The inclusion of the 9 parameters related to the wells' productivity indexes and the flux regions multipliers are the main responsible for this match, as these PIs multipliers are the parameters directly related to this flowing step, affecting the flowing pressure and simulating the skin factor effects in the wells vicinity. PROD10's water breakthrough is still a little late compared with the observations, as shown in Figure 19, but after the well closure on day 1092, compared to Figure 9, the water cut is better match, this time obtaining values in the lower bound of the observation error interval.

From Figure 17 and the previous result obtained from the first ES-MDA run, the final ensemble spread is very similar with the one observed in Figure 10, with good matches on the static pressure and, in the flowing pressure, converging to smaller values compared to the observation data. Due to the high complexity of the reservoir's geology, it is possible to have elementary channels with different properties then the ones defined from ULIKE™ inside the uncertain parameter list, which are permeability, porosity, and net-to-gross ratio.

Figure 20 shows the evolution of the mean OF over the five data assimilation iterations. The value of the OF is bigger in this second run because the oil production rate of PROD10 was included as one term of the OF calculation. The picture shows an initial high OF value with a big uncertainty, but decreasing from iteration to iteration, with its variance following the same pattern. The evolution of the OF and the uncertain parameters spread over the iterations is shown in Figure 22, where is possible to visualize the convergence of the method over the iterations to a certain region of parameters values where the OF has low values.

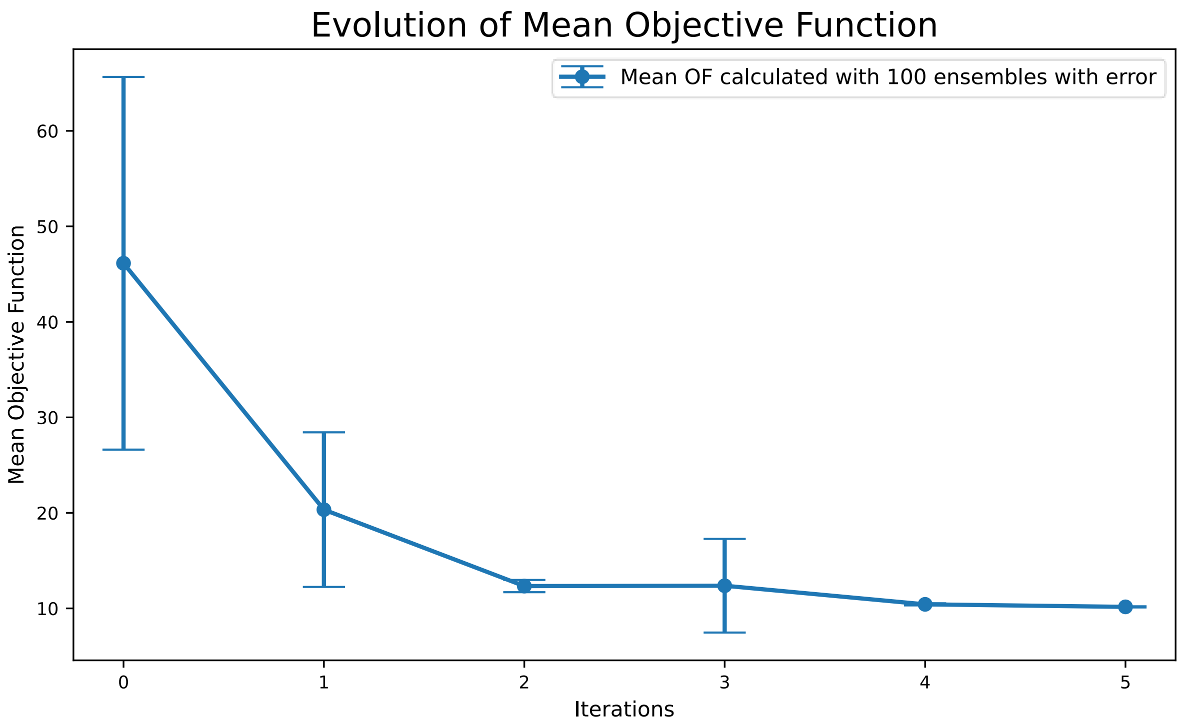


Figure 20 - ES-MDA’s second run mean OF statistics

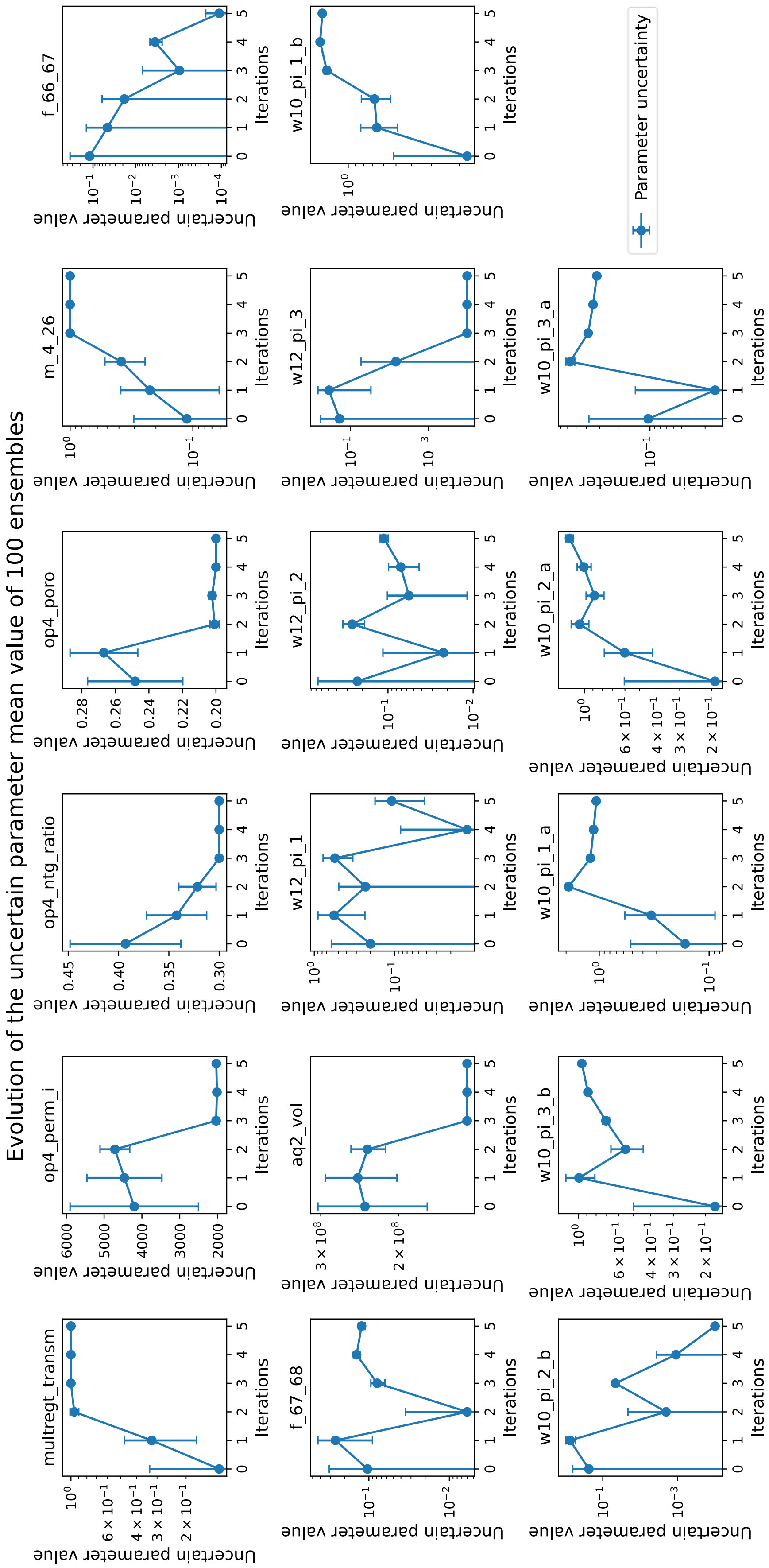


Figure 21 - Uncertain parameter evolution of second ES-MDA run on Alpha-East

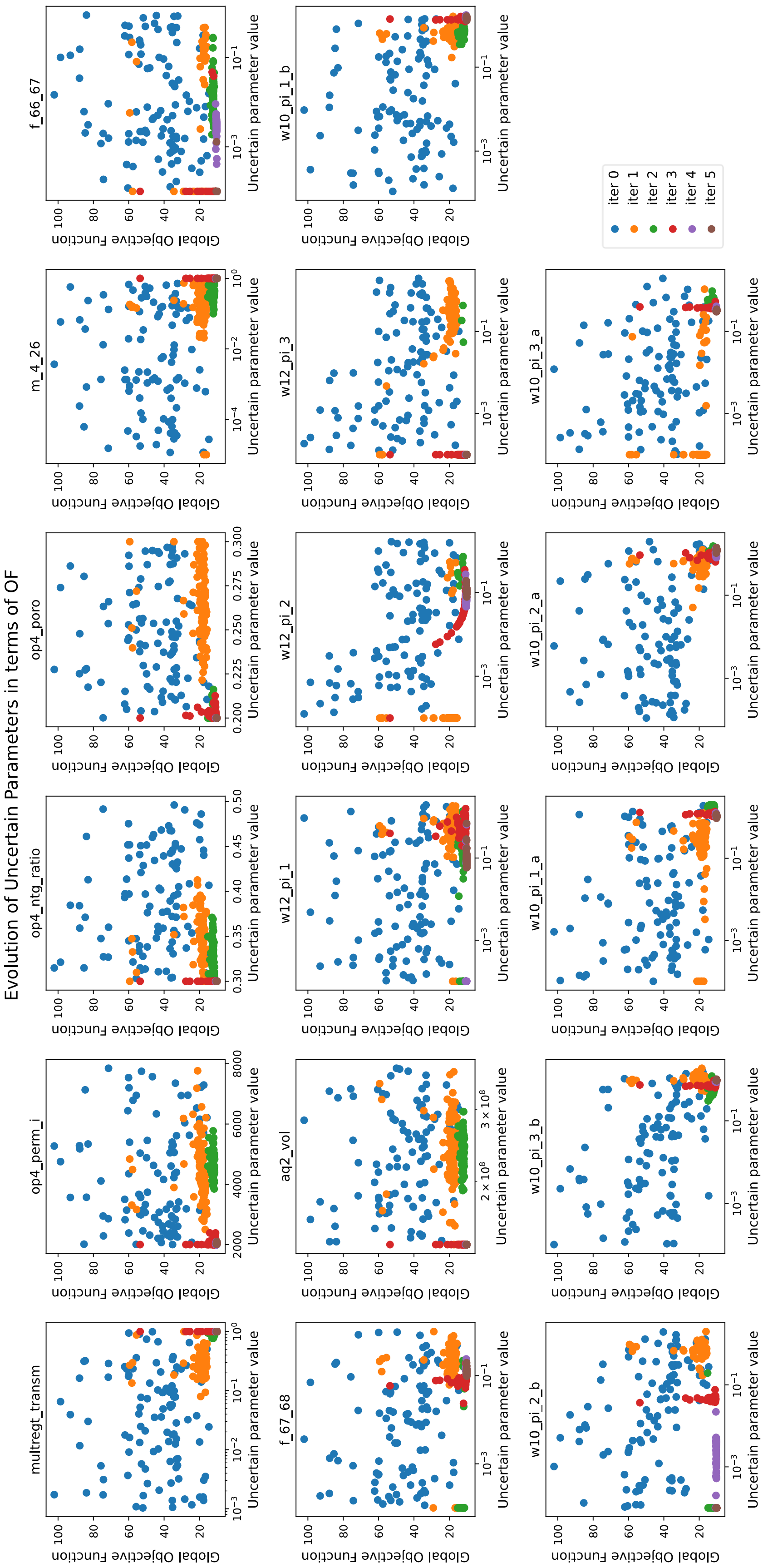


Figure 22 - Uncertain parameters scatter over the iterations of second ES-MDA run on Alpha-East

The scatter plots in Figure 22 show the quick convergence of ES-MDA for some uncertain parameters of choice, such as: *w10_pi_2_a*, *w10_pi_2_b* and *w10_pi_3_b*. For these parameters, mainly for iteration 2 and iteration 3, the scatter data tends to follow a vertical line with a very small variance (Figure 21). In other words, similar values of these uncertain parameters generate different OF values for different ensembles. This can happen as ES-MDA works with a large number of parameters combinations; hence, the same value of an uncertain parameter can generate an OF with small or big values, depending on the combination with another parameter that also contributed to this specific result. The uncertain parameters' scatter evolution over the iterations tend not to follow any specific or defined path in order to try to find the best combination for the defined OF. From an *a priori* scatter (iteration 0 of Figure 22) and following ES-MDA's update equations, their values are updated minimizing the OF and the ensemble variance (spread).

5.2 Alpha-Central: PROD6 – INJ8

An injector well (INJ8) and a producer well (PROD6) make up the Central block. Multiple faults have a significant impact on this section, reducing transmissibility within the core region but not blocking the flow. The aquifer and the elementary channels generated by ULIKE™ are used in the production mechanism.

The uncertain parameters considered in this section and their respective definition is shown in Table 4. For this analysis, the number of ensembles used was $N_e = 100$ and with 5 iterations ($N_a = 5$) and the observations to match were the same as the second ES-MDA run.

