UNIVERSIDADE DE SÃO PAULO

ESCOLA POLITÉCNICA

DEPARTAMENTO DE ENGENHARIA DE MINAS E DE PETRÓLEO

PAULO HENRIQUE RANAZZI

Iterative Ensemble Smoother with adaptive localization and deep learning parameterization for carbonate reservoirs

São Paulo

2023

PAULO HENRIQUE RANAZZI

Iterative Ensemble Smoother with adaptive localization and deep learning parameterization for carbonate reservoirs

Corrected Version

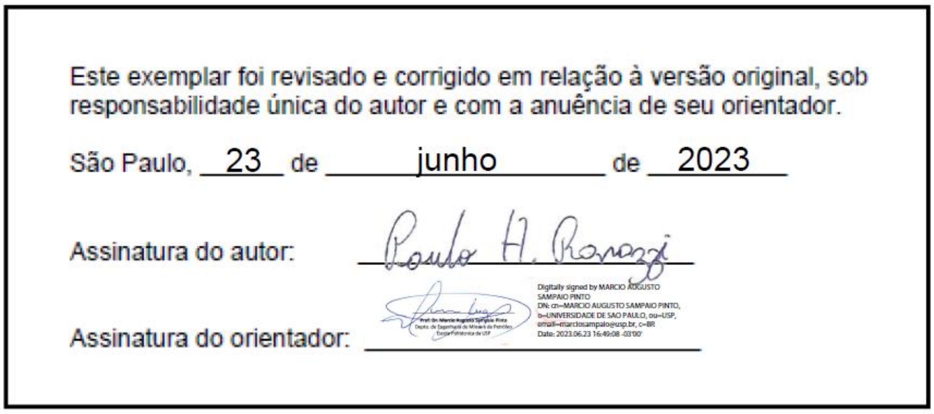
Thesis presented to the Graduate Program in Mineral Engineering at the Polytechnic School, University of Sao Paulo to obtain the degree of Doctor of Science.

Concentration Area: Mineral Engineering

Advisor: Marcio Augusto Sampaio Pinto, PhD

São Paulo

2023



Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 23 de junho de 2023

Assinatura do autor: Paulo H. Ranazzi

Prof. Dr. Marcio Augusto Sampaio Pinto
Depto. de Engenharia de Minas e de Petróleo
Escola Politécnica da USP

Digitally signed by MARCIO AUGUSTO SAMPAIO PINTO
DN: cn=MARCIO AUGUSTO SAMPAIO PINTO, o=UNIVERSIDADE DE SAO PAULO, ou=USP, email=marciosampaio@usp.br, c=BR
Date: 2023.06.23 16:49:08 -03'00'

Assinatura do orientador:

Catalogação-na-publicação

Ranazzi, Paulo Henrique
Iterative Ensemble Smoother with adaptive localization and deep learning parameterization for carbonate reservoirs / P. H. Ranazzi – versão corr. -- São Paulo, 2023.
145p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Minas e de Petróleo.

1.Data Assimilation 2.Iterative Ensemble Smoother 4.Adaptive Localization 5.Generative Adversarial Networks I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Minas e de Petróleo II.t.

ACKNOWLEDGEMENTS

I would like to express my heartfelt appreciation to my family and friends for their unwavering support and encouragement throughout this journey. Their belief in me and their constant presence have been a source of strength and motivation.

I am deeply grateful to my advisor, Prof. Marcio Sampaio, for his continuous support, guidance, and unwavering commitment to my academic and professional development. His expertise, invaluable insights, and willingness to share his knowledge have significantly influenced the outcome of this research.

I would like to extend my sincere thanks to the Energy Production Innovation Center (EPIC), FAPESP and Equinor for their generous financial support throughout the course of this research. Their investment in this project has enabled me to pursue in-depth investigations and achieve meaningful results.

I am grateful to the UNISIM group for providing the simulation models that were crucial for conducting this study. Their contribution has greatly enhanced the accuracy and reliability of the research outcomes. Additionally, I would like to acknowledge the Computer Modelling Group (CMG) for providing the essential licenses of the reservoir flow simulator.

Lastly, I want to express my gratitude to the Polytechnic School of the University of São Paulo (USP) and the Laboratory of Petroleum Reservoir Simulation and Management (LASG) for providing the necessary infrastructure and resources that facilitated the progress of this research. Their commitment to academic excellence and dedication to fostering a conducive research environment have been invaluable.

“Explore the world. Nearly everything is really interesting if you go into it deeply enough.”

(Richard P. Feynman)

ABSTRACT

In last decade, iterative ensemble smoother methods have become the standard approach to perform history matching of petroleum reservoirs. However, their applicability is limited when it comes to more complex reservoirs. These limitations can be summarized by two main assumptions during the ensemble-based methods formulation: the use of an ensemble with finite size to represent the distributions and the assumption of Gaussianity in both parameter and data uncertainties. Unfortunately, Gaussian assumption is particularly significant in carbonate reservoirs where many uncertain parameters are non-Gaussian, such as matrix permeability containing Super-k layers - very thin layers with extremely high permeability. The standard approaches in the petroleum literature for handling the limited ensemble size and non-Gaussianity are Kalman gain localization and parameterization, respectively. Kalman-gain localization involves tapering the Kalman-gain matrix to reduce the effect of spurious correlations during the uncertain parameter updates, while parameterization involves mapping non-Gaussian parameters to a Gaussian field before the update and then mapping them back to the original domain to forward the ensemble through the reservoir simulator. These limitations motivated the evaluation and development of specific methods to improve the data assimilation workflow in large-scale carbonate reservoirs. An adaptive localization method was developed, which showed an improvement in posterior ensemble preservation, consequently reducing the ensemble collapse effect. A novel parameterization method was also applied using a generative adversarial network with discriminator regularization, combined with a data augmentation technique, which demonstrated improved network performance even with a small ensemble. Finally, both methods were integrated into a two-step data assimilation workflow applied to the UNISIM-II-H benchmark, showing satisfactory results in terms of preserving geological features after the assimilation.

Keywords: Data Assimilation; Iterative Ensemble Smoother; Adaptive Localization; Parameterization; Generative Adversarial Networks.

RESUMO

Na última década, *Conjuntos Suavizados Iterativos* se tornaram o método padrão para a realização de ajuste de histórico de reservatórios de petróleo. Entretanto, sua aplicabilidade é limitada em reservatórios complexos. Estas limitações são relacionadas a duas das principais hipóteses durante a formulação destes métodos baseados em um conjunto de modelos: o uso de um conjunto de tamanho limitado para representar as distribuições e a hipótese de Gaussianidade em ambas as incertezas. Infelizmente, a hipótese de Gaussianidade é significativa em reservatórios carbonáticos onde muitos parâmetros incertos não possuem uma distribuição Gaussiana, como por exemplo a permeabilidade da matriz contendo camadas *Super-k* – camadas muito finas na direção vertical com uma permeabilidade extremamente elevada. Os métodos padrão para contornar os problemas relacionados com o tamanho do conjunto limitado e a não-Gaussianidade são Localização aplicada ao ganho de Kalman e parametrização, respectivamente. Localização aplicada ao ganho de Kalman refere-se à redução do impacto do ganho de Kalman para reduzir o efeito das correlações espúrias durante a atualização dos parâmetros incertos, enquanto a parametrização envolve um mapeamento do parâmetro não-Gaussiano para um domínio Gaussiano antes da etapa de atualização, então mapeando o parâmetro atualizado para seu domínio original para ser utilizado no simulador de reservatórios. Estas limitações motivaram a avaliação e desenvolvimento de métodos específicos para melhorar o processo de assimilação de dados em reservatórios carbonáticos de larga-escala. Um método de localização adaptativo foi desenvolvido, resultando em uma melhor preservação da variância do conjunto *a posteriori*, consequentemente reduzindo o efeito do colapso do conjunto. Um novo método de parametrização também foi avaliado, usando uma rede adversária generativa, em conjunto com uma técnica de aumento de dados, demonstrando melhorar no desempenho de treinamento mesmo com um conjunto de tamanho limitado. Finalmente, ambos os métodos foram integrados em uma assimilação de dados em duas etapas, aplicada ao benchmark UNISIM-II-H, resultando em resultados satisfatórios em termos da preservação das características geológicas após assimilação.

Palavras-chave: Assimilação de Dados; Conjunto Suavizado Iterativo; Localização Adaptativa; Parametrização, Redes Adversárias Generativas.

LIST OF FIGURES

LIST OF TABLES

SUMMARY

CHAPTER 1 - INTRODUCTION

One of the most challenging tasks in the petroleum industry is to estimate the reservoir performance during its entire production life. This estimated behavior is used to define many practical aspects of the petroleum industry, for example, production systems sizing, exploitation resources priority, risk assessment, etc. Numerical reservoir simulation is the most used technique to predict this behavior nowadays, where computational advances are making possible the evaluation of large-scale and complex reservoir systems. The reservoir simulator is composed by a set of input parameters that describes the reservoir, and a set of partial differential equations that represents the fluid flow through the reservoir porous media, which are discretized in both space and time to be solved numerically. Several properties are used to construct the reservoir simulation model, some examples are porosity, permeability, fluid-contact depths, fault properties, relative permeability curves, etc. In this scenario of reservoir modeling and forecasting, many properties used as input data to the building of the reservoir simulation model have uncertainties. These uncertainties may arise from indirect acquisition methods, approximations to reduce the computational cost to solve the porous flow equations (e.g., upscaling), and even approximations in the set of equations, neglecting some physics of the phenomena (COATS, 1969). These uncertainties and approximations are guaranteed to result in discrepancies between the reservoir simulator response and its measurements (which also contains measurement errors), making this reservoir simulation model unable to predict the future reservoir performance. History matching might be defined as the process which modifies the uncertain properties values to reduce the discrepancy between the reservoir simulator response and the measurements. In this context, it is important to mention that the overwhelming increase in computational power is making possible the development of robust methodologies to build complex reservoir simulation models and robust methodologies to perform the history matching in this kind of reservoir.

The history matching methods can be divided into two main categories: manual and automatic (or assisted) history matching. Manual history matching refers to the trial-and-error procedure, where a few global reservoir parameters (such as zone property multiplier) are adjusted manually by the reservoir engineer until the reservoir response achieves satisfactory results to obtain a single matched reservoir model (RWECHUNGURA; DADASHPOUR; KLEPPE, 2011). Due to the time-consuming process of manual history matching and because of the non-uniqueness of the solution (OLIVER; CHEN, 2011), manual history matching is generally impractical and inefficient. Automatic history matching represents the class of methods that modifies the parameters automatically in an optimization problem, making it possible to adjust several reservoir uncertain parameters simultaneously. In this context, the main methods to perform the history matching are described in Rwechungura, Dadashpour, and Kleppe (2011), and Oliver and Chen (2011).

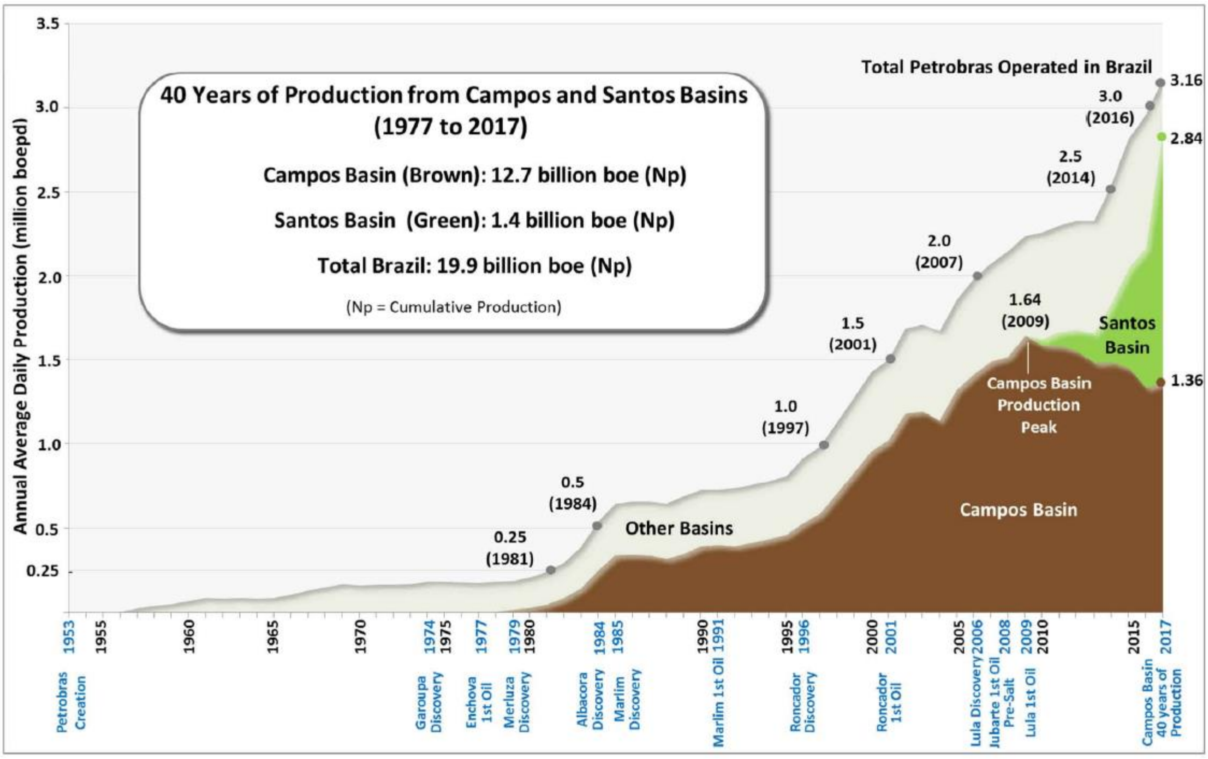
Automatic history matching might be considered as a type of data assimilation problem. Data assimilation is the science popularized by the meteorological and oceanography community which combines available data to improve the numerical model (AANONSEN et al., 2009). Generally, ensemble-based methods refer to the class of Monte Carlo methods which uses an ensemble to represent the model parameters and response uncertainties (AANONSEN et al., 2009; EVENSEN, 2009; OLIVER; CHEN, 2011). The advantages of the ensemble-based methods are its ease of application with relatively low computational cost, and the possibility to perform an uncertainty assessment since in this kind of method an ensemble of reservoir models is used to represent the model parameters and forecast uncertainties. Since its introduction with the Ensemble Kalman Filter – EnKF (EVENSEN, 1994), ensemble-based methods became the state-of-the-art approach in the petroleum industry. Among the ensemble-based methods, the class of Iterative Ensemble Smoothers (IES) has become popular in petroleum reservoir applications (EVENSEN, 2018). The main advantages of IES compared to other ensemble-based methods is its simultaneous assimilation scheme, instead the recursive EnKF workflow.

Although IES has been used extensively in large-scale realistic reservoir models(EMERICK, 2016; BRESLAVICH; SARKISOV; MAKAROVA, 2017; LORENTZEN et al., 2020), this kind of method still has some limitations. These limitations are mainly related with the limited ensemble size used to represent the uncertainties and the Gaussian assumptions during the formulation of the update equations. The former results in spurious correlations which induces a severe variance reduction of the posterior ensemble variance, (namely as ‘ensemble collapse' by the data assimilation community), and a divergence resulting in unrealistic posterior parameter values. Besides that, the Gaussian assumption makes the assimilation process to vanish any distribution which do not have a prior Gaussian, converting it to a Gaussian one, resulting in posterior fields that have no geological meaning.