Table 4 - Uncertain parameter list for ES-MDA run on Alpha-Central

Parameters	Explanations
multregt_transm	Transmissibility multiplier between channel and proximal levee
op4_perm_i	Elementary channel permeability in i-direction (assumed isotropy and permeability in k-direction is 10% of i-direction)
op4_poro	Elementary channel porosity
op4_ntg_ratio	Elementary channel net-to-gross ratio
m_2_38	Transmissibility multipliers between faulted areas
m_38_39	
f_21_22	Flux multiplier on vertical connection between Maximum Flooding Surfaces
f_22_23	
p6_1	PI multiplier of PROD6 first completion
p6_2	PI multiplier of PROD6 second completion
p6_3	PI multiplier of PROD6 second completion
p8_1	PI multiplier of INJ8 first completion
p8_2	PI multiplier of INJ8 second completion
p8_3	PI multiplier of INJ8 third completion

The results obtained are on Figures 23, 24, 25 and 26 shown below. In them, the initial ensemble spread is represented in grey scale, and the last ensemble is on red. The black curve is the best ES-MDA run given the global OF. The observation data is in blue, and its uncertain region is also represented given the observation data variance.

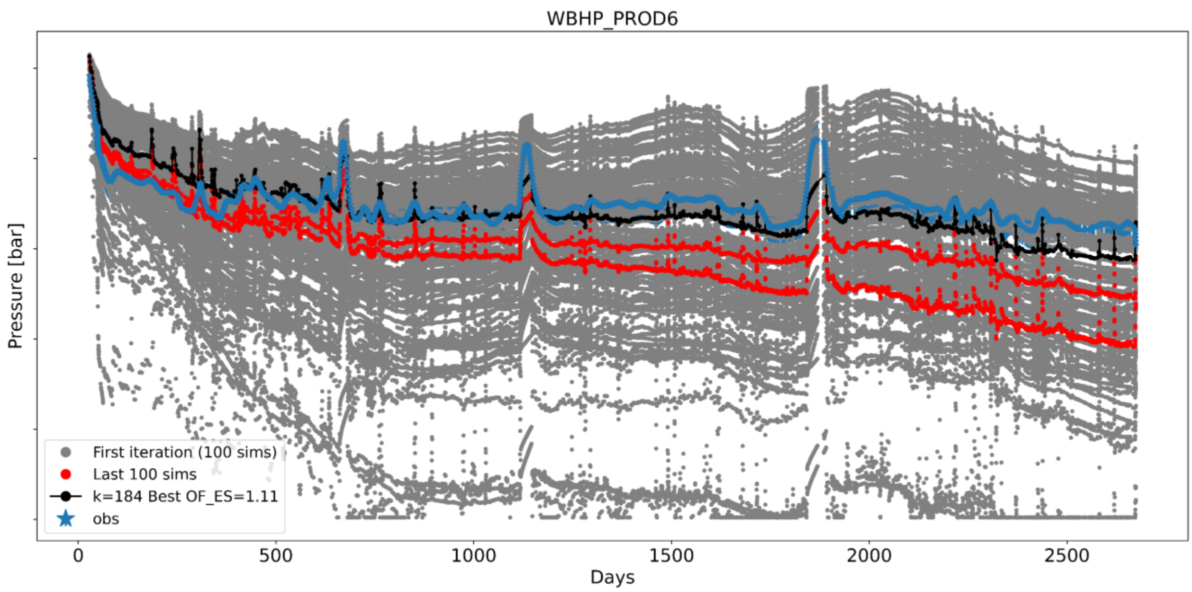


Figure 23 - Well bottom hole pressure simulation results from ES-MDA run for the PROD6. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.

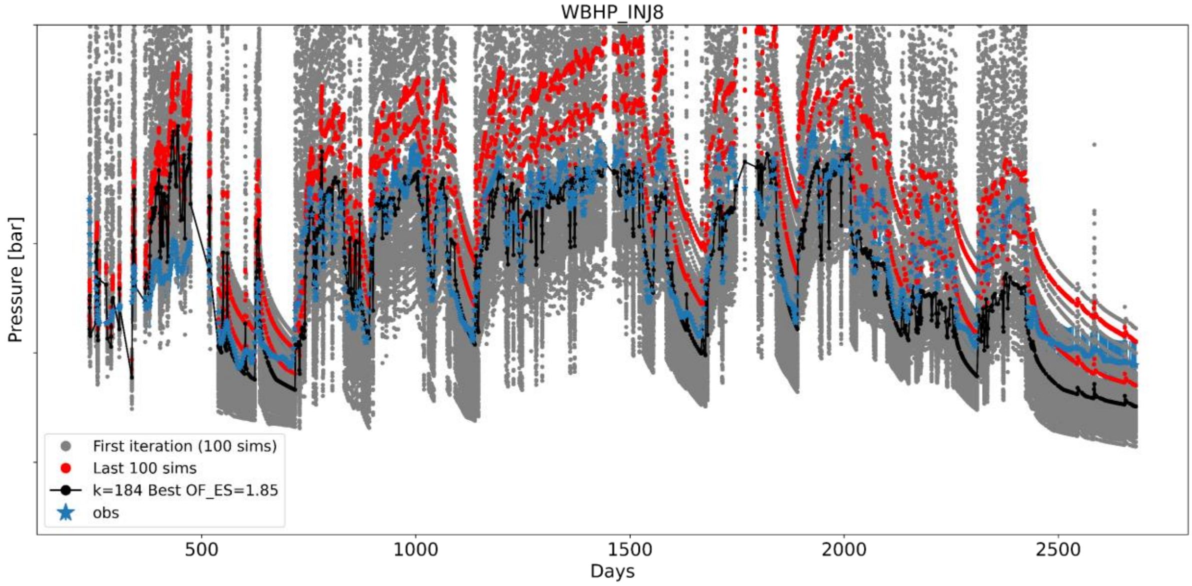


Figure 24 - Well bottom hole pressure simulation results from ES-MDA run for the INJ8. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.

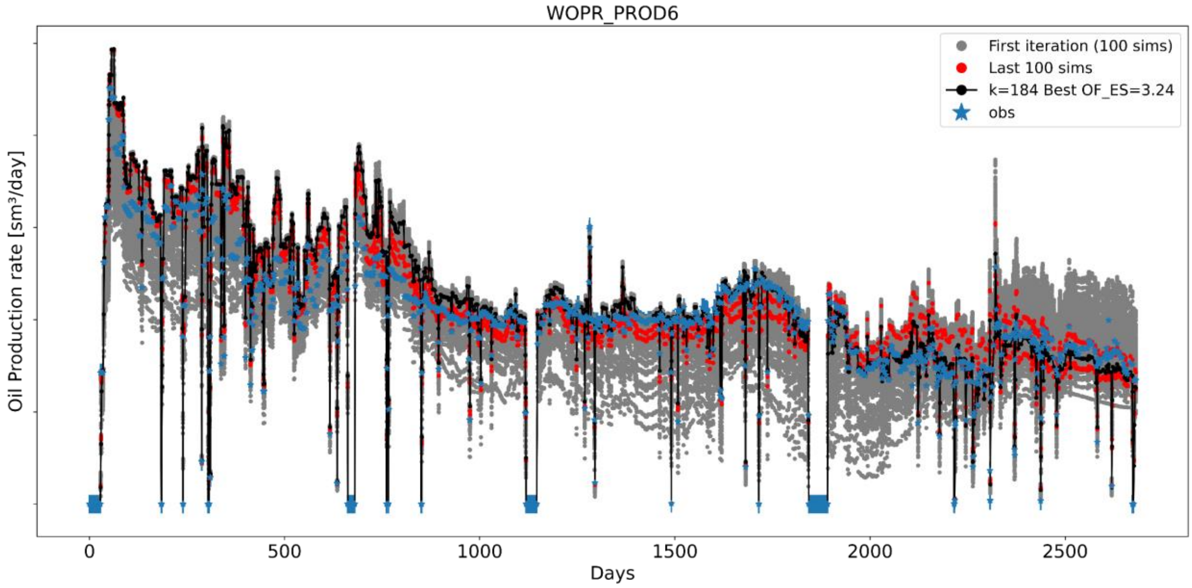


Figure 25 - Oil production rate simulation results from ES-MDA run for the PROD6. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.

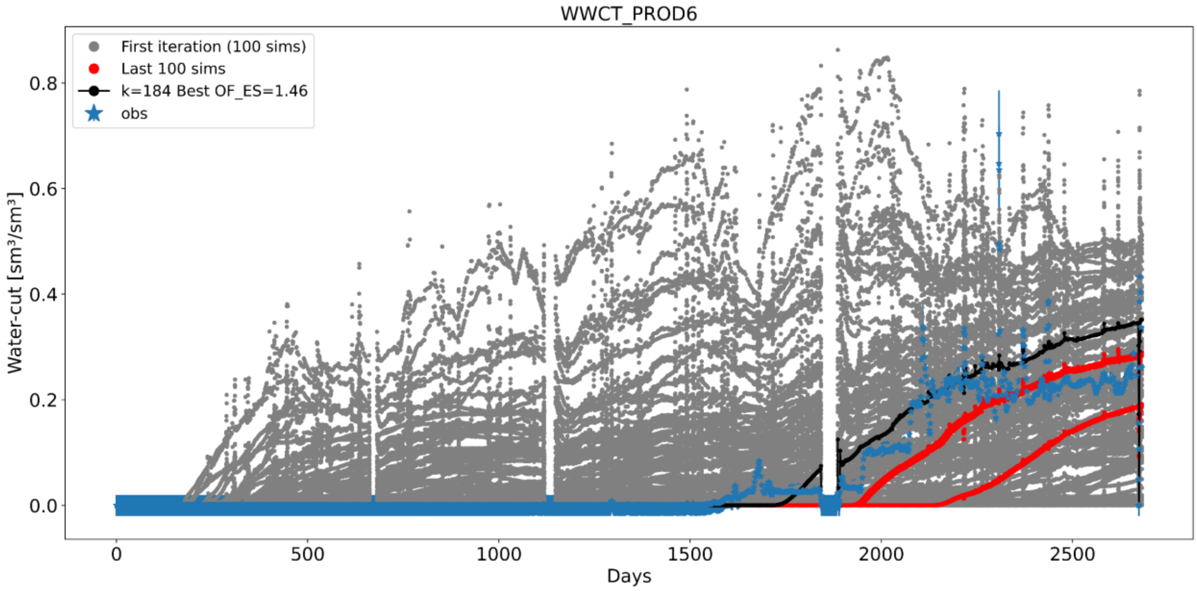


Figure 26 - Water cut simulation results from ES-MDA run for the PROD6. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.

The BHP of PROD6 (Figure 23) shows that the simulated values are very close to the observation, mainly for the flowing part. Regarding the static pressure (all three build-ups), there is gap between simulated and observed data, which could indicate a lack of connected volume between the injector INJ8 who helps to maintain the pressure for the producer. About INJ8 BHP, we see in Figure 24 that ES-MDA manages to get a good match on both fluid and static pressure. For instance, as observed mainly in the three falloffs between days 1000 and 2000, the short-term pressure drop is well matched, meaning that the behavior in the well's vicinity is correctly guessed with the parameters of choice: in this case, the productivity index multiplier on the three completion layers. In the final days of the simulation, mainly for INJ8, there is relatively large mismatch on the pressure drop. This could also be an indicative that there is still some volume connectivity missing inside the model which could be improved in further analysis.

Figures 25 and 26 show, respectively, the oil production rate and the water breakthrough of PROD6. Both properties are well matched during the whole simulated timesteps. Nonetheless, in Figure 26 we observe two more families of simulated ensembles (red curves) in the final ES-MDA iteration. This means that, if continued, the assimilation processes could be able to find not only a more significant representation of the water cut then it already is, but also for the others three production vector properties analyzed. The water cut of a production well is not only function of the injection rate of the injector well helping to maintain a good pressure value for the oil recovery from the producer, but also on the faults' transmissibility, which are all considered inside our uncertain parameter lists and the geological model applied for this ES-MDA run.

Figure 27 shows the evolution of the mean OF over the five data assimilation iterations. As previously mentioned, in the final iterations, ES-MDA generated some new families of ensembles that do not follow the same mean behavior, hence the value increase of OF from iteration 3 to 4. These new ensemble families generated from iteration 4 to 5 are not so dispersed around the mean, meaning that the variance estimated from the ensembles are reduced, as shown in the reduction of the error bar from the picture. Nevertheless, for the history matching results and analysis of the uncertain parameters, the data assimilation process should stop at iteration 2, as it reached a possible minimum of the OF.

The evolution of both, mean uncertain parameters' values and OF in respect with the uncertain parameters spread over the iterations are shown in Figures 28 and 29, respectively.

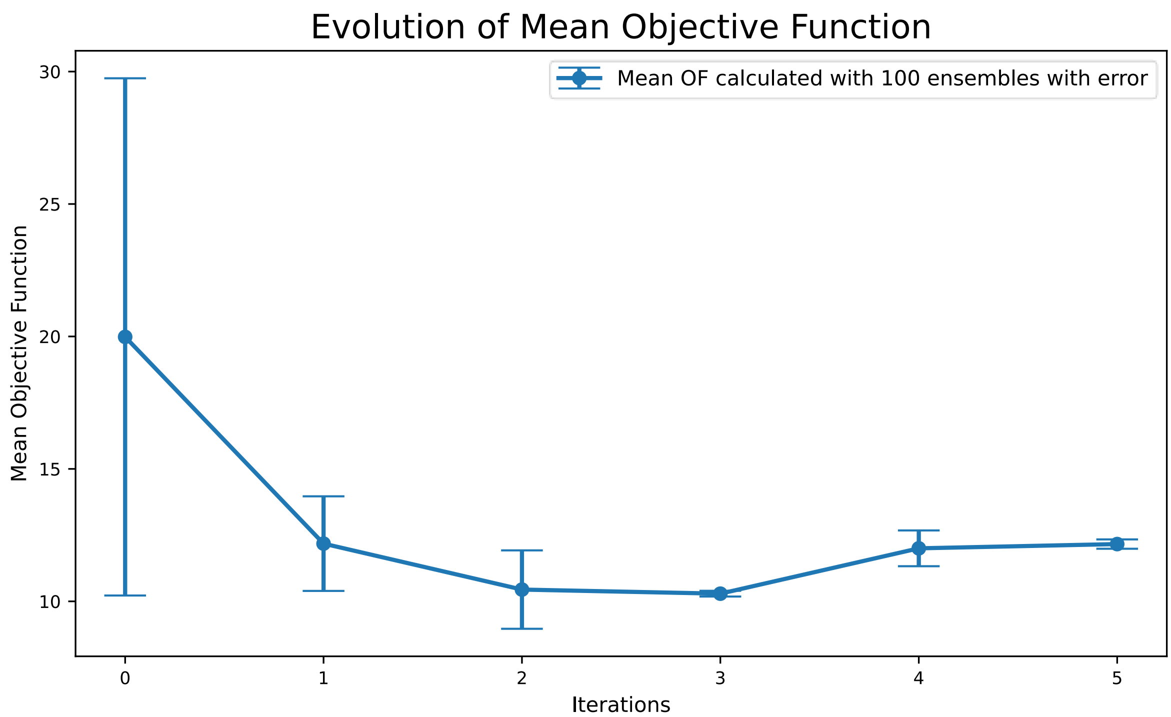


Figure 27- ES-MDA run mean OF statistics on Alpha-Central

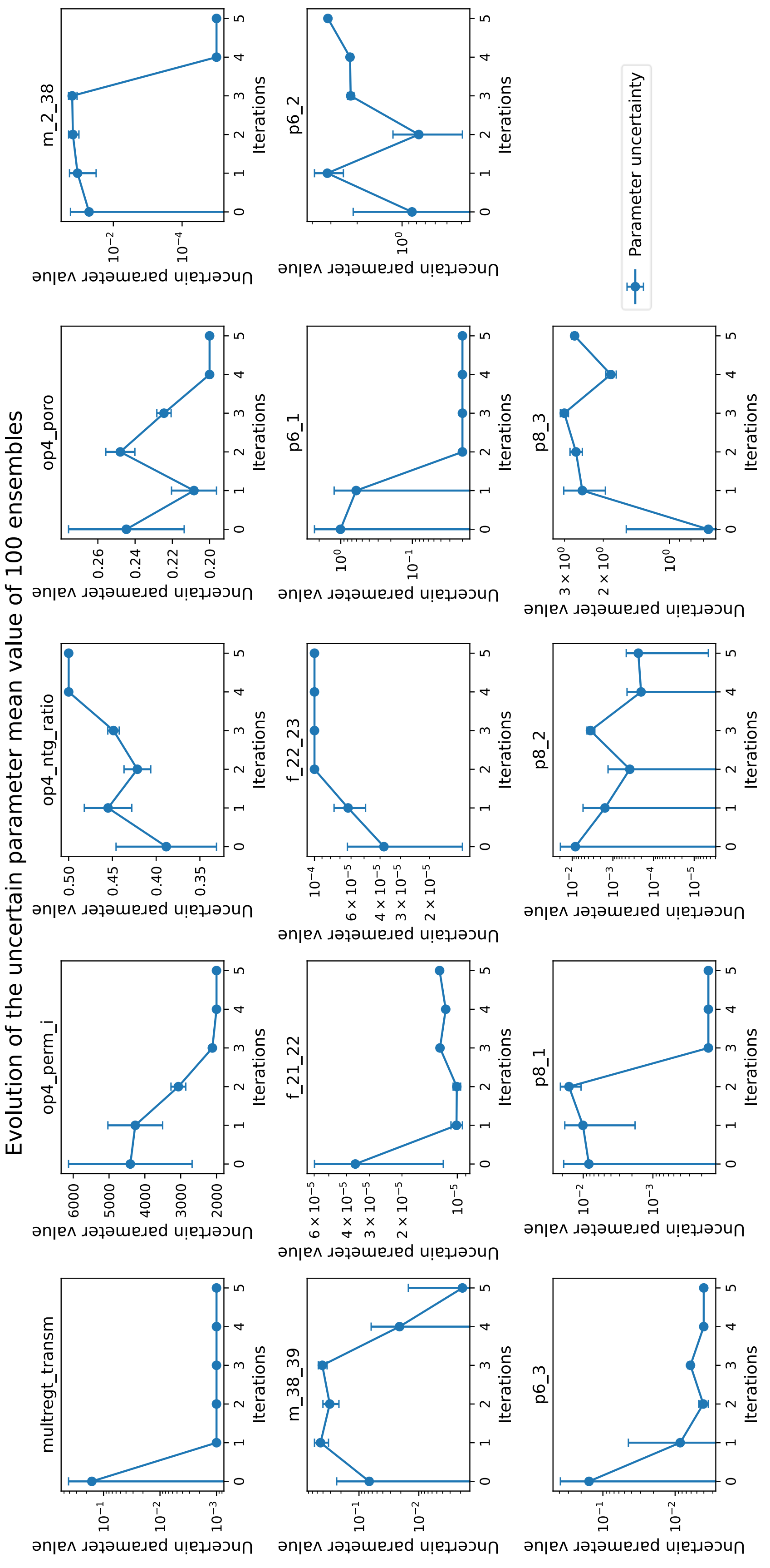


Figure 28- Uncertain parameter evolution of ES-MDA run on Alpha-Central

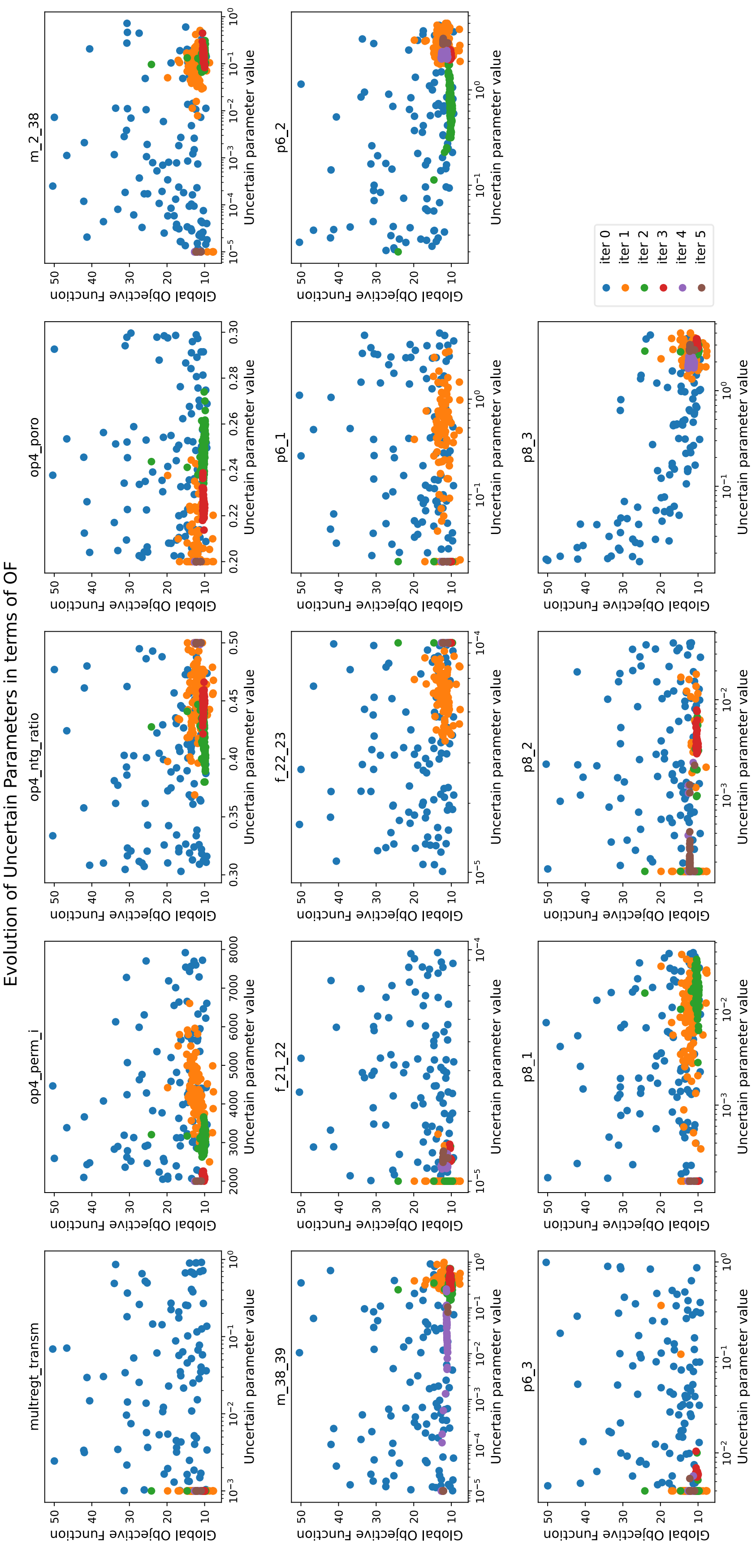


Figure 29 - Uncertain parameters scatter over the iterations on Alpha-Central

Figure 29 shows the same process of uncertain parameters' scatter and evolution of the iterations of Figures 15 and 22. However, in Figure 29 it is noticeable a slower convergence of certain uncertain parameters values over the iteration, as shown in the parameters: *op4_perm_i, op4_ntg_ratio, op4_poro, m_38_39* and *p_8_2*. In contrast, other parameters such as *multregt_transm* and *f_21_22* converge early in iteration 1 and do not undergo further significate updates over the next iterations. This can be an example of spurious correlation which can interfere with the data assimilation problem. Besides, as shown in Figures 23, 24 and 26, the last iteration did not generated one family of red curves with a relatively small ensemble variance, but at least two or three. This directly impacts the process of finding a global minimum and the best combination of uncertain parameters for the problem.

5.2.1 Analysis of Alpha-Central using EST's manual history matching with interpolated responses

As detailed in *Section 2.3.3*, ES-MDA is a multiple data assimilation method used in inverse problems where the mismatches between simulated and observation data are reduced along the data assimilation, hence giving the best combination of a number of uncertain parameters defined for the forecast of the model response. One advantage of ES-MDA is that there is no limitation in the number of uncertain parameters defined to be used on the model to obtain the simulated data, as long as it is provided for each member of the uncertain parameter list the statistics necessary in order to build a probability distribution function for the initial random sampling to initiate the process.

In this section of the work, the software EST – an internal software from TotalEnergies used for history matching – was used to also try to obtain a good match on the Alpha-Central. The motivation of this comparison was the difficulty in finding a good range for the PROD6 and INJ8 productivity index for the initial ensemble of ES-MDA, which often generated spurious correlation values and parameters collapses in earlier iterations, not obtaining a good first spread over the observation data and, thus, failing to get a match. The goal of EST is to find a good response of the simulation data that fits the chosen parameters as well as the desired model responses. This model could be a kriging surface or a polynomial regression.

The kriged responses inside EST are created using second-degree polynomials. Therefore, a minimum, maximum, median value and, thus, a distribution of logarithmic or linear parameters must be established for each parameter. The median is required to calculate the second-degree coefficients for polynomial regression. After that, the regression should be as near to the experimental response points as possible. It will be required to determine a combination of factors in order to discover the maximum likeness to the experimental points.

As opposed to ES-MDA, which is a method without limitation on the numbers of uncertain parameters, the maximum handled by EST are 11. Therefore, some parameters were removed (*op4_ntg_ratio, f_22_23, f_21_22*) from the list in Table 4 to make this comparison and, in the INTERSECT™ input file, their values were replaced with the best ones found from the previous ES-MDA. In Figure 30, the simulations from which EST finds the polynomial regression parameters are represented in blue, with a total of 151 curves. From that, it is necessary to manually find the best combination of parameters that approximated most the observation data. That kriged response is shown in the red line.

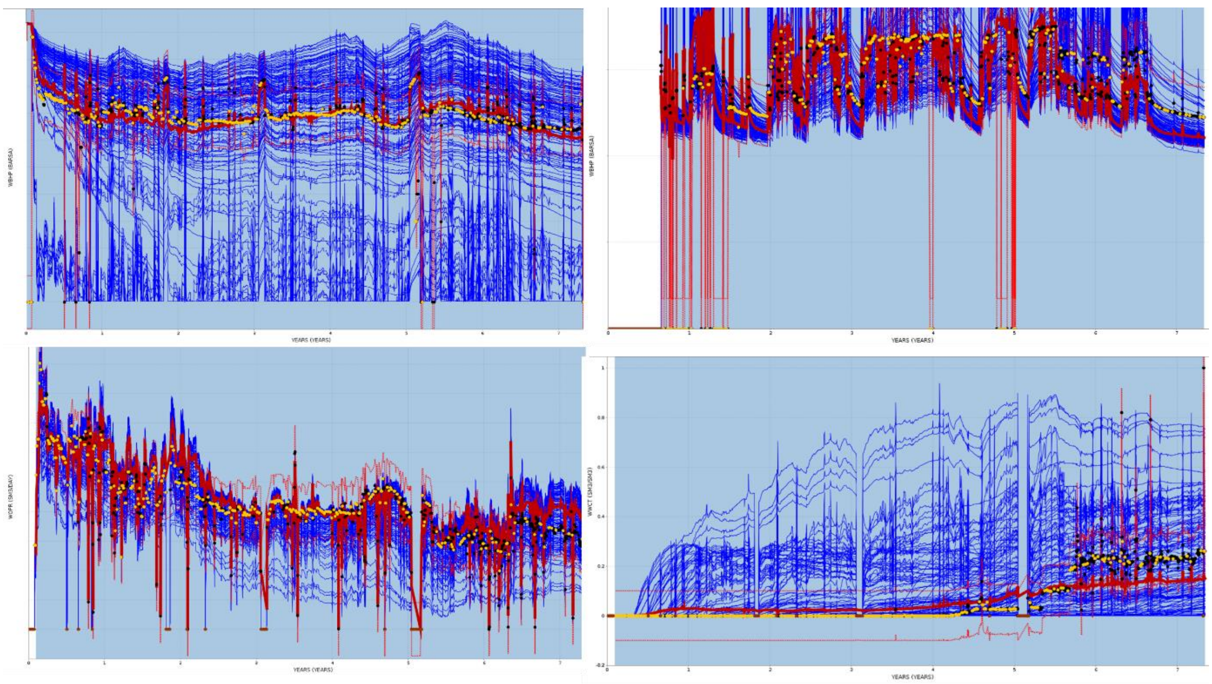


Figure 30 - From top to bottom – left to right: PROD6 bottom-hole pressure; INJ8 bottom-hole pressure; PROD6 oil production rate; PROD6 water cut. The blue curves are the 151 reservoir simulations. The red line is the best kriged response found.