The discoveries of the Brazilian Pre-salt reservoirs are among the most relevant discoveries in recent years. These reservoirs under the thick salt-layer are giant fields with light oil located in ultra-deep waters (approximately 1900-2400m), far from the coast (approximately 240 km from Rio de Janeiro state), consisting of heterogeneous carbonate rocks. In 2018, these reservoirs already accounted for almost half of all Brazilian oil production (BRUHN et al., 2017; PETROBRAS, 2022). Figure 1 shows the annual average daily production from all Brazilian gas and oil fields operated by Petrobras, the main Brazilian petroleum company, being possible to observe the growing importance of the pre-salt reservoirs.

The characteristics of this kind of reservoir converges towards the problems aforementioned about the ensemble-based methods to perform the data assimilation, making its evaluation challenging (DA ROCHA et al., 2021).

Figure 1 – Annual average daily production from all Brazilian gas/oil fields operated by Petrobras.



Source: BRUHN *et al.* (2017)

Regarding the reservoir simulation models, carbonate reservoirs are among the most complex reservoir types to perform efficient forecasting. These difficulties arise from the fact that carbonate reservoirs are generally naturally fractured, requiring a dual-porosity, dual-permeability system to represent these fractures (MAUCEC et al., 2016). Furthermore, Super-K layers are present in several carbonate reservoirs (ALQAM; NASR-EL-DIN; LYNN, 2001), which refers to thin layers with exceptionally high permeability (MEYER; PRICE; AL-RAIMI, 2000). These layers are modeled using an object-based modeling procedure (CORREIA et al., 2015), making the uncertain parameter distribution as non-Gaussian. However, as mentioned by Maschio and Schiozer (2019), despite the increasing body of literature on history matching in complex reservoirs, the majority of studies have focused on channelized systems. Consequently, there is a lack of applications specifically tailored for conducting history matching in carbonate reservoirs.

In this scenario, we might argue that the application of IES methods in large-scale carbonate reservoirs is one of the researcher's nowadays goals. There are some techniques to remedy the limitations mentioned before. However, none of them has been developed or evaluated specifically for carbonate reservoirs.

1.1 Objectives

As mentioned in the previous section, the main objective of this thesis is to improve the history matching process with iterative ensemble smoother when it is applied in carbonate reservoirs.

The objective of this work is to improve the state-of-the-art IES application in carbonate reservoir and might be summarized as follows:

1. Developing a novel localization method that may be computationally inexpensive and suitable to be applied in carbonate reservoirs and parameterized properties.
2. Developing a novel parameterization method that will be able to preserve the Super-K layers during the assimilation process.
3. Integrating the previous methods to develop a robust workflow to apply IES in history matching in large-scale carbonate reservoirs.

CHAPTER 2 - LITERATURE REVIEW

2.1 Ensemble-based methods

To introduce the ensemble-based theory, we need first introduce briefly the original Kalman Filter (KF). The Kalman Filter (KALMAN, 1960) is an optimal recursive parameter-state estimator for linear systems where the parameter and state estimates are ‘corrected’ according to available measurements contaminated with noise. Kalman Filter workflow might be divided into two main steps: forward and analysis (or update). In the forward step, the linear model is integrated forward in time, then the parameter, state, and model covariance are updated in the analysis step whenever measurements are available through the analysis equations. Several methods aim to adapt the Kalman Filter to non-linear models. For example, in the Extended Kalman Filter (MCELHOE, 1966) the nonlinear model is linearized, and the approximated equation is used to estimate the error statistics (EVENSEN, 2003). However, these equations might be impractical to be linearized in high-dimensional systems, and the errors might be significant in highly non-linear systems.

In this scenario, the Ensemble Kalman filter (EnKF) was developed by Evensen (1994), a Monte Carlo approximation of the traditional KF, being an attractive method because it avoids the forward operator linearization, and the huge computation efforts required to evolve and store the covariances matrices in high-dimensional systems. Briefly, the Ensemble Kalman filter uses an ensemble of members to represent the prior statistics and estimate the covariances, in such a way that these covariances do not need to be evolved analytically. Thus, this ensemble is updated recursively whenever measurements are available, similarly to KF, in a parameter-state estimation problem. In the first known application of EnKF in reservoir literature, near-well permeabilities were updated using the continuous recursive methodology (NÆVDAL; MANNSETH; VEFRING, 2002). Since then, the literature of EnKF in field-scale history matching problems is vast (SKJERVHEIM et al., 2007; HAUGEN et al., 2008; SEILER et al., 2009; EMERICK; REYNOLDS, 2011a; ZHANG; OLIVER, 2011). However, the recursive methodology may lead to some issues in history matching. As the EnKF is a recursive methodology where both the parameter and states are updated, the frequent simulation restarts increase the required time to read/write simulation files. In addition, the state updates at each update step may lead to unphysical values. For more information about EnKF in history matching, Aanonsen et al. (2009) review the developments and applications of EnKF until 2009 and serve as an EnKF reference document.

To avoid the recursive methodology, van Leeuwen and Evensen (1996) developed the Ensemble Smoother (ES): a variance-minimizing analysis like EnKF update equation, however with a single update performed over the whole measurement domain in a parameter-estimation problem, avoiding the state updates. However, as mentioned by the authors, ES results in lower performance in comparison to the recursive methodology in non-linear systems, maybe because the recursive assimilations keep the ‘model on track’ (EVENSEN; EIKREM, 2018). This was also verified in the first ES experiment in history matching performed by Skjervheim et al. (2011), where the authors compared the EnKF and ES and concluded that ES resulted in poorer results, however requiring approximately 10% of the simulation time required by EnKF.

To overcome the poor quality of ES results, some iterative variants were developed. After the introduction of the Ensemble Randomized Maximum Likelihood (EnRML) (CHEN; OLIVER, 2012) and the Ensemble Smoother with Multiple Data Assimilations (ES-MDA) (EMERICK; REYNOLDS, 2013), the applications of iterative ensemble smoothers have become possible in large-scale field reservoirs. Some examples of field applications are found in Chen and Oliver (2013a) and Emerick (2016). Here, we recall the class of methods based on EnRML and ES-MDA as Iterative Ensemble Smoother (IES) methods. Since then, IES has become the preferred choice to perform history matching using ensemble-based methods (CHEN; OLIVER, 2017; EMERICK, 2018; EVENSEN, 2018; EVENSEN; EIKREM, 2018). Finally, after EnRML and ES-MDA introduction, several stochastic iterative ensemble smoothers variants were developed to overcome some IES drawbacks. Some examples are the methodologies developed by Chen and Oliver (2013b), Luo et al. (2015), Emerick (2016), Le et al. (2016), Rafiee and Reynolds (2017), Ma and Bi (2019), Luo (2021). Although these methods have revolutionized the history matching process, several issues still arise, mainly because even the iterative variants rely on Gaussian assumptions and have limited ensemble sizes. The following section introduces the basic theory of Iterative Ensemble Smoothers.

2.2 Kalman Gain Localization

To address the effects of limited ensemble size when computing covariances in ensemble-based methods, various approaches have been developed, including covariance inflation, local analysis and localization (covariance or Kalman gain). Covariance inflation involves artificially inflating the variance of the ensemble by using an inflation factor slightly greater than one to counteract the ensemble collapse (ANDERSON; ANDERSON, 1999; EVENSEN, 2009). Some examples of covariance inflation methods are found in Wang and Bishop (2003), Anderson (2007), Sacher and Bartello (2008), Anderson (2009). Local analysis decomposes the model into local domains and updates it only with data located within each subdomain (EVENSEN, 2003). Some examples of such works that used local analysis include (HOUTEKAMER; MITCHELL, 1998; OTT et al., 2004).

Localization generally refers to the tapering of either covariance matrices (HAMILL; WHITAKER; SNYDER, 2001; HOUTEKAMER; MITCHELL, 2001) or Kalman gain (CHEN; OLIVER, 2010, 2014). This tapering is performed using a localization matrix that is built based on some relationship between each parameter and data. In this context, Chen and Oliver (2017) compared the performance between localization and local analysis and concluded that both methods give equivalent results when applied in iterative ensemble smoothers. In reservoir history matching, localization is typically performed by applying the localization matrix directly to the Kalman gain. As a results, there is significant ongoing research in developing methods for constructing the tapering matrix (also known as localization matrix) for Kalman gain localization.

Developing a robust and efficient localization scheme is crucial for achieving optimal performance in data assimilation. For instance, the localization scheme can have a significant impact on the total number of iterations required for some adaptive IES variants (RANAZZI; SAMPAIO, 2019a). Therefore, it is important to consider the significance of a well-designed localization scheme to obtain accurate results.

Currently, the methods available to build the localization matrix can be classified into three main categories: distance-based; drainage-based and adaptive methods. Since its introduction by Houtekamer and Mitchell (2001), distance-based methods have become the standard approach for applying Kalman gain localization in ensemble-based methods. This approach uses a monotonic covariance function to determine tapering values based on the distance between each model parameter and data. Two widely used functions to compute the tapering (or localization) values are the Gaspari and Cohn function (GASPARI; COHN, 1999) and the isotropic covariance function defined by Furrer and Bengtsson (2007). Distance-based localization has been applied in various reservoir applications, as demonstrated in Chen and Oliver (2010, 2014), Emerick and Reynolds (2011b), Silva et al. (2017), Emerick (2018, 2019), Ranazzi and Sampaio (2019a, 2019b).

Drainage-based methods aim to define the influence region based on the reservoir dynamics. They can be defined using streamlines (ARROYO-NEGRETE et al., 2008; DEVEGOWDA; ARROYO-NEGRETE; DATTA-GUPTA, 2010; WATANABE; DATTA-GUPTA, 2012), pseudo-tracer (DAMIANI, 2007), block velocities (YEO; JUNG; CHOE, 2014).

In addition, some methods combine distance- and drainage-based localization. For instance, Emerick and Reynolds (2011a, 2011b) combine the prior field correlation lengths with the range obtained by the pseudo-tracer. Another similar example is the method proposed by Soares et al. (2018), who use each producer-injector well pair to determine the influence area used in the localization matrix definition.

In adaptive methods, the assimilation system itself (together with some statistical method) is used to determine the relationship between each parameter and data (EVENSEN, 2009). Examples in this aspect include the non-isotropic equation proposed by Furrer and Bengtsson (2007), the hierarchical localization method proposed by Anderson (2007), the smoothed ensemble correlations raised to a power (SENCORP) proposed by Bishop and Hodyss (2007), the bootstrap localization method developed by Zhang and Oliver (2010), and the sampling error correction (SEC) method (ANDERSON, 2012, 2016).

Luo et al. (2018; 2019), proposed the correlation-based localization method, where sample correlation is used to estimate the tapering values. Later, in Luo and Bhakta (2020), this method was improved by using the random-shuffle method in such way that the tapering values are computed automatically without tuning parameters.

Lacerda et al. (2019, 2021) recalled the non-isotropic equation proposed by Furrer and Bengtsson (2007), as ‘pseudo-optimal’ localization, compared it with the other adaptive localization methods, while introducing a sensitivity-analysis based localization method to perform localization for non-local parameters.

Although distance-based methods are the most used in the literature, they have several noticed limitations. Distance-based methods rely on the existence of physical locations of both model parameters and data, which makes it challenging to apply them to non-local parameters, e.g., relative permeability curves, fluid-contacts, and property multipliers. The need for a robust localization method further uprise when incorporating into reservoir data assimilation more sophisticated parameterization methods which map either model parameters or data to other non-physical domains (SEBACHER; HANEA, 2020; CANCHUMUNI et al., 2021).

Moreover, the values of tapering coefficients should also depend on a few other factors, e.g., the type of field data and the time instance (which means that even for the same type of data, the tapering coefficients at distinct time instances should be different). The drainage-based methods can address the data-type and time-dependence. However, its implementation might be difficult in certain circumstances. For example, the streamline tracing requires the availability of a streamline simulator. As for the block velocities, our previous experiments (not reported here for conciseness) showed that the localization matrix obtained through this method was not able to represent the tapering property. Furthermore, drainage-based methods do not solve the need to develop a robust workflow to perform the localization in non-local parameters.

Regarding the adaptive methods, Lacerda et al. (2019) mentioned its inferior performance compared to the distance-based methods for certain experiments therein. However, adaptive localization methods appear to be more promising to remedy the issues described above (OLIVER et al., 2021). Some works that apply adaptive localization methods (at the current stage, only correlation-based) to large-scale reservoir data assimilation problems appear to support this point of view (LUO et al., 2019; LORENTZEN et al., 2020; SOARES et al., 2021).

2.3 Parameterization

Despite the great popularity and wide use of the iterative ensemble smoothers to perform data assimilation, limitations still rely upon the Gaussian assumptions in the actual IES formulations, making its applicability limited in complex petroleum reservoirs. The standard technique to remedy the non-Gaussianity in the uncertain parameters is to perform a mapping between the original parameter domain and a Gaussian one, performing the data assimilation in this ‘transformed’ ensemble, denoted as ‘parameterization’ or ‘re-parameterization’. Since the introduction of IES, several methods have been developed to perform parameterization, such as pluri-Gaussian methods (LIU; OLIVER, 2005; SEBACHER; HANEA; HEEMINK, 2013; SEBACHER; HANEA; STORDAL, 2017; SEBACHER; STORDAL; HANEA, 2015), normal-score transform (ZHOU et al., 2011; LI et al., 2018), level-set (LORENTZEN; FLORNES; NÆVDAL, 2012; MORENO; AANONSEN, 2011), principal component analysis methods (PCA) methods (SARMA; DURLOFSKY; AZIZ, 2008; SARMA; CHEN, 2009; VO; DURLOFSKY, 2014, 2015; EMERICK, 2017), discrete cosine transform (DCT) method (ZHAO; FOROUZANFAR; REYNOLDS, 2017), high order singular value decomposition (SEBACHER; HANEA, 2020), support vector machine (JUNG et al., 2018), among others.

Due to the huge computational advances in the last decade, the application of deep learning techniques has become possible in the parameterization of complex reservoir models. One of the first methods which applied deep learning techniques to the parameterization problem was the application of a standard autoencoder (CANCHUMUNI; EMERICK; PACHECO, 2017), showing encouraging results. Another autoencoder application might be found in Kim et al. (2020a), which applied a stacked autoencoder (SAE) together with a distance-based model selection. However, as noticed in Canchumuni et al., (CANCHUMUNI; EMERICK; PACHECO, 2019a), these models fail to represent complex reservoirs, leading to the investigation of other deep learning techniques such as deep belief networks (CANCHUMUNI; EMERICK; PACHECO, 2019a). One drawback of fully connected layers structures is the high number of parameters, making them infeasible to solve large problems. To address this issue, the use of convolutional layers has been explored as a more efficient alternative, improving the performance of autoencoder in the parameterization problem, as reported in some works (LALOY et al., 2017; CANCHUMUNI; EMERICK; PACHECO, 2019b; KIM et al., 2020b).