With the combination of the parameters that generated the best kriged response, it is also necessary to confirm with the fluid flow simulator the real effect of those set of parameters, as there is a high uncertainty related to these interpolation processes and also the high non-linearity of the physics related to the problem. Figure 31 depicts the response from the combination of the uncertain parameters get from EST. We observe that in a general way, the simulation results follow the same trend of the observation data points, and the match is obtained for some production properties, like the PROD6 bottom hole pressure and oil production rate. However, is it on the INJ8 bottom hole pressure where we can see the biggest discrepancies from EST kriged response and the real simulated one, especially on the flowing pressure (days 1100 to 1600, approximately).

To improve the real response from EST kriged curves or even find a better combination of uncertain parameters, the process just described above for this workflow could be done iteratively. In other words, with the real response given by the fluid flow simulator, new ranges of uncertain parameters could be defined and retuned to EST to make a new analysis to obtain better matches.

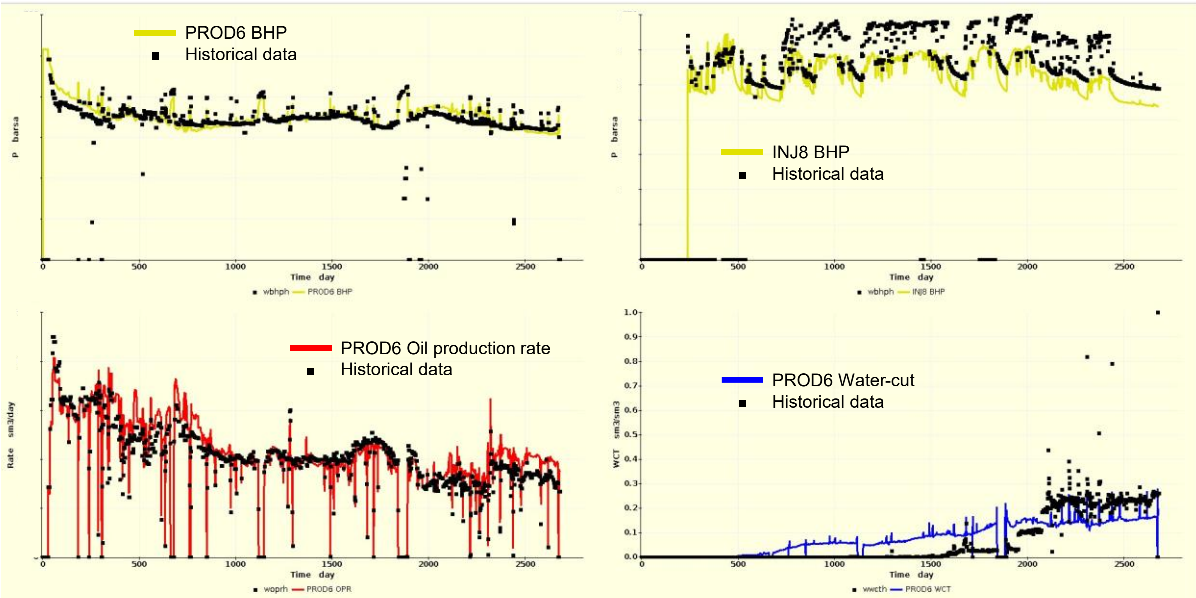


Figure 31 - From top to bottom – left to right: PROD6 bottom-hole pressure; INJ8 bottom-hole pressure; PROD6 oil production rate; PROD6 water cut. The continuous lines indicate the simulated production vector property. The black squares are the historical data.

6 CONCLUSION

This final graduation work presented a history matching workflow application for a real scale reservoir model with complex geological models, using data assimilation methods, more specifically an ensemble-based method. Even while ensemble-based methods have been effectively used in many applications over the past decades, inverse problems with significant nonlinearity, such as reservoir history matching, continue to pose difficulties. The Gaussian assumption that underlies this class of approaches are primarily to blame for these issues.

This work analyzed the real scale reservoir, denominated Alpha Field, into compartments due to the faults located in the area, which divided the turbiditic reservoir into different areas and regions. As mentioned, reservoir modeling is considered to be a complicated and time-consuming activity during the research and development process. This fact is mainly due to the geological complexity of petroleum reservoirs and the necessity of modeling the best was possible the geological features that may impact the fluid flow simulator results and, thus, the exploration activity. That said, it was proved the efficiency of the rule-based modeling tool ULIKE™, used to characterize the elementary sand channels and its respective fairways where the fluid flow is taken place and respecting all wells constraints of the field. These results affect directly in the history matching process of the analyzed area, as all the properties observed and the results obtained from the reservoir simulator is linked with the geology and how it is represented in the model, for instance, the connectivity between wells and how this impacts the bottom hole pressure.

Using data assimilation based on ensemble-based methodology to perform the history matching of the study area, the ES-MDA algorithm proved to be functional for the study case, being capable of yielding ensembles of models that honor the geological model and observation data, generating sets of history-matched models of good quality, both qualitatively and quantitatively. From the results of the analysis of both reservoir compartments, the following conclusions are deduced:

- In order to be able to check all the uncertain parameters responses for every ensemble member, it was necessary a high number of simultaneous reservoir fluid flow simulations, obtaining, thus, an ensemble of the forward operator used to the data assimilation problem. In average, for each reservoir compartment analysis using 5 assimilations iterations and an ensemble size of 100, it was necessary a minimum of 600 reservoir simulations to obtain a good representativity from the model responses and then, observe the quality of the assimilations;

- For all ES-MDA procedures made during this work, it was noted how much the initial ensembles models, resulted from the reservoir responses with the uncertain parameters sampled from the *a priori* distributions, affects the assimilation and history matching results. Many initial ensembles that did not have a big spread and did not cover the observation data failed to converge to a good history-matched ensemble of models, despite the use of inflation coefficients factors chosen in the beginning of the process. Therefore, the *a priori* distributions and the randomly sampled uncertain parameters contribute to achieving good results, whereby their ranges, mean/standard deviation values should be carefully chosen and analyzed previously during screening studies and sensitivity analysis;
- From the results shown in Figures 16 − 19 and Figures 23 − 26, it was shown the quick convergence of the method. This is mainly due to the use of inflation coefficient factors and how ES-MDA was formulated, minimizing the ensemble variance in each assimilation iteration. Hence, the choice of these factors and the assimilations numbers also have a direct impact on the results. For both reservoir compartments analyzed, five data assimilations iterations were enough for providing good data matches;
- The ensemble size also impacted on the history matching results. Small ensembles size failed to result in good history-matched models and also did not manage to give good uncertainty representations, regardless of the method's convergence. In this work, for the dataset analyzed, an average ensemble size of 100 was the best value in which was possible to better analyze the results, even if the ES-MDA processes were not completed, due to computational issues or any other reason. Small ensembles size, however, can give a good first representation of the uncertain parameters choice and how will them behave with the reservoir simulator responses;
- From the total of uncertain parameters used, the ones that impacted the most during the reservoir fluid flow simulator were the wells productivity index

multipliers. The productivity index of the well is one parameter that represents the behavior of the fluid flow near the well, and the multiplier on them tries to simulate the skin factor. Therefore, it can be concluded that these parameters directly affected the dynamic pressure values, but also it was confirmed a possible non-linearity regarding these parameters, in which very small changes generated ensembles with completely different responses. These responses affected some simulations where the reservoir simulator did not completely finish the run, and, in order to continue the data assimilation method, those failed simulations had to be re-run.

The ES-MDA approach's overall performance is encouraging, and it offers great perspectives for the use of this data assimilation method in a real scale reservoir with a considerate geological complexity. The easy implementation of this type of ensemble smoother is desirable, being able not only to be used for reservoir engineering applications, but also for other data assimilation processes. The major difficulties and deficiencies encountered are in relation to spurious correlations and large loss of variability of the model ensemble, which can result in a poor estimate of reservoir uncertainties causing a poor prediction of field life, even with a good historical fit. This fact can be mitigated by applying adaptive methodologies, where the damping of each iteration as well as the number of iterations are set automatically and using covariance localization methods to reduce the loss of variance due to possible sampling errors.

In addition to the data assimilation process using ES-MDA, it was used the internal software for history matching from TotalEnergies, EST, to analyze one of the compartments of the reservoir. Experimental designs models that use interpolations and proxy models in order to give an approximate response to inverse problems can drastically reduce the computational cost necessary to run some forward operator, such as the reservoir simulator. However, as these responses are considerate to be an estimate of the real problem, it is always necessary to analyze the results obtained and check if they are feasible, respecting the physics of the problem involved and if the uncertainty is still well represented. With the results obtained from the work using this method, some conclusion can be made:

- The dynamism of the software, giving instantaneous interpolated responses is its main attraction. Even with its limitations, mainly with the maximum numbers of uncertain parameters that can be considered, it is possible to have innumerous combinations, having the opportunity to find the best combination possible that best approximates the responses to the observation data. However, because it is a manual process of choosing the parameters values and visualizing the response in the interpolated model, this process can become extremely time-consuming, with the possibility in not finding a good match with the uncertain parameters and its ranges previously defined, being necessary to redo the process, either with other parameters or different ranges;

- Even though the kriging responses are instantaneous, it is only possible to analyze the response changing one parameter at a time. That said, and knowing that with petroleum reservoirs, many parameters are related with each other (e.g., permeability/porosity) this can become an issue if the production property vector in which we want to obtain a match is large and there is a significant number of uncertainty parameters that is known to be correlated or to have a possible correlation with each other;
- From the analysis of real response given by the reservoir simulator, it was observed some discrepancies between the kriged response given by the best combination found of the uncertain parameters. These differences are greater on the dynamic pressures values from the injector's bottom hole pressure. This could indicate a possible high non-linear effect of the productivity index's multipliers, as the kriged responses work with interpolations of second degree, which might not be able to correctly describe this production property.

With EST, the possibility of quickly screening the impact of some selected parameters as most influent that are used to calibrate the model with kriging responses can also give interactive insights about how the model is impacted with this specific parameter variation. EST history matching process was used considering the uncertain parameter as known. Due to the absence of the real reservoir response during the manual history matching process, EST multi solutions process could have been used. For that matter, this methodology is iterative after having checked with the real reservoir response, repeating the process until certain stop criteria arrive. The criteria are based on the cost function computation. Nevertheless, EST with screening options could have also been used during initial phases of reservoir history matching, in which the objective is not to get an history-matched model, but to determine the most influent parameters from a large number of uncertain parameters (roughly 100) choices and ranges from their impacts on the production property vector. Those most influent parameters could serve as inspiration and, possibly, inputs of the *a priori* distribution used afterwards in history matching of petroleum reservoirs using data assimilation methods, such as ES-MDA.

BIBLIOGRAPHY

AANONSEN, S. I. et al. The Ensemble Kalman Filter in Reservoir Engineering—a Review. **SPE Journal**, v. 14, n. 03, p. 393–412, 27 set. 2009.

ABADPOUR, A. et al. Integrated Geo-Modeling and Ensemble History Matching of Complex Fractured Carbonate and Deep Offshore Turbidite Fields, Generation of Several Geologically Coherent Solutions Using Ensemble Methods. **Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2018, ADIPEC 2018**, 12 nov. 2018.

AMUDO, C. et al. The Pains and Gains of Experimental Design and Response Surface Applications in Reservoir Simulation Studies. **SPE Reservoir Simulation Symposium Proceedings**, v. 1, p. 46–54, 2 fev. 2009.

ANDERSON, J. L. An adaptive covariance inflation error correction algorithm for ensemble filters. **New pub: Stockholm uni Press**, v. 59, n. 2, p. 210–224, mar. 2016.

ANTERION, F.; EYMARD, R.; KARCHER, B. Use of parameter gradients for reservoir history matching. **Society of Petroleum Engineers of AIME, (Paper) SPE**, 1989.

BISSELL, R.; KILLOUGH, J. E.; SHARMA, Y. Reservoir History Matching Using the Method of Gradients on a Workstation. p. 61–66, 24 maio 1992.

BJARKASON, E. et al. **Uncertainty Quantification of Highly-Parameterized Geothermal Reservoir Models Using Ensemble-Based Methods**. World Geothermal Congress. **Anais**...Iceland: jul. 2021.

BRASSEUR, P. Ensemble-Based Data Assimilation Methods. Em: **Operational Oceanography in the 21st Century**. Dordrecht: Springer Netherlands, 2011. p. 381–393.