Generative Adversarial Networks (GAN) were introduced by Goodfellow et al. (2014) in the field of deep learning, and have gained attention for their potential to learn high-dimensional data. Unlike autoencoders, which use a single network to minimize the difference between the original and the reconstructed samples using a single network, GANs introduced the concept of adversarial learning using two distinct networks: a generator and a discriminator. The generator produces data from a latent space and tries to deceive the discriminator, which classifies whether a sample is real or generated from the generator. The main difference between GANs and other generative methods is the fact that in adversarial training the original distribution is not defined explicitly, with the generator being trained based on the discriminator outputs. For a comprehensive review of the GANs basic theory and its applications in several science fields, readers are referred to Lucic et al. (2018), Gonog and Zhou (2019), Hong et al. (2020), and Iglesias et al. (2022). Furthermore, a general overview of different deep generative models might be found in Bond-Taylor (2022).

Like the other science fields, GANs also gained attention in the data assimilation field. Maybe the first work which applied generative adversarial networks to the generation of geological facies was performed by Chan and Elsheikh (2017), with an extended version presented later (CHAN; ELSHEIKH, 2019a). This work applied a Wasserstein GAN (ARJOVSKY; CHINTALA; BOTTOU, 2017) to the parameterization problem, with later improvements to the conditional generation to honor hard data (CHAN; ELSHEIKH, 2019b). Zhang et al. (2021) proposed a different method to include conditional data in the realization generation, considering a U-Net as the generator network, where the inputs now are the latent space and a field containing the conditioning data. Canchumuni et al. (2021) analyzed different GANs architectures, including transfer learning techniques, in the data assimilation quality of facies and production data, showing promising results for its application in data assimilation. The authors also introduced two localization strategies in this kind of problem: a local analysis and the application of GANs containing the latent input with spatial relationship regarding the original facies. Since then, several examples of GANs applications might be found in the literature (MOSSER; DUBRULE; BLUNT, 2018, 2019; FENG et al., 2022; RAZAK; JAFARPOUR, 2022; ZHANG et al., 2022). Finally, Bao et al. (2022) compared the GAN and variational autoencoder (VAE) performance in the data assimilation problem, where GAN showed poor performance concerning the data assimilation, however with GAN generating more realistic facies, probably because the aforementioned differences between VAE and GAN formulation.

Despite the several works regarding GANs to facies generation, some questions remain open. First, the effect of different GAN losses and architectures might be evaluated quantitively with respect to. the generated distribution. Second, in many cases, the available size of the dataset is prohibitive to train a Generative Adversarial Network properly, being the evaluation of data augmentation techniques necessary. Finally, the performance of GAN is not well established since most works apply its methodologies in the same two facies channelized toy model, being required the performance evaluation in other reservoir types.

Data augmentation is the set of techniques that artificially increases the size of the training dataset to improve the training process. There are several methods to perform the data augmentation, for instance, geometric transformation, color space augmentations, mixing images, even GAN itself to generated new dataset samples (SANDFORT et al., 2019). For a comprehensive review about data augmentation techniques, reader are referred to Wong et al. (2016) and Shorten and Khoshgoftaar (2019). Although data augmentation is a well-established technique in several deep learning fields, such as text data (SHORTEN; KHOSHGOFTAAR; FURHT, 2021), vocal speech (JAITLY; HINTON, 2013; KIM et al., 2019), medical imaging (MOTAMED; ROGALLA; KHALVATI, 2021; ZHANG et al., 2023), nuclear physics (BAHTIYAR; SOYDANER; YÜKSEL, 2022), its application in generative adversarial networks is not well established. This is primarily due to the fact that when standard data augmentation techniques are applied, the generative models learn to generate the augmented distribution, rendering the need for explicit data augmentation less apparent in GANs.Karras et al. (2020) called this phenomenon as ‘leaky’ augmentations. To address this issue, various works have introduced state-of-the-art augmentation methods in GANs, having as principle the application of invertible transformations in both the real and fake samples before the discriminator in the workflow. One example is the Data Augmentation Optimized for GAN method – DAG

(TRAN et al., 2021), which uses different transformation operators in different discriminators, where the generator and the discriminator losses now includes the loss of the discriminator without any transformation as well as the losses from the discriminators that receives transformed samples with different operators. For our understanding, this more complex network structure might be prohibitive, especially for users with lower computational power. Other methodologies introduce these transformations without changing the loss structure. Zhao et al. (2020a) compared three cases: augment only real samples, augment all samples when training the discriminator, and augment all samples when training the discriminator and the generator, proposing the *DiffAugment* method. Zhao et al. (2020b) proposed a similar method, called Contrastive Loss for GAN Training, where the contrastive loss is applied to regularizing the discriminator on the augmented samples, the authors also analyzed the strength effect of each augmentation in the training, concluding that Instance Noise (SØNDERBY et al., 2017) does no showed good improvements during the GAN training. Regarding the strength of the data augmentation, it has been shown that it can vary depending on the dataset, network settings, and even over iterations during the training, as demonstrated by Karras et al. (2020). To address this issue, they introduced the Adaptive Discriminator Augmentation (ADA) method, which applies all possible augmentations to real and generated images only with a certain probability. This approach ensures that the generator avoids the leaking effect. Despite all these advances in data augmentation applied in GAN training, its performance is not known when applied to petroleum reservoir studies (to our knowledge, no work has applied data augmentation in petroleum reservoirs). His evaluation is crucial as the augmentation behavior can vary from the current state-of-the-art approaches due to the distinct nature of the reservoir samples. For instance, reservoir samples typically consist of single color channels containing integer values, which differ from the standard GAN applications that often involve diverse subjects like human faces, cars, and landscapes.

2.4 History Matching Problem

2.4.1 Iterative Ensemble Smoothers

In this section, we will introduce the mathematical background of the Iterative Ensemble Smoothers (IES) theory, and the localization methods used nowadays. First, in the general history matching formulation, the imperfect forward operator prediction $\mathbf{d} \in \Re^{N_d}$ depends nonlinearly on the input model parameters $\mathbf{m} \in \Re^{N_m}$ (EVENSEN, 2019):

$$\mathbf{d} = \mathbf{g}(\mathbf{m}). \tag{1}$$

Here, the forward operator $\mathbf{g}$ represents the reservoir simulator (i.e., usually a 'black-box' operator). In the history matching inverse problem, we have some set of noisy measurements $\mathbf{d}_{\mathrm{obs}}$ of the true data value $\mathbf{d}_{\mathrm{true}}$:

$$\mathbf{d}_{\mathrm{obs}} = \mathbf{d}_{\mathrm{true}} + \epsilon, \tag{2}$$

where the error ϵ is usually drawn from $\epsilon \sim \mathcal{N}(0, \mathbf{C}_{\mathrm{D}})$, with $\mathbf{C}_{\mathrm{D}} \in \Re^{N_d \times N_d}$ representing the covariance matrix of measurement errors. Thus, is straightforward to define the history matching as the inverse problem $\mathbf{d}_{\mathrm{obs}} = \mathbf{g}(\mathbf{m})$. From Bayes' theorem, Evensen (2018, 2019) demonstrated that the posterior marginal distribution can be obtained from:

$$f(\mathbf{m}|\mathbf{d}_{\mathrm{obs}}) \propto \exp\left(-\frac{1}{2}J(\mathbf{m})\right), \tag{3}$$

where function $J(\mathbf{m})$ is given by:

$$J(\mathbf{m}) = [\mathbf{d}_{\mathrm{obs}} - \mathbf{d}]^{\mathrm{T}} \mathbf{C}_{\mathrm{D}}^{-1} [\mathbf{d}_{\mathrm{obs}} - \mathbf{d}] + \left[\mathbf{m} - \mathbf{m}_{\mathrm{pr}}\right]^{\mathrm{T}} \mathbf{C}_{\mathrm{M}}^{-1} \left[\mathbf{m} - \mathbf{m}_{\mathrm{pr}}\right], \tag{4}$$

where the subscript pr denotes the priors. Thus, it is possible to conclude that the maximum pdf (3) is obtained minimizing the objective function (4). In ensemble-based methods we can sample the priors to represent the covariances, defining the cost function for each sample realization (or ensemble member). In this sense, the cost function can be derived by the stochastic smoother by van Leeuwen (2020):

$$\begin{aligned} J(\mathbf{m}_j) = {} & \left[\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_j - \epsilon_j\right]^{\mathrm{T}} \mathbf{C}_{\mathrm{D}}^{-1} \left[\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_j - \epsilon_j\right] \\ & + \left[\mathbf{m}_j - \mathbf{m}_{\mathrm{pr},j}\right]^{\mathrm{T}} \mathbf{C}_{\mathrm{M}}^{-1} \left[\mathbf{m}_j - \mathbf{m}_{\mathrm{pr},j}\right], \end{aligned} \tag{5}$$

for an ensemble $\{\mathbf{m}_j\}_{j=1}^{N}$ with N members. In the following, we will analyze the general solution by the Ensemble Smoother with Multiple Data Assimilation.

2.4.2 Ensemble Smoother with Multiple Data Assimilations

The Ensemble Smoother with Multiple Data Assimilation (ES-MDA) introduced by Emerick and Reynolds (2013) is a variant of the original Ensemble Smoother which performs multiple ES iterations using an inflated covariance measurement error to damp the iterations. To solve (5), ES-MDA uses a tempering procedure of the likelihood function (EMERICK, 2016; EVENSEN, 2018, 2019) over N_i multiple assimilations:

$$f(\mathbf{m}|\mathbf{d}_{\mathrm{obs}}) = \prod_{i=1}^{N_i} f(\mathbf{m}|\mathbf{d}_{\mathrm{obs}})^{\frac{1}{\alpha_i}}, \tag{6}$$

where $\alpha_1, \alpha_2, \dots, \alpha_{N_i}$ is the inflation factor which damps each iteration i. From (6), we can conclude that the posterior variance is correct only if the following condition is satisfied:

$$\sum_{i=1}^{N_i} \frac{1}{\alpha_i} = 1. \tag{7}$$

Now, combining the solution introduced by Evensen (2019) and the stochastic ensemble smoother introduced by van Leeuwen (2020). The cost function of the weak-constraint ES-MDA is given by recursive minimizations over N_i iterations:

$$\begin{aligned} J(\mathbf{d}_{i+1}) = {} & \left[\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_j - \sqrt{\alpha_i}\epsilon_{j,i}\right]^{\mathrm{T}} \alpha_i \mathbf{C}_{\mathrm{D}}^{-1} \left[\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_j - \sqrt{\alpha_i}\epsilon_{j,i}\right] \\ & + \left[\mathbf{m}_{j,i+1} - \mathbf{m}_{j,i}\right]^{\mathrm{T}} \mathbf{C}_{\mathrm{M},i}^{-1} \left[\mathbf{m}_{j,i+1} - \mathbf{m}_{j,i}\right]^{\mathrm{T}}. \end{aligned} \tag{8}$$

Generally, in history matching problems, the number of model parameters is much higher than the number of ensemble members ($N_m \gg N$). In this context, the ES-MDA solution will be (EMERICK; REYNOLDS, 2013; EVENSEN, 2019):

$$\mathbf{m}_{j,i+1} = \mathbf{m}_{j,i} + \mathbf{C}_{\mathrm{MD},i}\left(\mathbf{C}_{\mathrm{DD},i} + \alpha_i \mathbf{C}_{\mathrm{D}}\right)^{-1}\left(\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_{j,i} - \sqrt{\alpha_i}\epsilon_j\right), \tag{9}$$

where $\epsilon_j \sim N(0, \mathbf{C}_{\mathrm{D}})$. In ensemble-based methods, covariances are estimated around the ensemble mean with size N:

$$\mathbf{C}_{\mathrm{MD}} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\mathbf{m}_{j,i} - \bar{\mathbf{m}}_i\right)\left(\mathbf{d}_{j,i} - \bar{\mathbf{d}}_i\right)^{\mathrm{T}}, \tag{10}$$

and,

$$\mathbf{C}_{\mathrm{DD}} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\mathbf{d}_{j,i} - \bar{\mathbf{d}}_i\right)\left(\mathbf{d}_{j,i} - \bar{\mathbf{d}}_i\right)^{\mathrm{T}}, \tag{11}$$

where overbars denote ensemble mean. In this stochastic iterative smoother, the model is integrated forward in time, and model parameters and errors are updated recursively at each iteration step until the condition in (7) is achieved. Thus, a critical choice parameter in ES-MDA is the inflation factors α_i definition for all iterations, which determines the damping of each iteration and the algorithm termination. For simplicity, its common to define a predetermined number of N_i updates with $\alpha_i = N_i$ for all iterations. However, the inflation factor definition is an intensive research subject, with the development of methodologies to optimize this selection. For example, geometric selection (RAFIEE; REYNOLDS, 2017; EMERICK, 2019) and adaptive methodologies where either number of iterations or the inflations factors are determined automatically (LE; EMERICK; REYNOLDS, 2016; MA; BI, 2019). Independently of the inflation factor selection method, the general stochastic ES-MDA algorithm may be summarized as follows:

Algorithm 1 General Stochastic ES-MDA

```
Set $\{\alpha_i\}_{i=1}^{N_i}$ definition scheme
while $\sum_{i=1}^{N_i} \frac{1}{\alpha_i} < 1$
    if $i = 1$
        Sample $\mathbf{m}_{j,1}$
    end if
    for $j = 1{:}N$
        Compute $\mathbf{d}_{j,i} = \mathbf{g}(\mathbf{m}_{\boldsymbol{j,i}})$
    end for
    for $j = 1{:}N$
        Compute $\epsilon_j \sim N(0, \mathbf{C}_\mathrm{D})$
        Compute $\mathbf{m}_{j,i+1} = \mathbf{m}_{j,i} + \mathbf{C}_{\mathrm{MD},i}\left(\mathbf{C}_{\mathrm{DD},i} + \alpha_i \mathbf{C}_\mathrm{D}\right)^{-1}\left(\mathbf{d}_\mathrm{obs} - \mathbf{d}_{j,i} - \sqrt{\alpha_i}\epsilon_j\right)$
    end for
    $i \leftarrow i + 1$
end while
```

In the following sections, we will introduce the existing techniques to reduce the effects of limited ensemble size and non-Gaussian parameters.

2.4.3 Kalman Gain Localization

In Kalman gain localization, the analysis equation is rewritten adding the localization matrix which is applied to the entire Kalman gain matrix by the Schur product (element-wise matrix multiplication):

$$\mathbf{m}_{j,i+1} = \mathbf{m}_{j,i} + \mathbf{P} \circ \left[\mathbf{C}_{\mathrm{MD},i}\left(\mathbf{C}_{\mathrm{DD},i} + \alpha_i \mathbf{C}_\mathrm{D}\right)^{-1}\right]\left(\mathbf{d}_\mathrm{obs} - \mathbf{d}_{j,i} - \sqrt{\alpha_i}\epsilon_j\right) \tag{12}$$

where $\mathbf{P}$ represents the localization matrix, which is given by:

$$\mathbf{P} = \begin{bmatrix} \rho_{\mathrm{m}_1,\mathrm{d}_1} & \rho_{\mathrm{m}_1,\mathrm{d}_2} & \cdots & \rho_{\mathrm{m}_1,\mathrm{d}_{N_d}} \\ \rho_{\mathrm{m}_2,\mathrm{d}_1} & \rho_{\mathrm{m}_2,\mathrm{d}_2} & & \rho_{\mathrm{m}_2,\mathrm{d}_{N_d}} \\ \vdots & & \ddots & \vdots \\ \rho_{\mathrm{m}_{N_m},\mathrm{d}_1} & \rho_{\mathrm{m}_{N_m},\mathrm{d}_2} & \cdots & \rho_{\mathrm{m}_{N_m},\mathrm{d}_{N_d}} \end{bmatrix}. \tag{13}$$

where each $\rho_{m,d}$ represents the tapering value corresponding to m parameter and d data.