BRUYELLE, J.; GUÉRILLOT, D. Proxy Model Based on Artificial Intelligence Technique for History Matching - Application to Brugge Field. **Society of Petroleum Engineers - SPE Gas and Oil Technology Showcase and Conference 2019, GOTS 2019**, 21 out. 2019.

BURGERS, G.; JAN VAN LEEUWEN, P.; EVENSEN, G. Analysis Scheme in the Ensemble Kalman Filter. **Monthly Weather Review**, v. 126, n. 6, p. 1719–1724, jun. 1998.

CARRASSI, A. et al. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. **Wiley Interdisciplinary Reviews: Climate Change**, v. 9, n. 5, p. e535, 1 set. 2018a.

CARRASSI, A. et al. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. **WIREs Climate Change**, v. 9, n. 5, 9 set. 2018b.

CHEN, H. et al. Assessing probabilistic modelling for wind speed from numerical weather prediction model and observation in the Arctic. **Scientific Reports 2021 11:1**, v. 11, n. 1, p. 1–11, 7 abr. 2021.

CHEN, Y.; OLIVER, D. S. History Matching of the Norne Full-Field Model With an Iterative Ensemble Smoother. **SPE Reservoir Evaluation & Engineering**, v. 17, n. 02, p. 244–256, 8 maio 2014.

CHO, D. et al. Comparative Assessment of Various Machine Learning-Based Bias Correction Methods for Numerical Weather Prediction Model Forecasts of Extreme Air Temperatures in Urban Areas. **Earth and Space Science**, v. 7, n. 4, p. e2019EA000740, 1 abr. 2020.

COHN, S. E. An introduction to estimation theory. **Journal of the Meteorological Society of Japan**, v. 75, n. 1 B, p. 257–288, 1997.

COSTA, L.; MASCHIO, C.; SCHIOZER, D. **Study of the influence of training data set in artificial neural network applied to the history matching process**. Rio Oil & Gas Expo and Conference. **Anais**...2010.

COURANT, R.; HILBERT, D. Methods of Mathematical Physics. **Methods of Mathematical Physics**, v. 1, p. 1–560, 24 dez. 2007.

E BUCKLEY, B. S.; LEVERETT, M. C.; AIME, M. Mechanism of Fluid Displacement in Sands. **Transactions of the AIME**, v. 146, n. 01, p. 107–116, 1 dez. 1942.

EIDE, A. L. et al. Automatic History Matching by use of Response Surfaces and Experimental Design. p. cp-233-00045, 7 jun. 1994.

EMERICK, A. **History Matching and Uncertainty Characterization Using Ensemble-based Methods**. Doctor of Philosophy—Tulsa: University of Tulsa, 2012.

EMERICK, A. A. Analysis of the performance of ensemble-based assimilation of production and seismic data. **Journal of Petroleum Science and Engineering**, v. 139, p. 219–239, 1 mar. 2016.

EMERICK, A. A. Analysis of geometric selection of the data-error covariance inflation for ES-MDA. **Journal of Petroleum Science and Engineering**, v. 182, p. 106168, 1 nov. 2019.

EMERICK, A. A.; MORAES, R.; RODRIGUES, J. History Matching 4D Seismic Data with Efficient Gradient Based Methods. **All Days**, 11 jun. 2007.

EMERICK, A. A.; REYNOLDS, A. C. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. **Computational Geosciences 2011 16:3**, v. 16, n. 3, p. 639–659, 26 jan. 2012.

EMERICK, A. A.; REYNOLDS, A. C. Ensemble smoother with multiple data assimilation. **Computers & Geosciences**, v. 55, p. 3–15, jun. 2013.

EVENSEN, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. **Journal of Geophysical Research: Oceans**, v. 99, n. C5, p. 10143–10162, 15 maio 1994.

EVENSEN, G. **Data Assimilation**. Second ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

EVENSEN, G. Analysis of iterative ensemble smoothers for solving inverse problems. **Computational Geosciences 2018 22:3**, v. 22, n. 3, p. 885–908, 3 mar. 2018.

FOROUD, T.; SEIFI, A.; AMINSHAHIDI, B. Assisted history matching using artificial neural network based global optimization method – Applications to Brugge field and a fractured Iranian reservoir. **Journal of Petroleum Science and Engineering**, v. 123, p. 46–61, 1 nov. 2014.

GERRITSEN, M. G.; DURLOFSKY, L. J. MODELING FLUID FLOW IN OIL RESERVOIRS. **Annual Review of Fluid Mechanics**, v. 37, n. 1, p. 211–238, 1 jan. 2005.

GOMEZ, S.; GOSSELIN, O.; BARKER, J. W. Gradient-Based History Matching With a Global Optimization Method. **SPE Journal**, v. 6, n. 02, p. 200–208, 1 jun. 2001.

GUERILLOT, D. R.; BRUYELLE, J. History matching methodology using an optimal neural network proxy and a global optimization method. **3rd EAGE Integrated Reservoir Modelling Conference**, 5 dez. 2016.

GUÉRILLOT, D. R.; BRUYELLE, J. . **Uncertainty Assessment in Production Forecast with an Optimal Artificial Neural Network**. Day 4 Thu, March 09, 2017. **Anais**...SPE, 6 mar. 2017.

GUPTA, A. D. et al. High-resolution monotonic schemes for reservoir fluid flow simulation. v. 15:3, 1991.

HEIMSUND, B.-O. **Mathematical and Numerical Methods for Reservoir Fluid Flow Simulation**. [s.l: s.n.].

HOUTEKAMER, P. L.; MITCHELL, H. L. Data Assimilation Using an Ensemble Kalman Filter Technique. **Monthly Weather Review**, v. 126, n. 3, p. 796–811, mar. 1998.

JABER, A. K.; AL-JAWAD, S. N.; ALHURAISHAWY, A. K. A review of proxy modeling applications in numerical reservoir simulation. **Arabian Journal of Geosciences**, v. 12, n. 22, p. 1–16, 1 nov. 2019.

JUNG, S. et al. Ensemble-Based Data Assimilation in Reservoir Characterization: A Review. **Energies 2018, Vol. 11, Page 445**, v. 11, n. 2, p. 445, 17 fev. 2018.

KABANIKHIN, S. et al. Definitions and examples of inverse and ill-posed problems. **Journal of Inverse and Ill-posed Problems - J INVERSE ILL-POSED PROBL**, v. 16, p. 317–357, out. 2008.

KALETA, M. P. et al. Model-reduced gradient-based history matching. **Computational Geosciences 2010 15:1**, v. 15, n. 1, p. 135–153, 5 ago. 2010.

KALMAN, R. E. A New Approach to Linear Filtering and Prediction Problems. **Journal of Basic Engineering**, v. 82, n. 1, p. 35–45, 1 mar. 1960.

LI, G.; REYNOLDS, A. C. Iterative Ensemble Kalman Filters for Data Assimilation. **SPE Journal**, v. 14, n. 03, p. 496–505, 27 set. 2009.

LIANG, X. et al. Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. **Quarterly Journal of the Royal Meteorological Society**, v. 138, n. 662, p. 263–273, jan. 2012.

LIU, M.; GRANA, D. Time-lapse seismic history matching with an iterative ensemble smoother and deep convolutional autoencoder. **GEOPHYSICS**, v. 85, n. 1, p. M15–M31, 1 jan. 2020.

LORENTZEN, R. J. et al. Underbalanced and Low-head Drilling Operations: Real Time Interpretation of Measured Data and Operational Support. **All Days**, 30 set. 2001.

MA, X.; BI, L. A robust adaptive iterative ensemble smoother scheme for practical history matching applications. **Computational Geosciences 2019 23:3**, v. 23, n. 3, p. 415–442, 27 mar. 2019.

MASSONNAT, G. Random walk for simulation of geobodies: A new process-like methodology for reservoir modelling. **4th EAGE Conference on Petroleum Geostatistics**, v. 2019, n. 1, p. 1–5, 2 set. 2019.

MAUCEC, M. et al. Ensemble-Based Assisted History Matching With Rigorous Uncertainty Quantification Applied to a Naturally Fractured Carbonate Reservoir. **Proceedings - SPE Annual Technical Conference and Exhibition**, v. 2016-January, 26 set. 2016.

NÆVDAL, G.; MANNSETH, T.; VEFRING, E. H. Near-Well Reservoir Monitoring Through Ensemble Kalman Filter. **All Days**, 13 abr. 2002.

OLIVER, D. S.; CHEN, Y. Recent progress on reservoir history matching: a review. **Computational Geosciences 2010 15:1**, v. 15, n. 1, p. 185–221, 20 jul. 2010.

PEACEMAN, D. W. **Fundamentals of Numerical Reservoir Simulation**. [s.l.] Elsevier Scientific Publishing Company, 1977. v. 6

PYRCZ, MICHAEL. J.; DEUTSCH, CLAYTON. V. **Geostatistical Reservoir Modeling**. Second Edition ed. New York, NY: [s.n.].

RANAZZI, P. H.; LUO, X.; SAMPAIO, M. A. Improving pseudo-optimal Kalman-gain localization using the random shuffle method. **Journal of Petroleum Science and Engineering**, v. 215, p. 110589, 1 ago. 2022.

RANAZZI, P. H.; SAMPAIO, M. A. Influence of the Kalman gain localization in adaptive ensemble smoother history matching. **Journal of Petroleum Science and Engineering**, v. 179, p. 244–256, 1 ago. 2019a.

RANAZZI, P. H.; SAMPAIO, M. A. Ensemble size investigation in adaptive ES-MDA reservoir history matching. **Journal of the Brazilian Society of Mechanical Sciences and Engineering**, v. 41, n. 10, p. 1–11, 1 out. 2019b.

RANAZZI, P.; SAMPAIO, M. A. **Análise da influência do tamanho do conjunto na aplicação do Conjunto Suavizado no processo de Ajuste de Histórico**. jun. 2018.

RODRIGUES, J. R. P. Calculating derivatives for history matching in reservoir simulators. **SPE Reservoir Simulation Symposium, Proceedings**, p. 527–535, 2005.

ROSA, A. J.; CARVALHO, R. DE S.; XAVIER, J. A. D. **Engenharia de reservatórios de petróleo**. Rio de Janeiro: Interciência, 2006.

ROSA, D. R.; SCHIOZER, D. J.; DAVOLIO, A. Impact of model and data resolutions in 4D seismic data assimilation applied to an offshore reservoir in Brazil. **Journal of Petroleum Science and Engineering**, v. 216, p. 110830, 1 set. 2022.

RWECHUNGURA, R.; DADASHPOUR, M.; KLEPPE, J. Advanced History Matching Techniques Reviewed. **SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings**, v. 3, p. 1729–1747, 25 set. 2011.

SCHLUMBERGER. **Eclipse: Reference Manual | PDF | Petroleum Reservoir**. Disponível em: <https://pt.scribd.com/document/546274215/Eclipse-Reference-Manual>. Acesso em: 7 out. 2022.

SILVA, T. M. D. et al. A new procedure for generating data covariance inflation factors for ensemble smoother with multiple data assimilation. **Computers & Geosciences**, v. 150, p. 104722, 1 maio 2021.

SKJERVHEIM, J.-A. .-A. et al. **An Ensemble Smoother for assisted History Matching**. All Days. **Anais**...SPE, 21 fev. 2011. Disponível em: <https://onepetro.org/spersc/proceedings/11RSS/All-11RSS/The%20Woodlands,%20Texas,%20USA/151153>

SLOTTE, P. A.; SMØRGRAV, E. **Response Surface Methodology Approach for History Matching and Uncertainty Assessment of Reservoir Simulation Models**. Europec/EAGE Conference and Exhibition. **Anais**...SPE, 9 jun. 2008.

SOARES, R. V.; MASCHIO, C.; SCHIOZER, D. J. Applying a localization technique to Kalman Gain and assessing the influence on the variability of models in history matching. **Journal of Petroleum Science and Engineering**, v. 169, p. 110–125, 1 out. 2018.

STENGEL, R. F. **Optimal Control and Estimation**. Mineola, New York: Dover Publications, 1994.

TADJER, A.; BRATVOLD, R. B.; HANEA, R. G. Efficient dimensionality reduction methods in reservoir history matching. **Energies**, v. 14, n. 11, 1 jun. 2021.

TIERNEY, L. Markov Chains for Exploring Posterior Distributions. **The Annals of Statistics**, v. 22, n. 4, 1 dez. 1994.

VAN LEEUWEN, P. J.; EVENSEN, G. Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. **Monthly Weather Review**, v. 124, n. 12, p. 2898–2913, 1996.

VOLKOV, O. et al. Gradient-based Pareto optimal history matching for noisy data of multiple types. **Computational Geosciences 2018 22:6**, v. 22, n. 6, p. 1465–1485, 17 ago. 2018.

WANTAWIN, M. et al. An Iterative Response-Surface Methodology by Use of High-Degree-Polynomial Proxy Models for Integrated History Matching and Probabilistic

Forecasting Applied to Shale-Gas Reservoirs. **SPE Journal**, v. 22, n. 06, p. 2012–2031, 14 dez. 2017.

WILLIAMS, M. A.; KEATING, J. F.; BARGHOUTY, M. F. The Stratigraphic Method: A Structured Approach to History-Matching Complex Simulation Models. **SPE Reservoir Engineering (Society of Petroleum Engineers)**, v. 1, n. 2, p. 169–176, 1998.

XUE, L. et al. Ensemble-based optimization of hydraulically fractured horizontal well placement in shale gas reservoir through Hough transform parameterization. **Petroleum Science**, v. 18, n. 3, p. 839–851, 1 jun. 2021.

YETEN, B. et al. **A Comparison Study on Experimental Design and Response Surface Methodologies**. SPE Reservoir Simulation Symposium. **Anais**...SPE, 31 jan. 2005.