In distance-based, some monotonical covariance function is used to compute the entries of the localization matrix. The Gaspari and Cohn correlation (1999) is defined by:

$$\rho(z) = \begin{cases} -\frac{1}{4}\left(\frac{z}{L}\right)^5 + \frac{1}{4}\left(\frac{z}{L}\right)^4 + \frac{5}{8}\left(\frac{z}{L}\right)^3 - \frac{5}{3}\left(\frac{z}{L}\right)^2 + 1, & 0 \leq z \leq L \\ \frac{1}{12}\left(\frac{z}{L}\right)^5 - \frac{1}{2}\left(\frac{z}{L}\right)^4 + \frac{5}{8}\left(\frac{z}{L}\right)^3 + \frac{5}{3}\left(\frac{z}{L}\right)^2 - 5\left(\frac{z}{L}\right) + 4 - \frac{2}{3}\left(\frac{L}{z}\right) & L \leq z \leq 2L \\ 0, & 2L \leq z \end{cases} \tag{14}$$

where z represents the Euclidean distance between each parameter and data, and L is a normalizing parameter denoted as critical length. Another example of correlation function is the isotropic covariance function given by Furrer and Bengtsson (2007). This function states that the tapering strength also depends on the ensemble size, such that the tapering values tends to one as ensemble size tends to infinity:

$$\rho(z) = \frac{1}{1 + [1 + f(0)^2/f(z)^2]/N}. \tag{15}$$

Here, $f(\cdot)$ represents a covariance function.

2.4.4 Adaptive localization

This chapter will introduce two preexisting adaptive localization techniques: pseudo-optimal and correlation-based localization. The motivation here is to further improve the adaptive localization scheme, by combining the strengths of these methods.

Furrer and Bengtsson (2007) derived a non-isotropic tapering function by minimizing the norm of the difference between the true covariance and the ensemble tapered covariance. This tapering function is in the following form:

$$\rho_{p,o} = \frac{c_{p,o}^2}{c_{p,o}^2 + \left(c_{p,o}^2 + c_{p,p}c_{o,o}\right)/N}, \tag{16}$$

where $\rho_{p,o}$ is the tapering value (entry of localization matrix $\mathbf{P}$) for a given parameter p and data o, and $c_{p,o}$ represents the true covariance between the pth model parameter and oth predicted data (LACERDA; EMERICK; PIRES, 2019). An important feature of this pseudo-optimal localization scheme is that this method will be optimal asymptotically (tapering values tend to one as the ensemble size N

approaches infinity). The authors propose an approximation that replaces $c_{p,o}$ by the ensemble covariance estimate instead to obtain an inexpensive non-isotropic localization method. In addition, Furrer and Bengtsson (2007) suggest that sparseness may be introduced by setting $\rho_{p,o} = 0$ whenever $c_{p,o} \approx 0$. In line with this idea, Lacerda et al. (2019) advocate to set:

$$\rho_{p,o} = 0, \text{ if } \frac{|c_{p,o}|}{\sqrt{c_{p,p}c_{o,o}}} < \theta, \tag{17}$$

where $|c_{p,o}|/\sqrt{c_{p,p}c_{o,o}}$ is the sample correlation (normalized covariance), and θ is a user-defined threshold. The authors analyzed different threshold values and concluded that a threshold value equal to $\theta = 10^{-3}$ results in reasonably good assimilation results with satisfactory data mismatch. However, as we will see in the following sections, the pseudo-optimal localization will not always be able to appropriately suppress spurious correlations following the form in equation(16).

Similarly, the correlation-based localization scheme suggests defining the localization matrix entries equal to one if the sample correlation between the pth model parameter and oth predicted data is higher than a user-defined threshold:

$$\rho_{p,o} = \mathbf{I}\left(|\hat{c}_{p,o}| > \theta\right), \tag{18}$$

where $\mathbf{I}$ is the indicator function (1 if the condition is satisfied, 0 otherwise), $\hat{c}_{p,o}$ is the sample correlation between the pth model parameter and the oth data, and θ is the threshold in correlation-based localization. Thus, analyzing equations (17) and (18), we can conclude that the same methodologies can be applied to estimate the threshold in both localization methods. Luo et al. (2018; 2019) suggest grouping similar model variables (e.g., the same type of petrophysical property) into a correlation field, to reduce the complexity to obtain this threshold. Furthermore, Luo et al. (2018) showed that the distribution of the sampling errors in this correlation field can be approximated by a Gaussian distribution with zero mean. Thus, it is possible to decompose the correlation field for a given group G, and a given data o as follows:

$$\hat{\boldsymbol{c}}_{G,o} = \hat{\boldsymbol{c}}_{G,o}^{\infty} + \boldsymbol{\varepsilon}_{G,o}. \tag{19}$$

Here, $\boldsymbol{\varepsilon}_G$ is the vector with the sampling errors for all parameters which belong to the group G. Thus, with an estimated noise level of $\boldsymbol{\varepsilon}_{G,o}$ we can calculate a unique threshold $\theta_{G,o}$ for this group of petrophysical property. More specifically, the authors proposed the application of the median absolute deviation (MAD) estimator and the universal rule (UR) (DONOHO; JOHNSTONE, 1994, 1995) to estimate the threshold for a given group G:

$$\sigma_{G,o} = \frac{\text{median}(|\boldsymbol{\varepsilon}_{G.o}|)}{0.6745}, \tag{20}$$

and,

$$\theta_{G,o} = \sqrt{2\ln\left(\#\boldsymbol{\varepsilon}_{G,o}\right)}\,\sigma_{G,o}, \tag{21}$$

where $\#\boldsymbol{\varepsilon}_{G,o}$ is the number of elements in group G. Luo and Bhakta (2020) suggested creating a substitute noise field $\tilde{\boldsymbol{\varepsilon}}_{G,o}$ to estimate the threshold, assuming that the noise field is induced by sampling errors. This substitute noise field can be generated by exploiting the fact that the true correlation between two statistically independent variables should be equal to zero. Using the ensembles $\mathbf{M} = \{\mathbf{m}_j\}_{j=1}^{N}$ and $\mathbf{D} = \{\mathbf{d}_j\}_{j=1}^{N}$, one can shuffle the positions of the members in the ensemble of model parameters to obtain an ensemble $\widetilde{\mathbf{M}}$ with the same statistics as the original ensemble $\mathbf{M}$, and with the additional property that the correlation between $\widetilde{\mathbf{M}}$ and $\mathbf{D}$ will be equal to zero for an infinite ensemble size. To do this, Luo and Bhakta (2020) suggest to randomly shuffle the column indexes of $\mathbf{M}$.

To avoid the binary behavior of (18), Luo and Bhakta (2020) also developed a continuous tapering rule based on the sample correlations and the estimated noise level. Specifically, the tapering values in this continuous tapering method are estimated by the following relationship:

$$\rho_{p,o} = f_t\left(\frac{1 - \left|\hat{c}_{p,o}\right|}{1 - \theta_{G,o}}\right), \tag{22}$$

where f_t is a given tapering function. In their work, the authors adopted the Gaspari and Cohn function (GASPARI; COHN, 1999).

Correlation-based localization will be suboptimal asymptotically as ensemble size tends to infinity (or computing tapering coefficients with the true correlation value). This occurs because, in its present form, we will only obtain tapering values equal to one if the correlation between a given pair of variables and data is one.

Furthermore, when $N \rightarrow \infty$, $\theta \rightarrow 0$ and $\left|\hat{c}_{p,o}\right| \rightarrow 0,$ one has $\rho_{p,o} \rightarrow 5/24$ instead of being 0.

2.4.5 Parameterization

The parameterization to perform the data assimilation using ensemble-based methods is based in the mapping between the original parameter domain to a domain which its value arises from a normal distribution. Mathematically, this might be represented by the operator p, responsible for the mapping between both fields:

$$p : \mathbf{m} \in \Re^{N_m} \rightarrow \mathbf{z} \in \Re^{N_z} \tag{23}$$

A fundamental aspect in parameterization is that this operator needs to be invertible, i.e., have an inverse transform p^{-1} to the original domain:

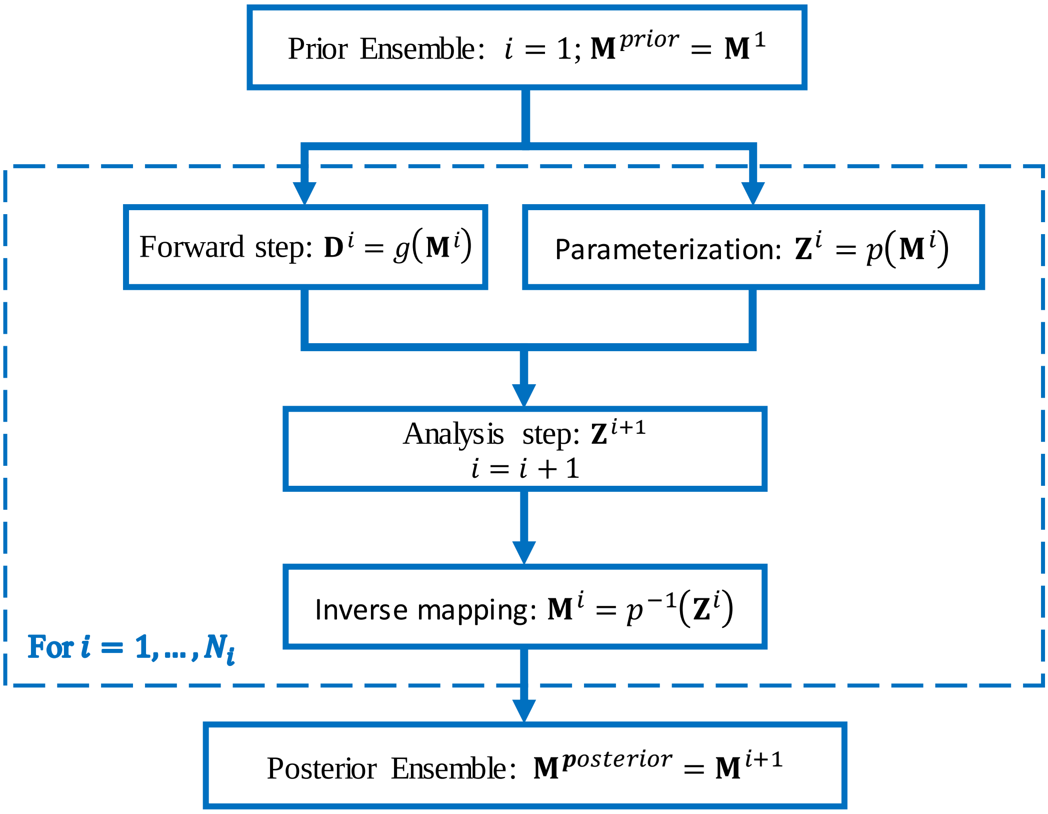
$$p^{-1} : \mathbf{z} \in \Re^{N_z} \rightarrow \mathbf{m} \in \Re^{N_m} \tag{24}$$

Denoting parameter and data ensemble as $\mathbf{M}^{i} = \left\{\mathbf{m}_{j}^{i}\right\}_{i=1}^{N}$ and $\mathbf{D}_{j}^{i} = \left\{\mathbf{d}_{j}^{i}\right\}_{j=1}^{N}$, the uncertain parameter ensemble is used to obtain the data ensemble and the parameterized one, then the analysis step is made using this parameterized ensemble as uncertain. In the ES-MDA solution this will be represented as:

$$\mathbf{z}_{j,i+1} = \mathbf{z}_{j} + \mathbf{C}_{\mathrm{ZD},i}\left(\mathbf{C}_{\mathrm{DD},i} + \alpha_{i}\mathbf{C}_{\mathrm{D}}\right)^{-1}\left(\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_{j,i} - \sqrt{\alpha_{i}}\epsilon_{j}\right), \tag{25}$$

where $\mathbf{z}$ represents the vector after the parameterization $\mathbf{z} = p(\mathbf{m})$. In this sense, the assimilation workflow considering the parameterized ensemble can be represented as Figure 2.

Figure 2 – Parameterization workflow.


Prior Ensemble: $i = 1; \mathbf{M}^{prior} = \mathbf{M}^1$
Forward step: $\mathbf{D}^i = g(\mathbf{M}^i)$
Parameterization: $\mathbf{Z}^i = p(\mathbf{M}^i)$
Analysis step: $\mathbf{Z}^{i+1}$
$i = i + 1$
Inverse mapping: $\mathbf{M}^i = p^{-1}(\mathbf{Z}^i)$
For $i = 1, \ldots, N_i$
Posterior Ensemble: $\mathbf{M}^{posterior} = \mathbf{M}^{i+1}$

2.4.6 Parameterization with Generative Adversarial Networks

Given the discriminator network $\mathcal{D}$ and the generator network $\mathcal{G}$, the groundbreaking work of Goodfellow et al. (2014) proposed an alternative rule where instead of a likelihood, where the discriminator serves as objective for a generative model, where the generator objective is to learn its output distribution p_g to the original data distribution p_d. In this point of view the learning framework might be considered a two-player game between the generator and the discriminator: the generator aims to generate real-like samples and the discriminator aims to distinguish between real samples from the original data and the samples generated by the generator, thus, each network has its own objective where the convergence of this two-player game will be the Nash-equilibrium (HEUSEL et al., 2018; LUCIC et al., 2018). The formulation presented the problem as the following min-max objective:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathbb{E}_{x \sim p_d}\left[\log\left(\mathcal{D}(x)\right)\right] + \mathbb{E}_{\hat{x} \sim p_g}\left[\log\left(1 - \mathcal{D}(\hat{x})\right)\right] \tag{26}$$

In the optimization problem from (26), the generator learns to generate samples that have a low probability of being fake. However, in the early training, the large difference between generated and real samples can make $\mathcal{D}$ reject samples. In this case, $\log\big(1 - \mathcal{D}(\hat{x})\big)$ saturates (GOODFELLOW et al., 2014). Thus, the authors proposed a variation in the generator objective, training $\mathcal{G}$ to maximize the probability of generated images being real: $\log\big(\mathcal{D}(\hat{x})\big)$. Here, we will refer to this version as Non-saturating GAN (NS-GAN), and we will only evaluate the non-saturating version of the standard GAN.

Some works state that the Jensen-Shannon divergence used in the original formulation of GAN (GOODFELLOW et al., 2014) might be not the best option, since this divergence might be discontinuous with respect to the parameters of the generator (MESCHEDER; GEIGER; NOWOZIN, 2018). Thus, (ARJOVSKY; CHINTALA; BOTTOU, 2017) replaced the Jensen-Shannon divergence with the Wasserstein-Distance, proposing the Wasserstein GAN (WGAN). In WGAN, the role of the discriminator is no longer output probabilities, now the discriminator (also called critic) aims to measure the Wassertein-1 distance between the original and generated distributions.

$$\min_{\mathcal{G}} \max_{\mathcal{D}\in\mathfrak{D}} \mathbb{E}_{x\sim p_d}[\mathcal{D}(x)] - \mathbb{E}_{\hat{x}\sim p_g}[\mathcal{D}(\hat{x})] \tag{27}$$

Where $\mathfrak{D}$ represents the set of 1-Lipschitz continuous functions. To enforce Lipschitz, Arjovsky et al. (2017) proposed a weight clipping method for the discriminator lies in a compact space.

Unregularized GANs are not guaranteed to converge to the Nash-equilibrium, as shown in Mescheder et al. (2018). To address this issue, various regularization techniques have been developed to stabilize the training process, as they have shown to improve its robustness (ROSS; DOSHI-VELEZ, 2017). One such technique is instance noise, which involves add a decaying over time Gaussian noise into the inputs of the discriminator (SØNDERBY et al., 2017). In the context of Wasserstein GAN, weight clipping was demonstrated as not being the best method to ensure Lipschitz. In this way, Gulrajani (2017) introduced an alternative to ensure Lipschitz by introducing a gradient penalty regularization in the discriminator loss:

$$R_{GP} = \lambda \mathbb{E}_{\tilde{x}\sim p_{\tilde{d}}}[(\|\nabla\mathcal{D}(\tilde{x})\| - g_0)^2] \tag{28}$$

where $\tilde{x}$ is implicitly defined sampling uniformly between the original and generated distributions, and γ is the penalty coefficient (a little modified from the original paper). In the original paper, the authors used $\lambda = 10$ and $g_0 = 1$ for all experiments, suggesting that these values work across different architectures and datasets.

2.5 Objective functions

In this section, we introduce some metrics, or objective functions used to verify the performance of the data assimilation methods.

2.5.1 Normalized data-mismatch objective function

For each ensemble member j, the normalized data-mismatch objective function represents the difference between simulator output and the measurements, normalized by the measurements error covariance matrix:

$$O_{N_d,j} = \left(\mathbf{d}_j - \mathbf{d}_{\mathrm{obs},j}\right)^{\mathrm{T}} \mathbf{C}_{\mathbf{D}}^{-1} \left(\mathbf{d}_j - \mathbf{d}_{\mathrm{obs},j}\right), \tag{29}$$

and its average:

$$\overline{O_{N_d}} = \frac{1}{N} \sum_{j=1}^{N} O_{N_d,j}. \tag{30}$$

2.5.2 Model-mismatch objective function

As mentioned in other works (CHEN; OLIVER, 2017; RANAZZI; SAMPAIO, 2019a), the objective function of model mismatch in a form similar to that in equation (30) is impractical in high-dimensional models, due to the high computational cost in calculating the inverse matrix $\mathbf{C}_{\mathbf{M}}^{-1}$. Thus, we adopt an approximation to compute the discrepancy between the reference model and posterior model parameters:

$$O_{N_m,j} = \frac{1}{N_m} \sum_{p=1}^{N_m} \left(\frac{\mathbf{m}_{p,j}^{posterior} - \mathbf{m}_{p,j}^{prior}}{\sigma_p^{prior}} \right)^2, \tag{31}$$

and its average:

$$\overline{O_{N_m}} = \frac{1}{N}\sum_{j=1}^{N} O_{N_m,j}. \tag{32}$$

Here, it is important to note that in real case applications, the reference model is unknown, being impossible to compute the difference between the posterior ensemble and the true reservoir. In this scenario, the difference is calculated between the prior and posterior ensemble: because the ill-posedness it is desirable that the posterior ensemble be as close as possible to the prior ensemble.

2.5.3 Normalized variance

Normalized variance represents the ratio between the variance of the posterior and prior parameter. Thus, averaged normalized variance is defined by the following formula (OLIVER; REYNOLDS; LIU, 2008; EMERICK, 2018):

$$\mathrm{ANV} = \frac{1}{N_m}\sum_{p=1}^{N_m} \frac{c_{p,p}^{posterior}}{c_{p,p}^{prior}}. \tag{33}$$

CHAPTER 3 - METHODOLOGY

3.1 Improving pseudo-optimal Kalman gain localization using the random-shuffle method

To improve the performance of adaptive localization methods, we propose a new localization scheme that merges the positive aspects of the aforementioned methods: the optimality of the pseudo-optimal localization and the spurious correlation removal of the correlation-based localization.

The methodology developed in this work is based on some assumptions. First, we can extend the idea behind the decomposition of the noise field with respect to the sample correlations to that with respect to the sample covariances. Following this notion, the sample covariances obtained through a finite ensemble size can be decomposed as:

$$c_{p,o} = c_{p,o}^{\infty} + \varepsilon_{p,o}, \tag{34}$$

where the obtained sample covariance $c_{p,o}$ is the sum of the true covariance obtained with an infinite ensemble size $c_{p,o}^{\infty}$ and the noise due to the limited ensemble size. Finally, a threshold can be computed using the same method as in the correlation-based localization, using the Median Absolute Deviation (MAD) and the Universal Rule (UR).

Following the rationale behind Luo and Bhakta (2020), a reasonable choice is to stay on the 'safe side' by setting the localization matrix values $\rho_{p,o} = 0$ whenever the sample covariance obtained is smaller than the noise level. In this regard, a simple idea is to set all sample covariances equal to 0 if their original values are lower than the threshold obtained, using (17). However, our experience with this truncation rule (not reported here for conciseness) indicates that it resulted in rough localization matrices (especially in high-dimensional systems) and consequently, poor assimilation results.

To perform a smoother tapering, an approximation of the pseudo-optimal localization needs to be performed, considering the hypothesis of Hamill, Withaker and Snyder (2001) and Lacerda et al. (2019) where the variances are correctly estimated, while the covariances contain more substantial sampling errors due to the limited ensemble size:

$$\mathbf{C} = \begin{bmatrix} c_{1,1} & c_{2,1} + \beta \\ c_{1,2} + \beta & c_{2,2} \end{bmatrix}, \tag{35}$$

where β represents the sampling errors in the estimated sample covariances.

Using this generic covariance matrix, we can add sampling errors into the pseudo-optimal localization (POL) by inflating the sample covariances in the objective function obtained by Furrer and Bengtsson (2007), when $i \neq j$ (please see Lemma 7 of Furrer and Bengtsson, 2007):

$$\min_{\rho_{p,o}} \left(\sum_{p,o} c_{p,o}^2 - 2\rho_{p,o} c_{p,o}^2 + \rho_{p,o}^2 \left(c_{p,o}^2 + \frac{c_{p,o}^2}{N} + \frac{c_{p,p} c_{o,o}}{N} + \beta_{p,o}^2 \right) \right), \tag{36}$$

where $\beta_{p,o}^2$ is the sampling errors, presented here as a penalty factor. Similarly, following the work of Furrer and Bengtsson (2007), the term-by-term minimization results in:

$$\rho_{p,o} = \frac{c_{p,o}^2}{c_{p,o}^2 + \left(c_{p,o}^2 + c_{p,p} c_{o,o}\right)/N + \beta_{p,o}^2}. \tag{37}$$

Therefore, it is reasonable to use the random-shuffle method of Luo and Bhakta (2020) to compute a threshold for the covariance field in equation (34), and then use the obtained threshold as the penalty factor in (37). Moreover, as pointed out by Lacerda et al. (2019), weaker correlations are harder to estimate, or in other words, weaker correlations are more affected by the sampling errors. In the sequel, we evaluate four different methods to calculate the threshold based on the correlations between model parameters and data. It is important to note that in the current work, the calculation of this penalty factor is an ad hoc procedure because the non-isotropic tapering function of Furrer and Bengtsson (2007) will be guaranteed to be optimal only if the true covariances are used therein.

3.1.1 Methods to compute the penalty factor

In the present work, we evaluate different methods to define the penalty term in equation (37), which is in a general form of

$\beta_{p,o}^{F} = F(c_{p,o}, c_{p,p}, c_{o,o}) \times \theta_{p,o}$, where $F(c_{p,o}, c_{p,p}, c_{o,o})$ is a “factor function” that potentially depends on $c_{p,o}$, $c_{p,p}$ and $c_{o,o}$. Here we consider four choices for the factor function $F(c_{p,o}, c_{p,p}, c_{o,o})$: 1) $F(c_{p,o}, c_{p,p}, c_{o,o})$ is a fixed constant (POL-F); 2) $F(c_{p,o}, c_{p,p}, c_{o,o})$ has a linear relation to the squared sample correlation $c_{p,o}^2/(c_{p,p}c_{o,o})$ (POL-L); 3) $F(c_{p,o}, c_{p,p}, c_{o,o})$ has a nonlinear relation to the squared sample correlation $c_{p,o}^2/(c_{p,p}c_{o,o})$ through the Gaspari and Cohn function (POL-GC); 4) $F(c_{p,o}, c_{p,p}, c_{o,o})$ is an empirical exponential function of the sample covariance $c_{p,o}$ (POL-EXP).

Specifically, in the first case, the penalty term is simply the threshold obtained by the random shuffle method:

$$\beta_{p,o}^{F} = \theta_{p,o}. \tag{38}$$

In the second choice, the penalty term equals the calculated threshold multiplied by a factor $1 - c_{p,o}^2/(c_{p,p}c_{o,o})$, i.e.:

$$\beta_{p,o}^{L} = \left(1 - \frac{c_{p,o}^2}{c_{p,p}c_{o,o}}\right)\theta_{p,o}, \tag{39}$$

which means that the closer the squared sample correlation $c_{p,o}^2/(c_{p,p}c_{o,o})$ to 1, the smaller the penalty term is.

In the third and the fourth methods, we consider the choice that the factor function is nonlinear. Specifically, in the third method, we let the factor function depend on the squared sample correlation through the Gaspari and Cohn (GC) function, as follows:

$$\beta_{p,o}^{GC} = f_t\left(2\frac{c_{p,o}^2}{c_{p,p}c_{o,o}}\right)\theta_{p,o}. \tag{40}$$

In the fourth method, we performed a univariate experiment to develop an empirical factor function and transfer the learned empirical function to higher-dimensional settings. In the experiment, for each different ensemble size N, we generate an ensemble of scalars m and d by drawing samples from a normal distribution:

$$\begin{bmatrix} m \\ d \end{bmatrix} \sim Normal\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & c_{\text{true}} \\ c_{\text{true}} & 1 \end{bmatrix}\right), \tag{41}$$

where c_{true} is the true covariance between m and d. Then, we computed the tapered Kalman gain and the true Kalman gain with $\mathbf{C_D} = 1$ for each ensemble size. The tapering coefficient, following equation (16), was computed using a penalty factor equal to $\beta = L.\theta$, with varying $c_{o,p}$ and L. For each $c_{o,p}$ and L we repeated the computation several times, obtaining $K_{c,L} = \left[K_{c,L}^{1}, K_{c,L}^{2}, \ldots, K_{c,L}^{t}\right]$ and computed the norm between these ensembles of tapered matrices and the true Kalman gain. We found that the curve which minimizes $\left|\rho \circ K_{c,L} - K_{\text{true}}\right|$ follows approximately:

$$\beta_{p,o}^{exp} = \theta_{p,o}.\exp\left(-\frac{6c_{p,o}}{L}\right), \tag{42}$$

with $L = 1.5/\sqrt{N}$. Figure 3 shows the obtained L values for different ensemble sizes. Furthermore, in Figure 4, it is possible to see the difference between the shape for each method to define the penalty factor for a general covariance varying between 0 and 1 for a threshold $\theta = 0.2$.

Figure 3 - Empirical L for different ensemble sizes.

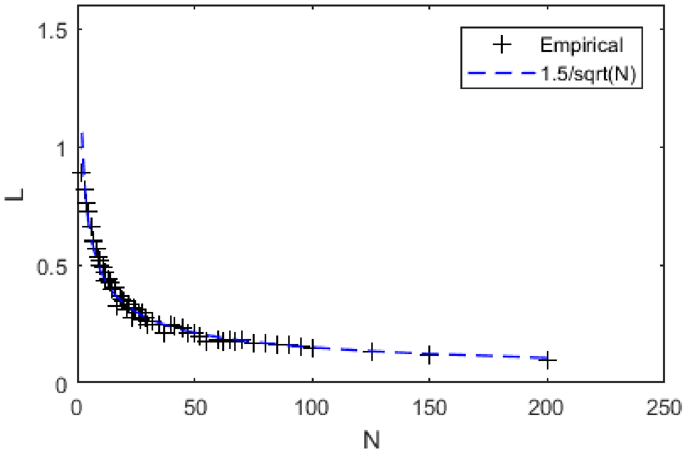
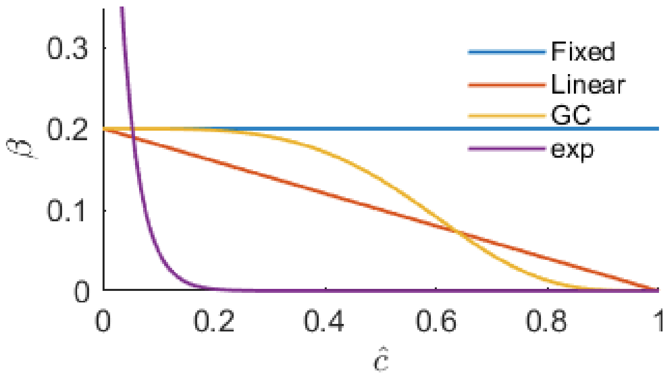


Figure 4 - Different methods to compute the penalty term. The blue line represents the fixed method, the red line is the linear method, the yellow line represents the 'GC' method, and the purple line represents the empirical exponential method.



3.1.2 An alternative method to compute the threshold

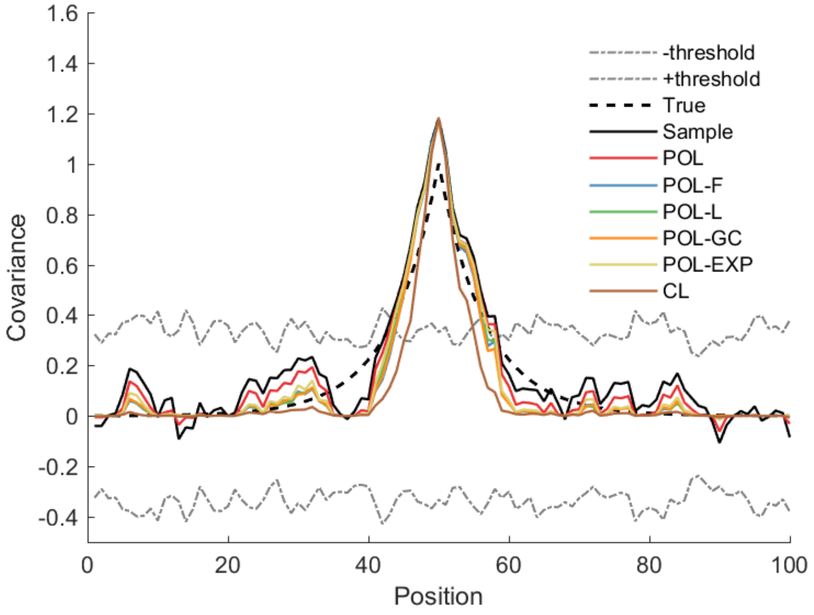
In real history matching applications, some uncertain model parameters cannot be grouped, making it awkward to obtain a single threshold θ. To avoid this problem, we propose an alternative rule: instead of estimating the sampling errors for a given group of parameters (shuffling a single time), we may shuffle the model parameters ensemble N_r times to obtain the noise level for a given model parameter m and predicted data o. Thus, we can generate an ensemble of sampling errors for each pth parameter and oth data as:

$$\tilde{\boldsymbol{\varepsilon}}_{p,o} = \left\{\varepsilon_{\tilde{p},o,r}\right\}_{r=1}^{N_r}. \tag{43}$$

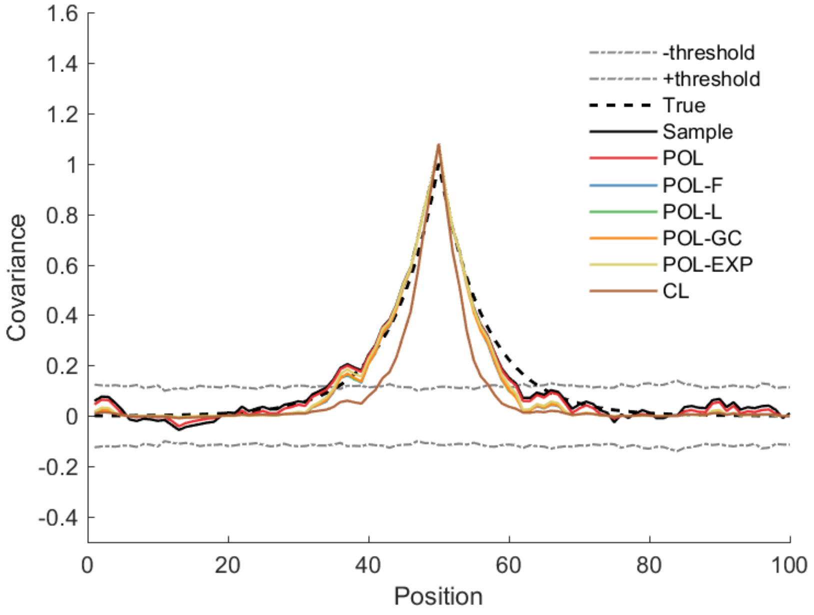
To illustrate the feasibility of this method, we generated one-dimensional unconditioned random field realizations for two different ensemble sizes ($N = 100$ and $N = 1000$), discretized in 100 gridblocks, using an exponential covariance function with a practical range equal to 20 gridblocks, and the prior mean and variance both equal to one. Next, we computed the sample covariance between each element of the model parameters and the element in the gridblock 50. For simplicity, we chose $N_r = N$ to estimate the noise levels.