ZACHARIASSEN, E. et al. **Integrated Workflow for Model Update Using Geophysical Reservoir Monitoring Data**. 4 abr. 2011.

ZHANG, Y. et al. Feature-based ensemble history matching in a fractured carbonate reservoir using time-lapse deep electromagnetic tomography. **Journal of Petroleum Science and Engineering**, v. 208, p. 109259, 1 jan. 2022.

ZUBAREV, D. I. **Pros and Cons of Applying Proxy-Models as a Substitute for Full Reservoir Simulations**. All Days. **Anais**...SPE, 4 out. 2009.

ANNEX A – SYNHTESIS ARTICLE (AS)

Universidade de São Paulo

Engenharia de Petróleo – Escola Politécnica

Número: 10335079 Data: 22/11/2022

History Matching on Real Case Study using Ensemble Smoother with Multiple Data Assimilation

Eduardo de Rossi Borin

Orientador: Prof. Dr. Leonardo Azevedo Guerra Raposo Pereira

Supervisor: Prof. Dr. Marcio Augusto Sampaio Pinto

Artigo Sumário referente à disciplina PMI3349 – Trabalho de Conclusão de Curso II
Este artigo foi preparado como requisito para completar o curso de Engenharia de Petróleo na Escola Politécnica da USP.

Abstract

History matching in reservoir engineering is one of the most important steps of the geo-modelling workflow during the production phase of any oil field. This technique allows to have a good knowledge about how the field is producing and also have a forecast model. The biggest challenge is to be able to generate models with the reservoir's level of complexity that represents reality in the best way possible. The objective of this work is to get a match on production vector properties in a real case study of a turbiditic reservoir. For that matter, it was used an ensemble-based method to assimilate the dynamic data, the Ensemble Smoother with Multiple Data Assimilation method, in order to find the best combination of uncertain parameters used as input to the reservoir simulator. The geological model used is from a real dataset and the elementary sand channels are generated with ULIKE™: a rule-based modeling process based on the random-walk principle. The data assimilation processes produced ensembles of history matched models that honor both dynamic data and the geological model. Furthermore, it was also used an history matching software based on experimental designs and proxy-models using interpolated surfaces/polynomials and kriging to inspect the impacts from the uncertain parameters, mainly the productivity index multipliers. Results show that using interpolation to estimate the reservoir response is a quicker and demands a lower computational cost but requires the confirmation with the numerical response from a reservoir simulator.

1. Introduction

Along with the large variety of studies that are made to fully understand the area in which a production activity will take place, reservoir studies and, more specific, fluid flow simulations play one of the most important roles in this whole hydrocarbon recovery process. To build a reservoir simulation model, it is necessary to have data about rock and fluid properties (petrophysical attributes), in order to characterize the fluid flow and the entire physical process behind it. Besides, it is also of utmost importance always bear in mind that the data available for characterization is not exact and entirely correct, as it is consisted of inaccuracy, inconsistency and uncertainty. As indirect methods of evaluation, such as seismic and well logs, are mainly used to acquire information about the subsurface formations' physical properties, reservoir models are built inside an uncertainty parameters domain, meaning that any prediction made from these models are also considered uncertain. That is the reason why history matching processes are being largely utilized in the industry in the last decades: improve the reliability of reservoir predictions using the models generated by available petrophysical data and statistical methods trying to incorporate field observation data in reservoir simulations models, allowing a better description of the uncertainty both in simulation predictions and reservoir parameters.

The main objective of this work is to obtain a model that can predict the future behavior of the reservoir in order to optimize the production process and oil recovery. For this, given the observation data from some of the production vector properties (i.e., bottom hole pressure, water cut and oil production rate) and the simulation results from the model previous created with the uncertain parameters (i.e., porosity, permeability, net-to-gross ratio, transmissibility multipliers between faulted regions and well productivity index's multipliers) as inputs, a history matching process will be made using mainly the ensemble smoother with multiple data assimilation (ES-MDA: EMERICK; REYNOLDS, 2013) to find the best combination of these uncertain parameters defined beforehand that gives the best match from the observed data.

1.1. Bibliographic Review

The petroleum industry has adopted ensemble-based data assimilation techniques since they produce several history-matched models and can also quantify the level of uncertainty in reservoir simulation results (AANONSEN et al., 2009;

EVENSEN, 2009). This group of data assimilation techniques includes ES-MDA, proposed by Emerick and Reynolds (2013). This method has been demonstrated to outperform existing ensemble-based approaches in both synthetic and real-world scenarios, with improved data matching and reduced computational costs (EMERICK, 2016).

The same data set is iteratively assimilated with an inflated measurement error covariance matrix in ES-MDA. For linear systems with Gaussian prior and Gaussian noise in the measurements, Emerick & Reynolds (2013) demonstrated that the inflation factors $\alpha's$ must satisfy the following requirement to achieve the following equivalence between single and multiple data assimilations:

$$\sum_{\ell=1}^{N_a} \frac{1}{\alpha_\ell} = 1, \tag{1}$$

where N_a is the number of data assimilation defined and α_ℓ is the inflation coefficient. Therefore, for every ES-MDA problem, it is needed to choose the number of assimilations that will be done and the inflations coefficients before each assimilation step.

From Eq. (1), it is noticeable that a simple choice for α is $\alpha_\ell = N_a$, for all ℓ. Nevertheless, it makes intuitive sense that selecting alpha in descending order can enhance the method's effectiveness (EMERICK; REYNOLDS, 2013). In this instance, on initial steps, data is assimilated with a high α value, which corresponds to reducing the amplitude of the initial updates, and then gradually decrease α as the iterations pass by. Nonetheless, Emerick (2012) shows that changing the inflation coefficients in a decreasing order result in only small improvements compared to using them constant and equal to the number of data assimilations.

Rosa et al. (2022) also employed the ES-MDA technique to assimilate 4D seismic data and production forecast in a deep-water oil field, located in offshore Southwest Brazil, in order to compare the impact of using grids with different resolutions, exploring options to incorporate 4D seismic maps in data assimilation and production forecast. The authors proposed the use of two types of grids: a coarser grid or a refined grid. Results shows that both grids can be used depending on the project's objectives. Since it can accurately reflect the immediate future with considerably less expensive models, the coarser grid is a useful choice for short-term investigations. Due to its safer risk analyses and improved representation of model uncertainty, the refined grid may be appropriate for long-term decisions.

The development of the ensemble-based history matching (EnHM) software in the TotalEnergies' geoscience platform known as Sismage-CIG is described in Abadpour et al. (2018). The ES-MDA approach with distance-based localization forms the basis of the software. A mature large carbonate oil field, a deep-offshore turbidite field, and a carbonate gas field were the three complex genuine fields that were used to demonstrate the performance of the methods.

2. Methodology

The methodology presented in this work aims to apply ES-MDA (EMERICK; REYNOLDS, 2013) to a typical history matching problem of a reservoir model using real data.

2.1. History matching using ES-MDA

Following the definition of the uncertain parameters and their input on the model, it is necessary to sample them given the type of pdf predefined as one of the inputs. Every set of the different combinations of parameters are called ensembles, and they are essential on the history matching process with ensembled-method data assimilation. Each member j of the ensemble with size N_e are containing the uncertain parameters of the model are defined as:

$$m_j^u = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{N_m} \end{bmatrix}, \tag{2}$$

where m^u is the vector containing all N_m uncertain parameters that will be updated, such as permeability and productivity multipliers indexes data. The superscript u stands for *uncertain*.

After the sampling of the parameters for the N_e ensembles, it is also generated N_e models which are taken to the forward operator, from time zero to the end of the historical period, in order to compute the vector of predicted data:

$$d_j^{\ell,f} = F\left(m_j^{\ell-1,u}\right), \quad for\ j = 1, 2, \cdots, N_e, \tag{3}$$

where $F(\cdot)$ is the forward operator (i.e., the non-linear model used to obtain the forecast observations); $d_j^{\ell,f}$ is the *forecast* (f) model response with size N_d, the total number of measurements in the historical period, resulted from the combination $m_j^{\ell-1,u}$ of uncertain parameters.

For this work, the INTERSECT™ Reservoir Simulator was employed to compute the non-linear forward problem, hence simulating the dynamic behavior of the reservoir response to the different combination of uncertainty parameters previously defined. INTERSECT™ is a high-resolution reservoir simulator that, when coupled with a high-performance computer, enables this work to fully parallelize the execution of dynamic simulations for every ensemble member.

ES-MDA requires the N_d measurements to be perturbed, i.e., apply Gaussian noise in order to be treated as random variables. The mean and standard deviation values are calculated from previous research of the study area during screening methods and single parameter research:

$$d_j^{\ell,f} = F\left(m_j^{\ell-1,u}\right), \quad for\ j = 1, 2, \cdots, N_e, \tag{4}$$

where $d^{\ell}_{obs_j,uc}$ is the perturbed vector of observations; z_d is sampled with $z_d \sim \mathcal{N}\left(0, I_{N_d}\right)$; C_D is the $N_d \times N_d$ covariance matrix of observed data measurement errors:

$$C_D = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{N_d}^2 \end{bmatrix}. \tag{5}$$

After that, the simulations results are compared with the observation data. As every process is based on statistical analysis and from observed data, there will certainly be uncertainty on the model, giving mismatches from the simulation results and the expectation (observations) data. Hence, these discrepancies will be numerically used inside the data assimilation based on correlation computation, and, for each iteration ℓ, the previous sampled parameters will be updated to a new set of uncertain parameters, in order to reduce the variance of the discrepancy, following:

$$m_j^{\ell,u} = m_j^{\ell-1,u} + \tilde{C}_{MD}^{\ell-1,f}\left(\tilde{C}_{DD}^{\ell-1,f} + \alpha_\ell C_D\right)^{-1}\left(d^{\ell}_{obs_j,uc} - d_j^{\ell,f}\right), \quad for\ j = 1, 2, \cdots, N_e, \tag{6}$$

where, $\tilde{C}_{MD}^{\ell-1,f}$ represents the cross-covariance matrix between the prior vector of model parameters, $m_j^{\ell-1,u}$, and the vector of predicted data, $d_j^{\ell,f}$; $\tilde{C}_{DD}^{\ell-1,f}$ is the $N_d \times N_d$ auto-covariance matrix of predicted data. As before, the tilde (~) notation indicates that these covariances matrices are estimated around the ensemble mean:

$$\tilde{C}_{MD}^{\ell-1,f} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left(m_j^{\ell-1,u} - \bar{m}^{\ell-1}\right)\left(d_j^{\ell,f} - \bar{d}^{\ell,f}\right)^T, \tag{7}$$

$$\tilde{C}_{DD}^{\ell-1,f} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left(d_j^{\ell,f} - \bar{d}^{\ell,f}\right)\left(d_j^{\ell,f} - \bar{d}^{\ell,f}\right)^T, \tag{8}$$

As mentioned, the ES-MDA processes will go on until the pre-defined number of assimilations (N_a) is reached. It is important to mention that the assimilation process is made after the end of the forward operator and that the observation data are randomly sampled according to the data error covariance. Furthermore, for the update part, ES-MDA uses an inflation factor α_ℓ for each assimilation step, and their choice must respect the condition stated in Eq. (1) for a correct *a posteriori* parameter's distribution. In this work, for all study cases using ES-MDA process, the inflation factor used were constant and equals to the number of iterations, i.e., $\alpha_\ell = N_a$.

Figure 1 illustrates the workflow containing the steps of applying ES-MDA.

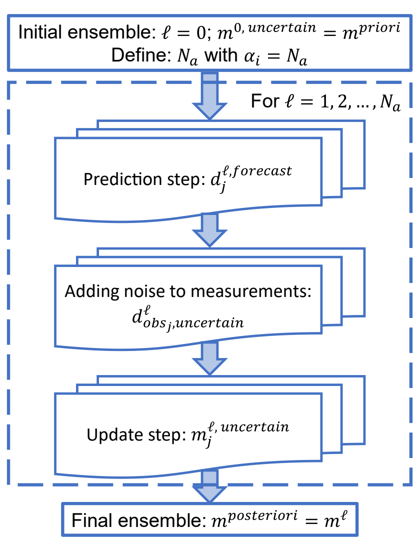


Figure 1 - Workflow for history matching with ES-MDA on Alpha-Field

Source: Adapted from Ranazzi & Sampaio (2018)

2.2. Objective function

The Objective Function (OF) for each data assimilation at iteration ℓ used in this work is through the quadratic deviation between simulated model forecast and the observation data, normalized by the inverse of the covariance matrix of observed data measurement errors, C_D and with the total number of observations N_d, as shown in Eq. (9)

$$OF^{\ell} = \sqrt{\frac{\sum_{j=1}^{N_e} \left(d_j^{\ell,f} - d_{obs_j}^{\ell}\right) C_D^{-2} \left(d_j^{\ell,f} - d_{obs_j}^{\ell}\right)^T}{N_D}} \quad (9)$$

3. Case Study

The study area of the work focuses on the analysis of a real case field with a reservoir consisting of turbiditic sediments composed of a loosely consolidated sand.

3.1. Synthesis of the case study and initial model

The studied field is called Alpha-Field, a multi-story erosive constructive channel complex. The reservoir is placed at the level of a turtle-back salt anticlinal created after turbidite reservoir deposition. This structure causes a large number of Keystone faults, which are normal faults that run perpendicular to the system's elongation. The studied area is divided according to the direction of the channels from East to West into 3 zones - Alpha-East, Alpha-Central and Alpha-West. This work focus on two areas (Figure 2):

- Alpha-Central, which contains 2 wells: one producer (PROD6) and one injector (INJ8);
- Alpha-East, which also contains 2 wells: one producer (PROD10) and one injector (INJ12).