Figure 5 shows the true covariance and tapered covariances for all methods described above. It is possible to verify, especially for the lower ensemble size, the inaccuracy of the standard pseudo-optimal localization to suppress spurious covariances. In addition, it is possible to verify that the spurious covariances observed in the sample covariances have a similar magnitude to the noise levels estimated using the random shuffle method. Also, it is possible to see that the POL methods with the random-shuffle-based penalty factor tend to suppress low magnitude spurious covariances, while avoiding excessive tapering as observed in the correlation-based localization (CL).

Figure 5 - Covariance values between model parameter on each gridblock and that gridblock 50, for two ensemble sizes ($N = 100$ and $N = 1000$). The black dashed lines represent the true covariance, black lines the sample covariance, gray lines the threshold obtained by the random shuffle method. The red lines represent the standard pseudo-optimal method, blue lines represent the pseudo-optimal method with POL-F, green lines represent the pseudo-optimal method with POL-L, orange lines represent the pseudo-optimal method with POL-GC, yellow lines represent the pseudo-optimal method with POL-EXP, and brown lines represent the tapered covariance with correlation-based localization.



(a) Covariance values for $N = 100$

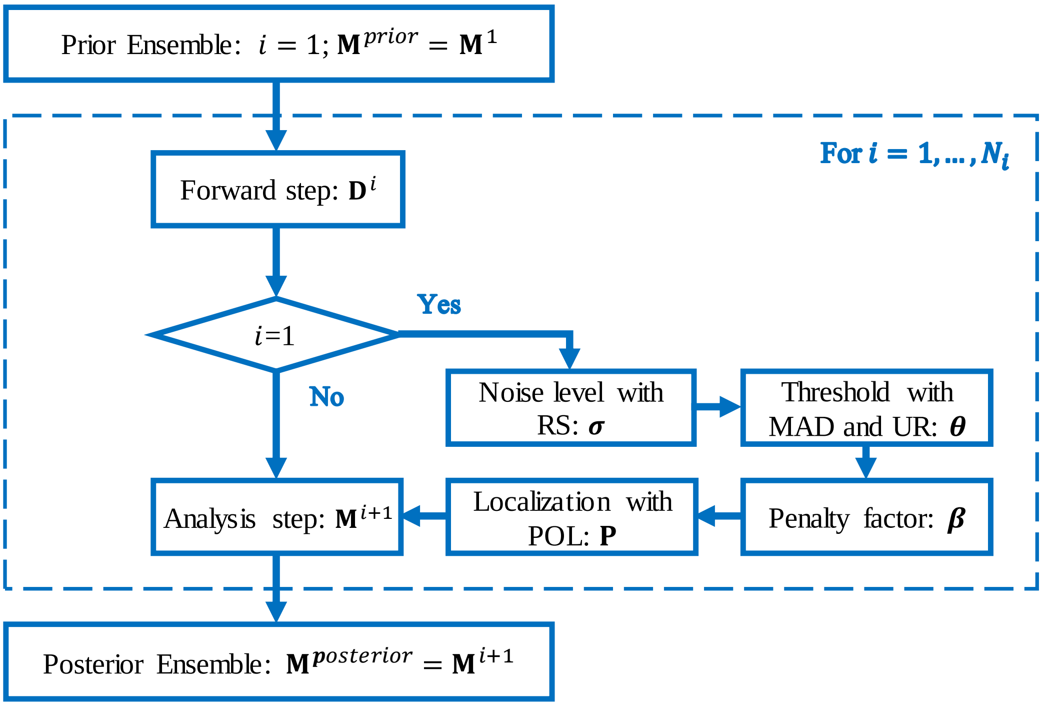


(b) Covariance values for $N = 1000$

3.1.3 IES with Random-Shuffle-Based Pseudo-Optimal Localization

In iterative ensemble smoother methods, intuitively it may seem reasonable to construct the localization matrix at every iteration step, since both model parameter and data spaces are updated. However, as mentioned by authors who applied adaptive localization methods, re-computing the localization matrix at every iteration showed similar or inferior results in comparison with the choice of constructing the localization matrix only at the first iteration (LUO; BHAKTA; NÆVDAL, 2018; LACERDA; EMERICK; PIRES, 2019). The reason for this is not completely understood, but a possible reason is the 'inbreeding' effect: the use of the same ensemble to compute the Kalman gain matrix and conduct the assimilation (HOUTEKAMER; MITCHELL, 1998; VAN LEEUWEN, 1999). Lacerda et al. (2019) pointed out that this effect is extended to the pseudo-optimal localization since we use ensemble covariances to compute the tapering values. Furthermore, computing the localization matrix at each assimilation step increases the computational time. Thus, in the present work, we only analyze the case where the localization matrix is computed from the prior ensemble. Finally, it is important to point out that the analysis steps in iterative ensemble smoothers with Kalman gain localization are generally computed row by row (EMERICK, 2016). Consequently, the Kalman gain matrix is never assembled in a practical implementation. For this reason, each entry of the localization matrix can be computed independently. The workflow of the pseudo-optimal localization with the random shuffle method can be seen in Figure 6.

Figure 6 - Workflow of the proposed localization methodology. Here, the localization matrix using the pseudo-optimal localization and the random shuffle method is computed only at the first iteration. After the construction of the localization matrix, the assimilation algorithm runs forward until it reaches the last iteration.


Prior Ensemble: $i = 1; \mathbf{M}^{prior} = \mathbf{M}^1$
For $i = 1, \dots, N_i$
Forward step: $\mathbf{D}^i$
i=1
Yes
No
Noise level with RS: $\boldsymbol{\sigma}$
Threshold with MAD and UR: $\boldsymbol{\theta}$
Analysis step: $\mathbf{M}^{i+1}$
Localization with POL: $\mathbf{P}$
Penalty factor: $\boldsymbol{\beta}$
Posterior Ensemble: $\mathbf{M}^{posterior} = \mathbf{M}^{i+1}$

3.2 GAN with R1 regularization and Adaptive Data Augmentation

Even with the WGAN with gradient penalty regularization, convergence is not always guaranteed (MESCHEDER; GEIGER; NOWOZIN, 2018). To address this, Mescheder et al. (2018) evaluated the use of zero-centered gradient penalties, suggesting improving the training convergence by penalizing the discriminator when it deviates from the Nash-equilibrium. This is done by the following regularization term:

$$R_1 = \frac{\gamma}{2}\mathbb{E}_{x\sim p_d}[(\|\nabla\mathcal{D}(x)\|^2)] \tag{44}$$

where γ is the R_1 regularization term.

Karras et al. (2019, 2020) applied the R_1 regularization in different datasets, and observed that the optimal value of γ is highly case-dependent. They found that a good starting point for the regularization term is $\gamma_0 = 0.0002n_p/b_s$, where n_p represents the number of pixels in the image and b_s represents the batch size.

However, this finding may not be applicable to GANs that generate reservoir realizations, which are typically represented as single-channel images, as shown later.

Table 1 summarizes the GAN losses used in this study. Although other GAN methods seem promising to generate facies realizations, we decided to evaluate only methods that do not change the discriminator/generator structure to perform a fair comparison. For instance, Boundary Equilibrium GAN (BEGAN), which uses an autoencoder as discriminator, being the loss the Wasserstein distance between autoencoding errors distribution (BERTHELOT; SCHUMM; METZ, 2017).

Table 1 – Loss functions used in this work.

	Discriminator Loss	Generator Loss
NS-GAN	$\mathcal{L}_{\mathcal{D}} = -\mathbb{E}_{x\sim p_d}\left[\log\left(\mathcal{D}(x)\right)\right] - \mathbb{E}_{\hat{x}\sim p_g}\left[\log\left(1-\mathcal{D}(\hat{x})\right)\right]$	$\mathcal{L}_{\mathcal{G}} = -\mathbb{E}_{\hat{x}\sim p_g}\left[\log\left(\mathcal{D}(\hat{x})\right)\right]$
NS-GAN-R1	$\mathcal{L}_{\mathcal{D}} = -\mathbb{E}_{x\sim p_d}\left[\log\left(\mathcal{D}(x)\right)\right] - \mathbb{E}_{\hat{x}\sim p_g}\left[\log\left(1-\mathcal{D}(\hat{x})\right)\right] + \frac{\gamma}{2}\mathbb{E}_{x\sim p_d}[\lVert\nabla\mathcal{D}(x)\rVert^2]$	$\mathcal{L}_{\mathcal{G}} = -\mathbb{E}_{\hat{x}\sim p_g}\left[\log\left(\mathcal{D}(\hat{x})\right)\right]$

3.2.1 Adaptive Data Augmentation applied to GAN

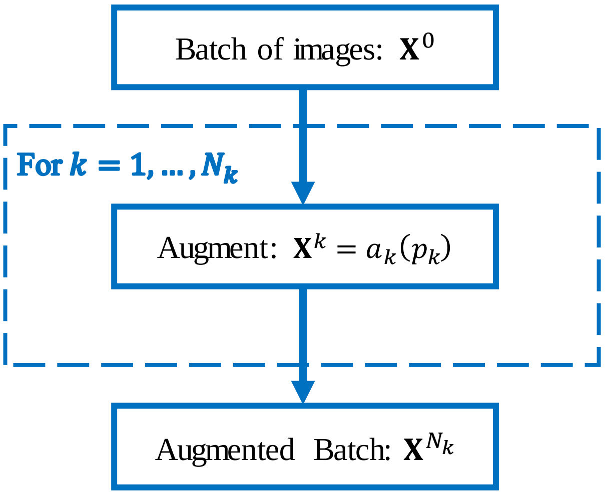
One of the main challenges regarding generative adversarial networks is the required dataset size to perform the training with reasonable quality. Regarding the data assimilation field, the available dataset size can make the GAN training prohibitive (CANCHUMUNI et al., 2021). Following the main idea that augmentations must be invertible to work with generative adversarial networks, to avoid the leaking behavior described by Karras et al. (2020). Here, we will call the operator containing all possible transformations as $\mathcal{A} = \{a\}_{k=1}^{K}$, where k is the set of all possible K augmentations. Furthermore, Karras et al. (2020) showed the conditions to ensure the non-leaking behavior: the main principle is when augmenting both real and generated samples using the same invertible (only if the operator is invertible) operator, the generator will try to match both distributions:

$$\mathcal{A}(\mathbf{x}) = \mathcal{A}(\mathbf{y}) \tag{45}$$

where $\mathbf{x}$ and $\mathbf{y}$ represents the original and generated distributions, respectively, and $\mathcal{A}$ denotes the augmentation operator. Karras et al. (2020) demonstrated that a wide range of augmentation techniques can be made invertible by applying them only with a probability p, ensuring that the leaking behavior is guaranteed to not happen. More details about the behavior of non-leaking augmentations might be found in Karras et al. (2020). Furthermore, a reasonable choice is to make the probability definition over the iterations adaptive: the required augmentation strength is not the same across the GAN training process since GAN performance deteriorates when the discriminator overfits.

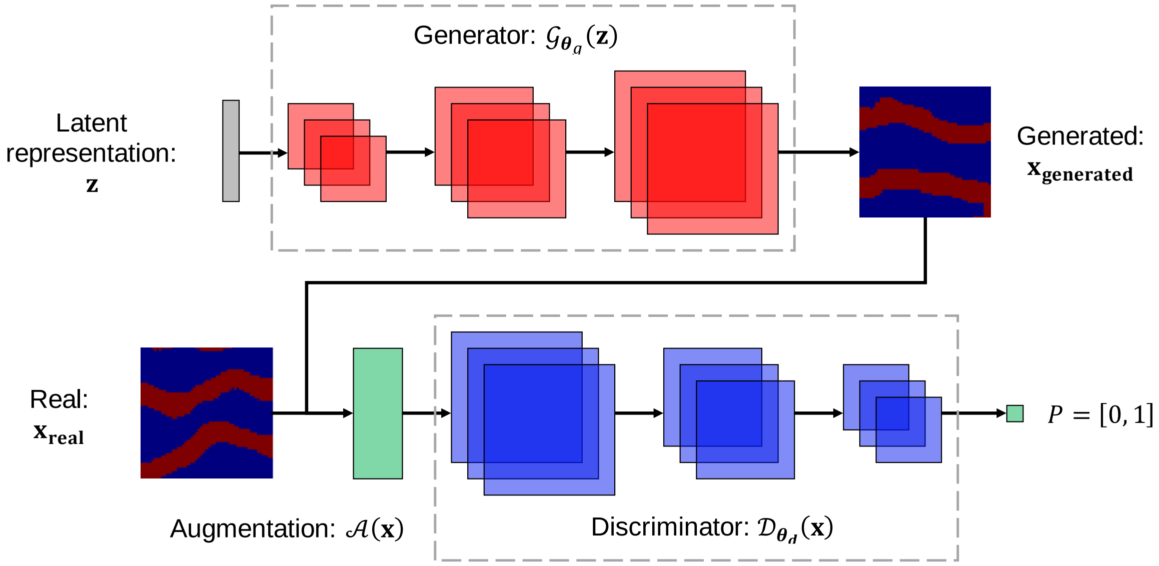
When the augmentation operator $\mathcal{A}$ represents a sequence of different augmentations, the resulting augmentation pipeline might be depicted as shown in Figure 7. For a given set of possible augmentations, each augmentation a_k is applied one at time with probability p_k. It is important to note that this probability value represents the fraction of the batch that will undergo augmentation. For instance, it represents the number of images that will be zoomed, rather than indicating the maximum zoom level in a random zoom augmentation operator.

Figure 7 – Augmentation pipeline.


Batch of images: $\mathbf{X}^0$
For $k = 1, \ldots, N_k$
Augment: $\mathbf{X}^k = a_k(p_k)$
Augmented Batch: $\mathbf{X}^{N_k}$

Furthermore, the adversarial training workflow with the invertible data augmentation pipeline might be represented as Figure 8.

Figure 8 – Generative adversarial training with adaptive data augmentation workflow.


Generator: $\mathcal{G}_{\boldsymbol{\theta}_g}(\mathbf{z})$
Latent representation: $\mathbf{z}$
Generated: $\mathbf{x}_{\mathbf{generated}}$
Real: $\mathbf{x}_{\mathbf{real}}$
Augmentation: $\mathcal{A}(\mathbf{x})$
Discriminator: $\mathcal{D}_{\boldsymbol{\theta}_d}(\mathbf{x})$
$P = [0, 1]$

In the following, we will introduce the set of individual augmentations that we evaluated. All possible augmentations might be divided into four categories: blitting transformations, geometric transformations, values transformations, and random noise. Blitting transformations move image pixels directly to another pixel of the image without any approximation. For example, x-flips and rotations by 90, 180 or 270 degrees in square images do not require any approximation or specific computation. Geometric transformations refer to the general geometric transformations, such as rotating an image by a given angle or zooming an image by a given factor. Value transformations change the pixel values, which can include modifying the brightness and contrast of an RGB image. Finally, random noise involves adding Gaussian noise to the image. Table 2 introduces each augmentation operation in details.

Table 2 – Augmentation pipeline (each applied only with probability p, except arbitrary rotations). Blitting and geometric transformations were based from Karras et al. (2020). Random numbers are generated for each image individually. In geometric pipeline, arbitrary rotation was split into two random rotations with probability $p_{rotation} = 1 - \sqrt{1-p}$, to make possible anisotropic zoom in both image directions.

Type	Operation	Description
Blitting	x-Flip	Flip images in $x\text{-direction}$
	90º rotations	Rotate image 90º degrees n times where $n \sim \mathcal{U}\{0, 3\}$
	Integer translation	Translate image $[\text{round}(t_x n_x), \text{round}(t_y n_y)]$ pixels, where $t_x, t_y \sim \mathcal{U}(-0.125, 0.125)$
Geometric	Isotropic zoom	Zoom image by a factor $[s_x, s_y]$, where $s_x, s_y \sim \mathcal{U}(-0.2, 0.2)$
	Arbitrary rotation (1)	Rotate image θ degrees, where $\theta \sim \mathcal{U}(0, 0.25)$, with $p_{rotation}$
	Anisotropic zoom	Zoom image by a factor $[s_x, s_y]$, where $s_x \sim \mathcal{U}(-0.04, 0.04)$ and $s_y \sim \mathcal{U}(-0.2, 0.2)$
	Arbitrary rotation (2)	Rotate image θ degrees, where $\theta \sim \mathcal{U}(0, 0.25)$, with $p_{rotation}$
	Fractional translation	Translate image $[t_x n_x, t_y n_y]$ pixels, where $t_x, t_y \sim \mathcal{U}(-0.125, 0.125)$
Value	Arbitrary sum	Add to the image v_s, where $v_s \sim \mathcal{U}(-0.2, 0.2)$
	Arbitrary product	Multiply image by v_m, where $v_m \sim \mathcal{U}(0.75, 1)$
	Facies inversion	Multiply image by -1
Noise	Gaussian noise	Add to the image a random field g_f, where $g_f \sim \mathcal{N}(0, \sigma)$ and $\sigma \sim \lvert\mathcal{N}(0, 0.1)\rvert$

Karras et al. (2020) demonstrated that the discriminator outputs during the training is a reliable indicator of overfitting. Therefore, a reasonable choice is to make the augmentation probability adaptive over the training, based on a desired discriminator behavior, by monitoring the discriminator outputs. To evaluate the overfitting, the accuracy of the discriminator on real images was demonstrated as a suitable metric:

$$r_t = \overline{\text{sgn}(D_{real})} \tag{46}$$

where $\text{sgn}(\cdot)$ represents the sign function and the overbar represents the mean. Here, r_t values have a range between -1 and 1, where a value of 1 represents complete overfitting (i.e., the discriminator identifies all real images as real). To maintain this accuracy within a target range, the probability is updated after a certain number of generator iterations, with the augmentation probability adjusted by a constant value determined by the following equation:

$$p_{t+1} = p_t + \frac{\text{sgn}\left(r_t - r_{target}\right)}{\ell} \tag{47}$$

where r_{target} is the desired target value and ℓ represents the integral gain that determines the velocity of change in the augmentation probability. In this work, we adopted a slightly different approach to update the augmentation probability. Instead, the original heuristic, we opted for a proportional-derivative (PD) controller (LIPTÁK, 2003, p. 109) to prevent the probability from exhibiting a high variance during the GAN training.

$$p_{t+1} = p_t + \frac{k_1\left(r_t - r_{target}\right) + k_2(r_t - r_{t-1})}{\ell} \tag{48}$$

where k_1 and k_2 represents the proportional and derivative coefficients, respectively.

3.2.2 Quality metrics

Evaluating the quality of a GAN generator and comparing different GANs can be a challenging task as it involves a combination of subjective evaluation and objective measurement. One aspect that requires subjective assessment is the individual assessment of the generated samples, while quantitative methods must be used to determine the similarity between statistics of the generated and training datasets, such as identifying mode collapse, where the generated images are identical or highly similar despite appearing realistic. To assess these two features in a single metric, Salimans et al. (2016) used the Inception model, a classifier network trained on the ImageNet dataset (DENG et al., 2009), to obtain the conditional label distribution of the generated samples. The authors argue that meaningful samples should produce a conditional label (output) distribution $p(y|x)$ with low-entropy. Additionally, the marginal probability density function should have high-entropy $p(y)$ when varied images are generated. In this setting, Salimans et al. (2016) proposed the Inception Score (IS), the computation of the Kullback–Leibler (KL) divergence between label distribution and the marginal distribution of the Inception v3 network (SZEGEDY et al., 2015). Although it presents reasonable results, the Inception Score (IS) has some limitations, such as not considering the statistics of the real dataset and comparing them with the generated ones. To address this issue, Heusel et al. (2018) proposed a new metric, called the Fréchet Inception Distance (FID), which overcomes this limitation. They achieved this by replacing the conditional output in the IS with the coding layer of the Inception Model, and then measuring the difference between the real and generated datasets by computing the Fréchet distance between the two distributions r^I and g^I obtained in the Inception coding layer:

$$\mathrm{FID}(r,g) = \left\|\mu_r^I - \mu_g^I\right\|_2^2 + Tr\left(C_r^I + C_g^I - 2\left(C_r^I C_g^I\right)^{\frac{1}{2}}\right), \tag{49}$$

where μ and C denote the mean and covariance of a given set, while Tr to the trace of the matrix. For further information about GAN metrics, reader is directed to Borji (2018). However, computing the FID metric to assess the quality of generated reservoir realizations poses challenges. Firstly, the Inception Model used to compute FID was trained on the ImageNet dataset, which consists of large quality (minimum size of 75x75) RGB color images and only includes natural images such as animals, food, and objects. As a result, the accuracy of FID in evaluating reservoir realizations may be limited. Additionally, the Inception model's high number of parameters (over 21 million) results in a substantial computational demand. An alternative approach to apply the Fréchet Inception Distance in our case is expand the image to three channels by repeating the values and scaling it to the appropriate size for input into the Inception Model. However, inspired by previous works that computed the FID on the MNIST dataset (BIŃKOWSKI et al., 2021; TRAN et al., 2021) we decided to replace the Inception Model to a single-channel reservoir classifier, designed to classify input images among several toy reservoirs. We call this method as Fréchet Reservoir Distance (FRD). Thus, the Fréchet distance is now computed between the two distributions r^R and g^R obtained in the Reservoir Classifier coding layer:

$$\mathrm{FRD}(r,g) = \left\|\mu_r^R - \mu_g^R\right\|_2^2 + Tr\left(C_r^R + C_g^R - 2\left(C_r^R C_g^R\right)^{\frac{1}{2}}\right). \tag{50}$$

We start the construction of the network architecture similarly to the Inception-v3 network (SZEGEDY et al., 2015), to get the maximum performance regarding the feature extraction task. Our network architecture comprises two primary parts: 1) feature extraction; 2) classification. The feature extraction part contains Inception blocks, including convolutional layers, max and average pooling layers, and concatenation layers. On the other hand, the classification part consists of fully connected layers and a Softmax layer, with a Dropout layer (SRIVASTAVA et al., 2014) before the fully connected layer. Figure 9 shows a summarized view of the Reservoir classifier architecture. When considering only the feature extraction block which will be used during the computation of the Fréchet Reservoir distance, the total number of parameters is equal to 83,806, whereas in the Inception-v3 network this number is equal to 21,802,784. This significantly increases the speed at which we can obtain the coding layer values. Additionally, we reduced the dimension of the last feature extraction layer from 2,048 to 96, which reduces the number of members required to compute the Fréchet Distance, since this distance is computed using the ensemble mean and covariances. Table 3 shows the Reservoir Classifier architecture in detail.

Figure 9 – Reservoir classifier architecture. The left and right dashed squares denote the feature extraction and classification blocks, respectively.

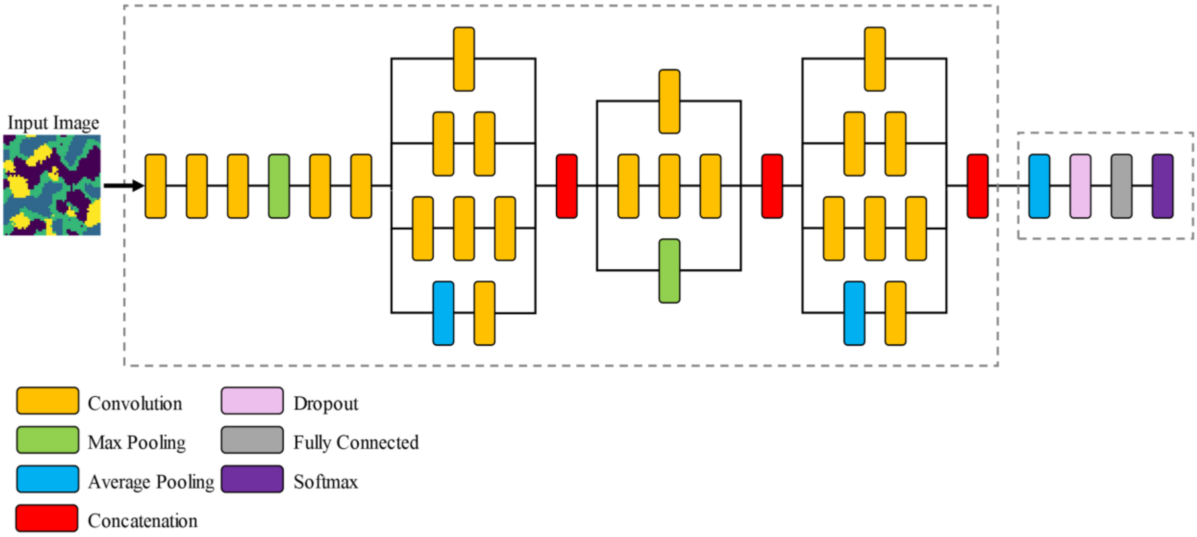


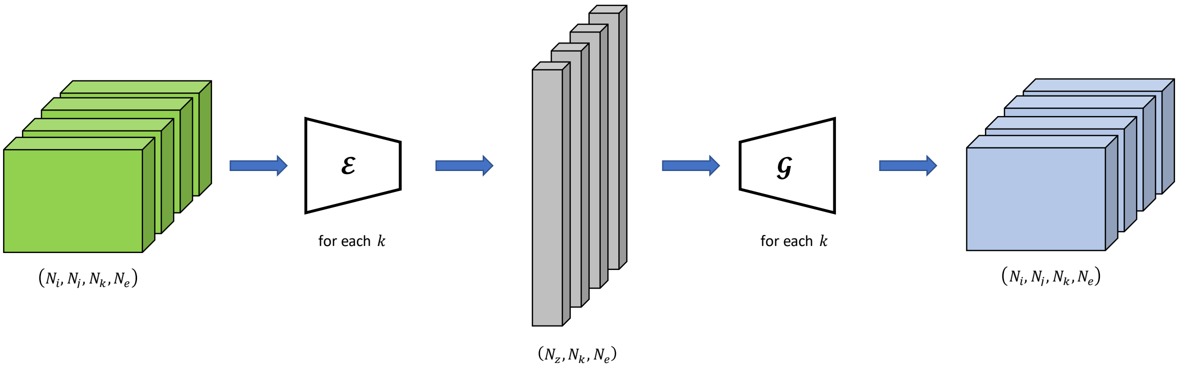
Table 3 – Reservoir Classifier network architecture. Here, nf refers to the number of base filters used, $nclasses$ the number of classes. Each convolutional block is composed by a sequence of a convolutional layer, batch normalization layer and Rectified Linear Unit (ReLU) activation function. Layers that lack connection information are linked to the layer in the row above them.