Within the system, structural and stratigraphic heterogeneities will exist. There is little transmissibility between regions because the degree of depletion varies by region. For the modelling of the elementary channels, ULIKE™ − a rule-based modeling process based on the random-walk principle – was used (MASSONNAT, 2019).

The uncertain parameters need to be chosen in order to assess their impact on the fluid flow and in the data assimilations processes. For the Alpha-system, lots of studies have already been carried out on the last couple of years in order to have a better geological understanding of the area and how to improve more and more this in a numerical representation. Hence, the uncertainty parameters considered for each section of the Alpha-field for this work is:

- Alpha-Central:
  - Transmissibility multipliers between: faulted areas; distal levee and channel fairways; channel and proximal levee;
  - Elementary channel’s porosity, permeability and net-to-gross ratio;
  - Flux multiplier on vertical connections between Maximum Flooding Surfaces;
  - Wells productivity index multipliers

- Alpha-East
  - Transmissibility multipliers between: distal levee and channel fairways; channel and proximal levee;
  - Elementary channel’s porosity, permeability and net-to-gross ratio;
  - Flux multiplier on vertical connections
  - Wells productivity index multipliers
  - Analytical aquifer’s initial volume

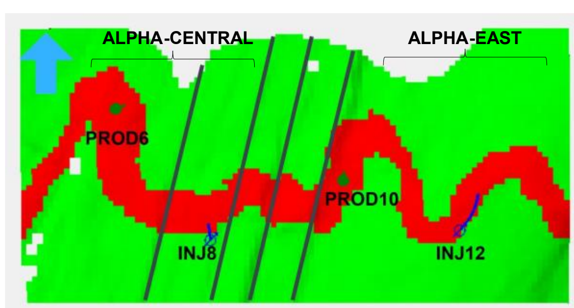


Figure 2 - Top view of the erosive channel complex containing an explanatory diagram of the studied field with the highlighted wells for future analysis

4. Results

4.1. Alpha-East: PROD10 − INJ12

ES-MDA is used to update the uncertain parameters in order to find a model that behaved similarly to the real reservoir. It is possible to present a few indicators that led to the correct behavior of the interest response. For this analysis, the number of ensembles used was $N_e = 100$ and with 5 iterations ($N_a = 5$). The production vectors properties to be matched are:

- Producer well (PROD10)
  - Bottom Hole Pressure (BHP)
  - Oil Production Rate
  - Water cut
- Injector well (INJ12)
  - Bottom Hole Pressure (BHP)

The results obtained are shown in Appendix 1. These figures show the response adjustments of the models that minimize the variance from of the ensembles from each iteration.

Analyzing the BHP of PROD10 and INJ12, it is possible to observe that a match was obtained for both flowing and static pressure, mainly after the long pressure build-up with PROD10's shut down (roughly day 1700). Also, the final ensemble spread (red curves) of PROD10 oil production rate is inside the uncertain region of the historical data. The water breakthrough is a little late compared with the observations, but after the well closure on day 1092, the water cut is better matched, obtaining values in the lower bound of the observation error interval. Due to the high complexity of the reservoir's geology, it is possible to have elementary channels with different properties then the ones defined from ULIKE™ inside the uncertain parameter list, which are permeability, porosity, and net-to-gross ratio.

Figure 3 shows the evolution of the mean OF over the five data assimilation iterations. It depicts an initial high OF value with a big uncertainty, but decreasing from iteration to iteration, with its variance following the same pattern.

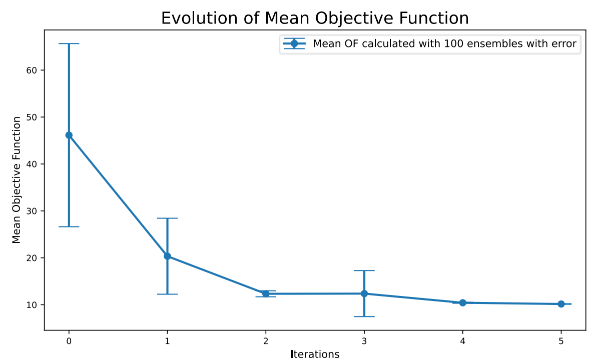


Figure 3 - OF statistics of ES-MDA run on Alpha-East

4.2. Alpha-Central: PROD6 – INJ8

Multiple faults have a significant impact on this section, reducing transmissibility within the core region but not blocking the flow. The aquifer and the elementary channels generated by ULIKE™ are used in the production mechanism. For this analysis, the number of ensembles used was $N_e = 100$ and with 5 iterations ($N_a = 5$). The production vectors properties to be matched are:

- Producer well (PROD6)
  - Bottom Hole Pressure (BHP)
  - Oil Production Rate
  - Water cut
- Injector well (INJ8)
  - Bottom Hole Pressure (BHP)

Appendix 2 shows the results obtained from the history matching and data assimilation process on the Alpha-Central section.

The BHP of PROD6 shows that the simulated values are very close to the observation, mainly for the flowing part. Regarding the static pressure (all three build-ups), there is gap between simulated and observed data, which could indicate a lack of connected volume between the injector INJ8 who helps to maintain the pressure for the producer. About INJ8 BHP, ES-MDA manages to get a good match on both fluid and static pressure. For instance, as observed mainly in the three fall-offs between days 1000 and 2000, the short-term pressure drop is well matched, meaning that the behavior in the well's vicinity is correctly guessed with the parameters of choice: in this case, the productivity index multiplier on the three completion layers. In the final days of the simulation, mainly for INJ8, there is relatively large mismatch on the pressure drop. This could also be an indicative that there is still some volume connectivity missing inside the model which could be improved in further analysis

Both oil production rate and water cut are well matched during the whole simulated timesteps. Nonetheless, we observe two more families of simulated ensembles (red curves) in the final ES-MDA iteration. This means that, if continued, the assimilation processes could be able to find not only a more significant representation of the water cut then it already is, but also for the others three production vector properties analyzed. The water cut of a production well is not only function of the injection rate of the injector well helping to maintain a good pressure value for the oil recovery from the producer, but also on the faults' transmissibility, which are all considered inside our uncertain parameter lists and the geological model applied for this ES-MDA run.

Figure 4 shows the evolution of the mean OF over the five data assimilation iterations. As previously mentioned, in the final iterations, ES-MDA generated some new families of ensembles that do not follow the same mean behavior, hence the value increase of OF from iteration 3 to 4. These new ensemble families generated from iteration 4 to 5 are not so dispersed around the mean, meaning that the variance estimated from the ensembles are reduced, as shown in the reduction of the error bar from the picture. Nevertheless, for the history matching results and analysis of the uncertain parameters, the data assimilation process should stop at iteration 3, as it reached a possible minimum of the OF.

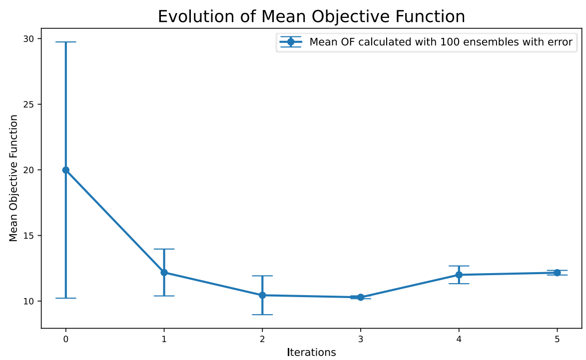


Figure 4 - OF statistics of ES-MDA run on Alpha-Central

4.2.1. Analysis of Alpha-Central using EST's manual history matching with interpolated responses

In this section of the thesis, the software EST – an internal software from TotalEnergies used for history matching – was used to also try to obtain a good match on the Alpha-Central. The motivation of this comparison was the difficulty in finding a good range for the PROD6 and INJ8 productivity index for the initial ensemble of ES-MDA, which often generated spurious correlation values and parameters collapses in earlier iterations, not obtaining a good first spread over the observation data and, thus, failing to get a match. The goal of EST is to find a good response of the simulation data that fits the chosen parameters as well as the desired model responses. This model could be a kriging surface or a polynomial regression. As opposed to ES-MDA, which is a method without limitation on the numbers of uncertain parameters, the maximum handled by EST are 11.

In Appendix 3 the simulations from which EST finds the polynomial regression parameters are represented in blue, with a total of 151 curves. From that, it is necessary to manually find the best combination of parameters that approximated most the observation data. That kriged response is shown in the red line.

With the combination of the parameters that generated the best kriged response, it is also necessary to confirm with the fluid flow simulator the real effect of those set of parameters, as there is a high uncertainty related to these interpolation processes and also the high non-linearity of the physics related to the problem. Appendix 4 depicts the response from the combination of the uncertain parameters get from EST. We observe that in a general way, the simulation results follow the same trend of the observation data points, and the match is obtained for some production properties, like the PROD6 bottom hole pressure and oil production rate. However, is it on the INJ8 bottom hole pressure where we can see the biggest discrepancies from EST kriged response and the real simulated one, especially on the flowing pressure (days 1100 to 1600, approximately).

5. Conclusions

This final graduation work presented a history matching workflow application for a real scale reservoir model with complex geological models, using data assimilation methods, more specifically an ensemble-based method. Even while ensemble-based methods have been effectively used in many applications over the past decades, inverse problems with significant nonlinearity, such as reservoir history matching, continue to pose difficulties. The Gaussian assumption that underlies this class of approaches are primarily to blame for these issues.

This work analyzed the real scale reservoir, denominated Alpha Field, into compartments due to the faults located in the area, which divided the turbiditic reservoir into different areas and regions. As mentioned, reservoir modeling is considered to be a complicated and time-consuming activity during the research and development process. This fact is mainly due to the geological complexity of petroleum reservoirs and the necessity of modeling the best was possible the geological features that may impact the fluid flow simulator results and, thus, the exploration activity. That said, it was proved the efficiency of the rule-based modeling tool ULIKE™, used to characterize the elementary sand channels and its respective fairways where the fluid flow is taken place and respecting all wells constraints of the field. These results affect directly in the history matching process of the analyzed area, as all the properties observed and the results obtained from the reservoir simulator is linked with the geology and how it is represented in the model, for instance, the connectivity between wells and how this impacts the bottom hole pressure.

Using data assimilation based on ensemble-based methodology to perform the history matching of the study area, the ES-MDA algorithm proved to be functional for the study case, being capable of yielding ensembles of models that honor the geological model and observation data, generating sets of history-matched models of good quality, both qualitatively and quantitatively. From the results of the analysis of both reservoir compartments, the following conclusions are deduced:

- In order to be able to check all the uncertain parameters responses for every ensemble member, it was necessary a high number of simultaneous reservoir fluid flow simulator, obtaining, thus, an ensemble of the forward operator used to the data assimilation problem. In average, for each reservoir compartment analysis using 5 assimilations iterations and an ensemble size of 100, it was necessary a minimum of 600 reservoir simulations to obtain a good representativity from the model responses and then, observe the quality of the assimilations;
- For all ES-MDA procedures made during this work, it was noted how much the initial ensembles models, resulted from the reservoir responses with the uncertain parameters sampled from the *a priori* distributions, affects the assimilation and history matching results. Many initial ensembles that did not have a big spread and did not cover the observation data failed to converge to a good history-matched ensemble of models, despite the use of inflation coefficients factors chosen in the beginning of the process. Therefore, the *a priori* distributions and the randomly sampled uncertain parameters contribute to achieving good results, whereby their ranges, mean/standard deviation values should be carefully chosen and analyzed previously during screening studies and sensitivity analysis;
- From the total of uncertain parameters used, the ones that impacted the most during the reservoir fluid flow simulator were the wells productivity index multipliers. The productivity index of the well is one parameter that represents the behavior of the fluid flow near the well, and the multiplier on them tries to simulate the skin factor. Therefore, it can be concluded that these parameters directly affected the dynamic pressure values, but also it was confirmed a possible non-linearity regarding these parameters, in which very small changes generated ensembles with completely different responses. These responses affected some simulations where the reservoir simulator did not completely finish the run, and, in order to continue the data assimilation method, those failed simulations had to be re-run.

The ES-MDA approach's overall performance is encouraging, and it offers great perspectives for the use of this data assimilation method in a real scale reservoir with a considerate geological complexity. The easy implementation of this type of ensemble smoother is desirable, being able not only to be used for reservoir engineering applications, but also for other data assimilation processes. The major difficulties and deficiencies encountered are in relation to spurious correlations and large loss of variability of the model ensemble, which can result in a poor estimate of reservoir uncertainties causing a poor prediction of field life, even with a good historical fit. This fact can be mitigated by applying adaptive methodologies, where the damping of each iteration as well as the number of iterations are set automatically and using covariance localization methods to reduce the loss of variance due to possible sampling errors.

In addition to the data assimilation process using ES-MDA, it was used the internal software for history matching from TotalEnergies, EST, to analyze one of the compartments of the reservoir. Experimental designs models that use interpolations and proxy models in order to give an approximate response to inverse problems can drastically reduce the computational cost necessary to run some forward operator, such as the reservoir simulator. However, as these responses are considerate to be an estimate of the real problem, it is always necessary to analyze the results obtained and check if they are feasible, respecting the physics of the problem involved and if the uncertainty is still well represented. With the results obtained from the work using this method, some conclusion can be made:

- The dynamism of the software, giving instantaneous interpolated responses is its main attraction. Even with its limitations, mainly with the maximum numbers of uncertain parameters that can be considered, it is possible to have innumerous combinations, having the opportunity to find the best combination possible that best approximates the responses to the observation data. However, because it is a manual process of choosing the parameters values and visualizing the response in the interpolated model, this process can become extremely time-consuming, with the possibility in not finding a good match with the uncertain parameters and its ranges previously defined, being necessary to redo the process, either with other parameters or different ranges;
- Even though the kriging responses are instantaneous, it is only possible to analyze the response changing one parameter at a time. That said, and knowing that with petroleum reservoirs, many parameters are related with each other (e.g., permeability/porosity) this can become an issue if the production property vector in which we want to obtain a match is large and there is a significant number of uncertainty parameters that is known to be correlated or to have a possible correlation with each other;
- From the analysis of real response given by the reservoir simulator, it was observed some discrepancies between the kriged response given by the best combination found of the uncertain parameters. These differences are

greater on the dynamic pressures values from the injector's bottom hole pressure. This could indicate a possible high non-linear effect of the productivity index's multipliers, as the kriged responses work with interpolations of second degree, which might not be able to correctly describe this production property.

Due to the absence of the real reservoir response during the manual history matching process, EST multi solutions process could have been used. For that matter, this methodology is iterative after having checked with the real reservoir response, repeating the process until certain stop criteria arrive. The criteria are based on the cost function computation. Nevertheless, EST with screening options could have also been used during initial phases of reservoir history matching, in which the objective is not to get an history-matched model, but to determine the most influent parameters from a large number of uncertain parameters (roughly 100) choices and ranges from their impacts on the production property vector. Those most influent parameters could serve as inspiration and, possibly, inputs of the *a priori* distribution used afterwards in history matching of petroleum reservoirs using data assimilation methods, such as ES-MDA.

Acknowledgements

The author would like to thank TotalEnergies S.A. for funding the internship period and providing all the necessary data and software licenses for the realization of this work.

6. References

AANONSEN, S. I. et al. The Ensemble Kalman Filter in Reservoir Engineering—a Review. **SPE Journal**, v. 14, n. 03, p. 393–412, 27 set. 2009.

ABADPOUR, A. et al. Integrated Geo-Modeling and Ensemble History Matching of Complex Fractured Carbonate and Deep Offshore Turbidite Fields, Generation of Several Geologically Coherent Solutions Using Ensemble Methods. **Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2018, ADIPEC 2018**, 12 nov. 2018.

EMERICK, A. **History Matching and Uncertainty Characterization Using Ensemble-based Methods**. Doctor of Philosophy—Tulsa: University of Tulsa, 2012.

EMERICK, A. A. Analysis of the performance of ensemble-based assimilation of production and seismic data. **Journal of Petroleum Science and Engineering**, v. 139, p. 219–239, 1 mar. 2016.

EMERICK, A. A.; REYNOLDS, A. C. Ensemble smoother with multiple data assimilation. **Computers & Geosciences**, v. 55, p. 3–15, jun. 2013.

EVENSEN, G. The ensemble Kalman filter for combined state and parameter estimation. **IEEE Control Systems**, v. 29, n. 3, p. 83–104, jun. 2009.

MASSONNAT, G. Random walk for simulation of geobodies: A new process-like methodology for reservoir modelling. **4th EAGE Conference on Petroleum Geostatistics**, v. 2019, n. 1, p. 1–5, 2 set. 2019.

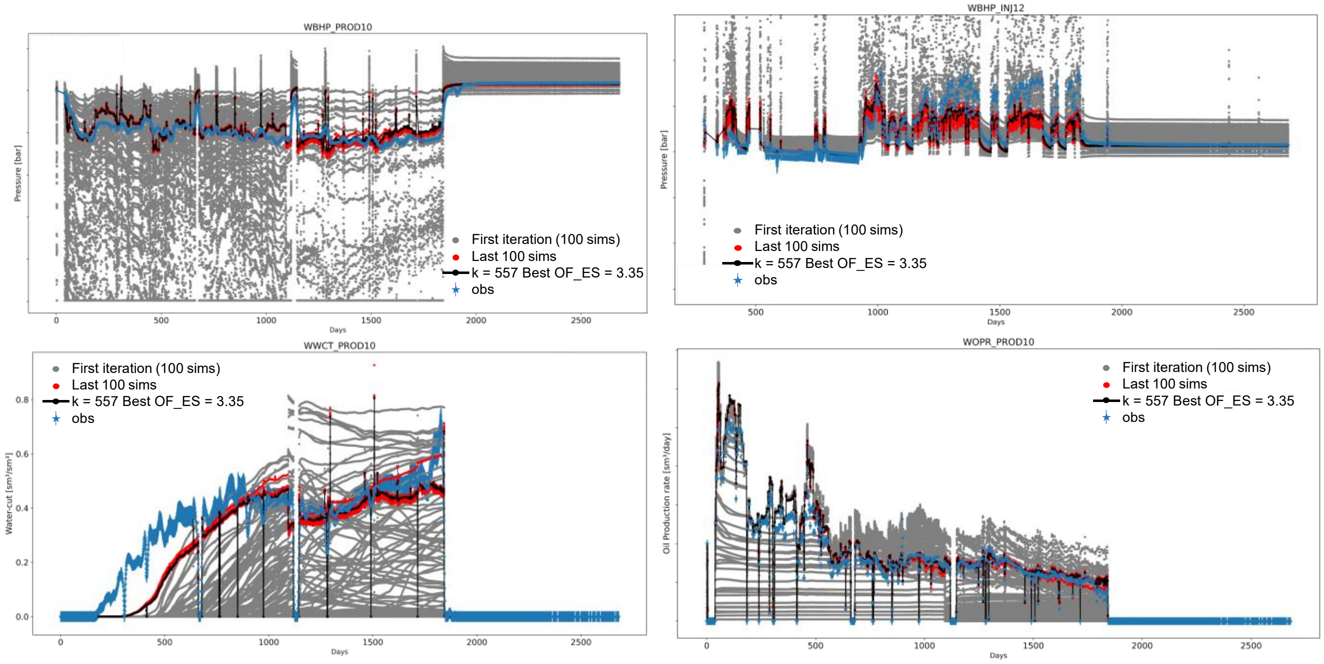
RANAZZI, P.; SAMPAIO, M. A. **Análise da influência do tamanho do conjunto na aplicação do Conjunto Suavizado no processo de Ajuste de Histórico**. jun. 2018.

ROSA, D. R.; SCHIOZER, D. J.; DAVOLIO, A. Impact of model and data resolutions in 4D seismic data assimilation applied to an offshore reservoir in Brazil. **Journal of Petroleum Science and Engineering**, v. 216, p. 110830, 1 set. 2022.

Appendix – Results obtained from ES-MDA and EST analysis on Alpha Field

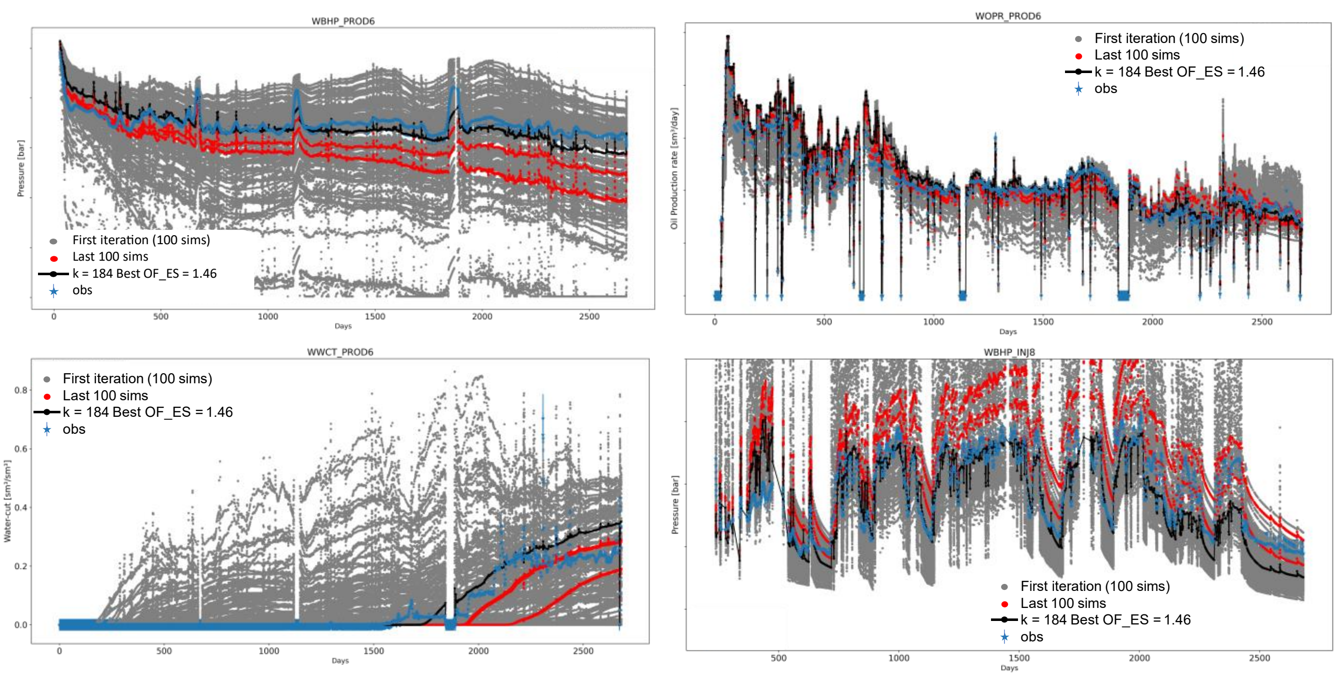
Appendix 1. Results obtained from ES-MDA on Alpha-East

From top to bottom – left to right: PROD10 bottom-hole pressure; PROD10 oil production rate; PROD10 oil water cut; INJ12 bottom-hole pressure. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.



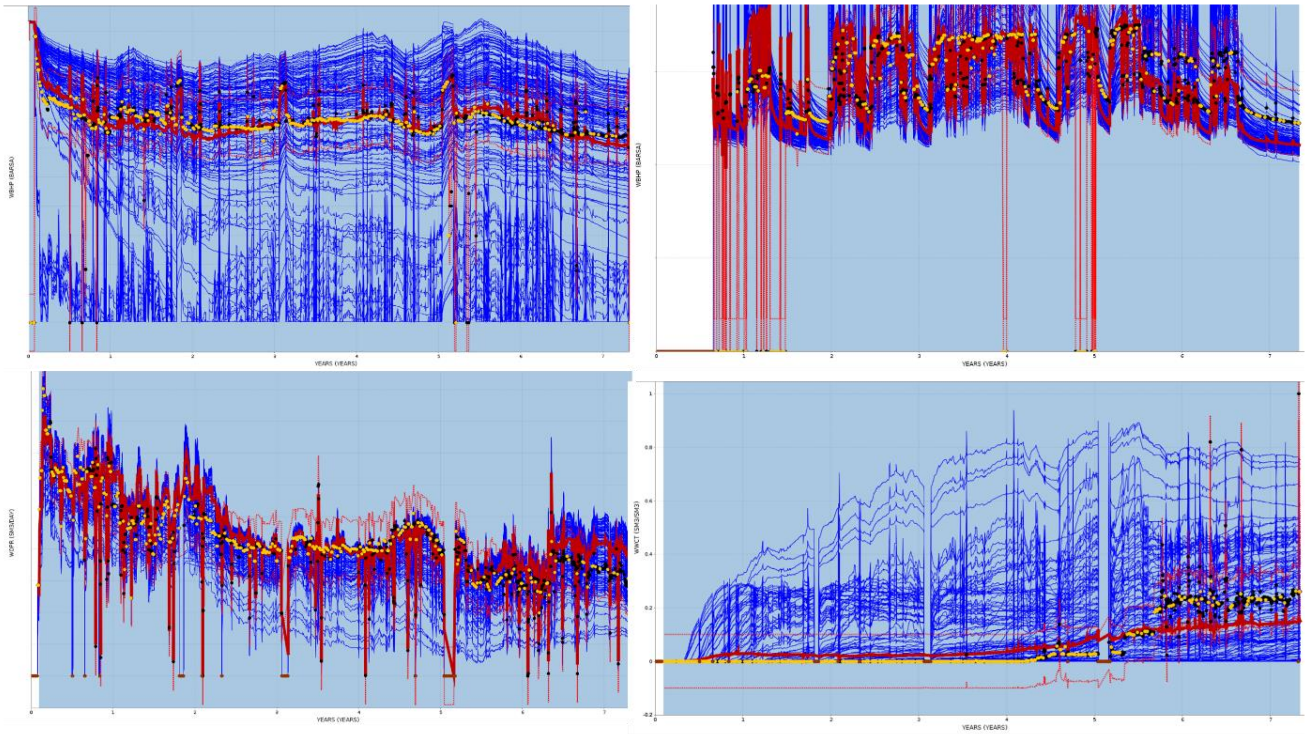
Appendix 2. Results obtained from ES-MDA on Alpha-Central

From top to bottom – left to right: PROD6 bottom-hole pressure; PROD6 oil production rate; PROD6 oil water cut; INJ8 bottom-hole pressure. The grey scale represents the initial ensemble spread. The red one represents the last ensemble. Historical data and their uncertain region (noise in measurements) are in blue.



Appendix 3. Best kriged response found using EST

From top to bottom – left to right: PROD6 bottom-hole pressure; INJ8 bottom-hole pressure; PROD6 oil production rate; PROD6 water cut. The blue curves are the 151 reservoir simulations. The red line is the best kriged response found.



Appendix 4. Fluid flow simulator response from the combination of uncertain parameters obtained from EST

From top to bottom – left to right: PROD6 bottom-hole pressure; INJ8 bottom-hole pressure; PROD6 oil production rate; PROD6 water cut. The continuous lines indicate the simulated production vector property. The black squares are the historical data.