Layer		Options	Obs.
Input		Shape = $(48, 48, 1)$	Input image
Conv. block		Filters = $1 \times nf$, size = $(3, 3)$, pad = same	
Conv. block		Filters = $1 \times nf$, size = $(3, 3)$, pad = valid	
Conv. block		Filters = $2 \times nf$, size = $(3, 3)$, pad = valid	
Max Pooling		Size = $(3, 3)$, strides = $(2, 2)$	
Conv. block		Filters = $2.5 \times nf$, size = $(1,1)$, pad = valid	
Conv. block (C_5)		Filters = $6 \times nf$, size = $(3, 3)$, pad = valid	
Inception block 1	Avg. Pooling	Size = $(3, 3)$, strides = $(1, 1)$, pad = same	Connected to C_5
	Conv. block (B_I1_1)	Filters = $1/4 \times nf$, size = $(1, 1)$, pad = same	
	Conv. block (B_I1_2)	Filters = $2/4 \times nf$, size = $(1, 1)$, pad = same	Connected to C_5
	Conv. block	Filters = $1.5/4 \times nf$, size = $(1, 1)$, pad = same	Connected to C_5
	Conv. block (B_I1_3)	Filters = $2/4 \times nf$, size = $(5, 5)$, pad = same	
	Conv. block	Filters = $2/4 \times nf$, size = $(1, 1)$, pad = same	Connected to C_5
	Conv. block	Filters = $3/4 \times nf$, size = $(3, 3)$, pad = same	
	Conv. block (B_I1_4)	Filters = $3/4 \times nf$, size = $(3, 3)$, pad = same	
	Concatenate (I1_c)	B_I1_1, B_I1_2, B_I1_3, B_I1_4	
Inception block 2	Max Pooling (B_I2_1)	Size = $(3, 3)$, strides = $(2, 2)$	Connected to I1_c
	Conv. block	Filters = $2/4 \times nf$, size = $(1, 1)$, pad = same	Connected to I1_c
	Conv. block	Filters = $3/4 \times nf$, size = $(3, 3)$, pad = same	
	Conv. block (B_I2_2)	Filters = $3/4 \times nf$, size = $(3, 3)$, pad = valid	
	Conv. block (B_I2_3)	Filters = $3 \times nf$, size = $(3, 3)$, strides = $(2, 2)$ pad = valid	Connected to I1_c
	Concatenate (I2_c)	B_I2_1, B_I2_2, B_I2_3	
Inception block 3	Avg. Pooling	Size = $(3, 3)$, strides = $(1, 1)$, pad = same	Connected to I2_c
	Conv. block (B_I3_1)	Filters = $6/4 \times nf$, size = $(1, 1)$, pad = same	
	Conv. block (B_I3_2)	Filters = $6/4 \times nf$, size = $(1, 1)$, pad = same	Connected to I2_c
	Conv. block	Filters = $1 \times nf$, size = $(1, 1)$, pad = same	Connected to I2_c
	Conv. block (B_I3_3)	Filters = $6/4 \times nf$, size = $(4, 4)$, pad = same	
	Conv. block	Filters = $1 \times nf$, size = $(1, 1)$, pad = same	Connected to I2_c
	Conv. block	Filters = $6/4 \times nf$, size = $(1, 4)$, pad = same	
	Conv. block (B_I3_4)	Filters = $6/4 \times nf$, size = $(4, 1)$, pad = same	
	Concatenate (I3_c)	B_I3_1, B_I3_2, B_I3_3, B_I3_4	
Global Average Pooling		Output Shape = (96)	
Dropout		Rate = 0.3	
Fully-connected		Units = $8 \times nf$	
Softmax		Units = $nclasses$	

3.3 Two-step data assimilation in a three-dimensional carbonate reservoir

Parameterization in three-dimensional models remains a challenge in data assimilation due to the need to maintain vertical continuity between the layers when mapping to the Gaussian field. To address this issue, we propose using the initial ensemble of the three-dimensional reservoir to project its latent representations onto a two-dimensional GAN. This projection has been already made in a two-dimensional reservoir (CANCHUMUNI et al., 2021) by using an encoder-decoder scheme. However, its behavior in three-dimensional cases is not known. The rationale behind this approach is that data assimilation only introduces minor changes to the ensemble, and projecting onto a two-dimensional space may preserve the vertical continuity of the three-dimensional realizations. Here, we aim to assess the validity of this assumption. This process might be shown by using a generic operator which map a reservoir k –th layer to a respective latent representation $\mathcal{E}: \mathbf{m}_k \rightarrow \mathbf{z}_k$ (Figure 10). By using this operation, we project the initial three-dimensional ensemble $\left(N_i, N_j, N_k, N_e\right)$ to the latent space performing a layer-per-layer operation, resulting in a (N_z, N_k, N_e) latent space.

Figure 10 – Methodology to preserve vertical correlation during the data assimilation.



The problem of project real samples into a latent representation is known. Actually, there are two main methods to perform this projection: training an encoder-decoder network structure using the trained generator (with frozen weights) as decoder, and from a random initial latent sample, apply a gradient-based technique (ABDAL; QIN; WONKA, 2019). As the encoder-decoder structure naturally introduces a bias when projected samples that are not used during the training, the development of techniques to improve the gradient-based projection has been proposed, for instance, Stochastic Clipping (LIPTON; TRIPATHI, 2017), and perceptual losses (JOHNSON; ALAHI; FEI-FEI, 2016; ABDAL; QIN; WONKA, 2019), however, the projection of real samples onto the latent space stills an open problem.

The problem of project real samples in the latent space is clearly an ill-posed inverse problem. Thus, we propose an alternative: the application of an ensemble-based method to obtain the latent representation of a real sample. To our knowledge, this will be the first attempt to perform this optimization problem by using a gradient-free method.

3.3.1 Ensemble-based real samples projection into the latent space

From a pre-trained generator $\boldsymbol{\mathcal{G}}$, we want to find its respective latent representation $\hat{\mathbf{z}} \in \mathfrak{R}^{N_z}$ of a given real sample $\mathbf{m} = \hat{\mathbf{d}}_{\mathrm{r}} \in \mathfrak{R}^{N_{\hat{d}}}$ (this sample might be used during the network training or not). This ill-posed inverse problem might be defined as find the vector $\hat{\mathbf{z}}$ which satisfies:

$$\hat{\mathbf{d}}_{\mathrm{r}} = \boldsymbol{\mathcal{G}}(\hat{\mathbf{z}}). \tag{51}$$

Following the solution obtained in section 2.4, the ES-MDA solution might be obtained from:

$$\hat{\mathbf{z}}_{j,i+1} = \mathbf{z}_{j,i} + \mathbf{C}_{\mathrm{Z}\widehat{\mathrm{D}},i}\left(\mathbf{C}_{\widehat{\mathrm{D}}\widehat{\mathrm{D}},i} + \alpha_i \mathbf{C}_{\mathrm{M}}\right)^{-1}\left(\hat{\mathbf{d}}_{\mathrm{r}} - \hat{\mathbf{d}}_{j,i} - \sqrt{\alpha_i}\epsilon_j\right). \tag{52}$$

Using only the generator output to compute the innovation (difference between target sample and the ensemble of generated samples) does not provide consistent results. It happens because the pixel-per-pixel minimization might neglect some image details, for example low proportion facies. This raises the application of Perceptual losses to reduce the discrepancy between high-level features also (JOHNSON; ALAHI; FEI-FEI, 2016). In gradient-based methods a loss network is used to augment the optimization problem by computing the differences between the real and generated outputs at different layers of this loss network. Regarding ensemble-based methods, is straightforward to consider this high-level feature, by augmenting both data and observation ensembles. Denoting this loss network and its layers as $\boldsymbol{\mathcal{C}} = \{\boldsymbol{c}_\ell\}_{\ell=1}^{N_\ell}$, both data and observation vectors might be redefined as:

$$\hat{\mathbf{d}}_{\mathrm{r}} = [\mathbf{m}, \boldsymbol{c}_1(\mathbf{m}), \boldsymbol{c}_2(\mathbf{m}), \dots, \boldsymbol{c}_\ell(\mathbf{m})]^{\mathrm{T}}, \tag{53}$$

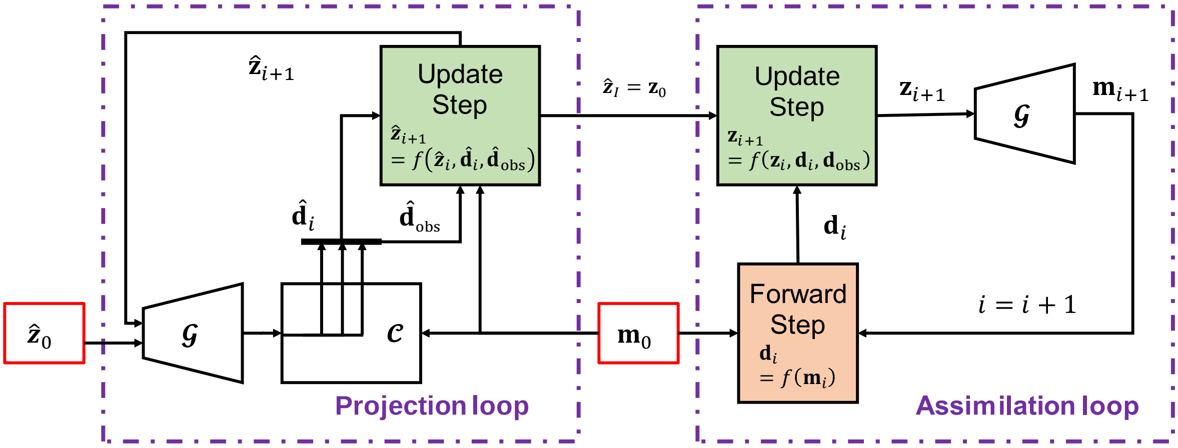
and,

$$\hat{\mathbf{d}}_j = \left[\boldsymbol{\mathcal{G}}(\mathbf{z}_j), \boldsymbol{c}_1\left(\boldsymbol{\mathcal{G}}(\mathbf{z}_j)\right), \boldsymbol{c}_2\left(\boldsymbol{\mathcal{G}}(\mathbf{z}_j)\right), \dots, \boldsymbol{c}_\ell\left(\boldsymbol{\mathcal{G}}(\mathbf{z}_j)\right)\right]^{\mathrm{T}}, \tag{54}$$

where $\boldsymbol{c}_\ell$ represents the flattened output of a given layer of the loss network used to extract high-level features. Usually, the VGG network (SIMONYAN; ZISSERMAN, 2014) is used as loss function to compute the perceptual losses in gradient-based applications (KARRAS et al., 2019). However, a more effective option in our problem is to employ the Reservoir Classifier network introduced before as this loss function.

Hence, the entire data assimilation workflow can be divided into two main steps. Starting from an initial ensemble $\mathbf{m}_0$, a projection loop is carried out to obtain the latent representations $\hat{\mathbf{z}}$ of this ensemble with respect to a pre-trained generator $\mathcal{G}$. In this step, high-level features are also assimilated using the reservoir classifier network $\mathcal{C}$ as loss function. Subsequently, the obtained latent ensemble is used as the initial ensemble for classical history matching to minimize the discrepancy between the simulated output and the measurements. Figure 11 illustrates the complete workflow. Here, it is important to note that the projection loop is applied to each layer for all reservoir properties that need to be parameterized.

Figure 11 – Two-step data assimilation workflow.


$\hat{\mathbf{z}}_{i+1}$
Update Step
$\hat{\mathbf{z}}_{i+1} = f(\hat{\mathbf{z}}_i, \hat{\mathbf{d}}_i, \hat{\mathbf{d}}_{\text{obs}})$
$\hat{\mathbf{z}}_I = \mathbf{z}_0$
Update Step
$\mathbf{z}_{i+1} = f(\mathbf{z}_i, \mathbf{d}_i, \mathbf{d}_{\text{obs}})$
$\mathbf{z}_{i+1}$
$\mathcal{G}$
$\mathbf{m}_{i+1}$
$\hat{\mathbf{d}}_i$
$\hat{\mathbf{d}}_{\text{obs}}$
$\mathbf{d}_i$
Forward Step
$\mathbf{d}_i = f(\mathbf{m}_i)$
$i = i + 1$
$\hat{\mathbf{z}}_0$
$\mathcal{G}$
$\mathcal{C}$
$\mathbf{m}_0$
Projection loop
Assimilation loop

CHAPTER 4 - CASE STUDIES

This section introduces the case studies used in the first two methodology sections (localization and parameterization) evaluated individually, the dataset used during the Reservoir Classifier training, and also the benchmark used in the complete workflow combining both adaptive localization and parameterization with GAN-ADA. Furthermore, this section introduces all settings used in all experiments.

4.1 Adaptive Localization

4.1.1 Simple model: ¼ five-spot

The case study presented in this work evaluates the performance of different methods to compute the penalty factor in a simple model, which represents the $1/4$ five-spot with one producer and one injector well. The reservoir model dimension is $51 \times 51 \times 1$ with 2601 active blocks, and the uncertain parameters are the horizontal log-permeabilities $\log(k)$ distributed on active gridblocks ($N_m = 2601$). We have monthly measurements for almost 10 years of production history (119 months) of oil and water production and water injection rates ($N_d = 357$). The prior ensemble of model parameters was generated using an isotropic exponential covariance function with a practical range equal to 20 gridblocks, the ensemble mean equal to 5, and the ensemble variance equal to 1, respectively. The reference model was generated in the same way as the prior ensemble, while white Gaussian noise with errors equal to 15% of the production data was added to the true data of the reference model to obtain noisy measurements.

In the experiments, we generated an initial large ensemble with 5000 members and defined it as the reference case. Then, we defined two distinct cases to evaluate the performance of data assimilation: the ensemble size $N = 50$ and $N = 200$, with a total number of iteration steps $N_i = 8$ for both cases. To mimic the setting in real field applications, we adopt a relatively small ensemble size to evaluate the behavior of the proposed adaptive localization scheme. We also choose a larger ensemble size $N = 200$ to verify how this method behaves as the ensemble size increases. In all the experiments below, initial ensembles consist of N members randomly drawn from the ensemble with $N = 5000$, and we repeated the experiments 10 times in total (resulting in 10 assimilated ensembles for each method).

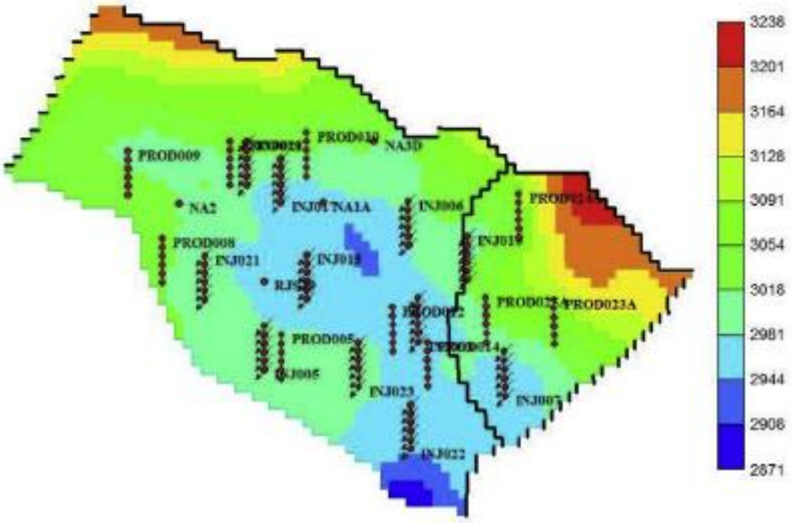
Data assimilation in all experiments was performed using the ES-MDA with an inflation factor equals to the total number of iterations for all iterations $\{N_i\}_{i=1}^{N_{iter}}$. In the analysis step, the inversion of the Kalman gain matrix was performed using the 'subspace inversion' method (EVENSEN, 2004; EMERICK, 2016) with 99% of singular values retained. Here, we evaluated the performance of different localization methods and classified them using: NO LOC: assimilation without using any localization method; POL: assimilation using the pseudo-optimal localization in its standard form; POL-F: pseudo-optimal localization with fixed penalty; POL-L: pseudo-optimal localization with linear penalty; POL-GC: pseudo-optimal localization with GC penalty; and POL-EXP: pseudo-optimal localization with the empirical exponential penalty.

4.1.2 UNISIM-I-H

Next, we evaluate the performance of the methods described earlier in the benchmark UNISIM-I-H (MASCHIO et al., 2015) for history matching (Figure 12). More information about the UNISIM-I benchmark can be found in Avansi and Schiozer (2015). The reservoir model is composed of a corner point grid (81x58x20), with 38,466 active gridblocks and one main fault separating the entire reservoir in two sections. The benchmark has both local and non-local uncertain parameters. In terms of local parameters, the prior ensemble consists of equiprobable realizations of petrophysical porosities on gridblocks, including net-to-gross ratio and permeabilities in three orthogonal directions. Meanwhile, non-local uncertain properties contain the water-oil contact of the east sector, reservoir rock compressibility, vertical permeability multiplier, and two parameters (Corey exponent and maximum water relative permeability) used in Corey function which models the relative permeability curve of the reservoir.

In this history matching problem, production wells are controlled by liquid rates, and injector wells control is performed by injection rates. The benchmark contains 11 years of monthly measurements for all production and injection wells. Oil production rate, water production rate, and water injection rate have measurement errors equal to 10% of the data value (minimum of 1 m³/d). Furthermore, bottom-hole pressures for both producers and injectors have measurement errors equal to 10 kgf/cm². To evaluate the performance of the methods described before, data assimilation was performed using an ensemble size equal to $N = 200,\ 350, or\ 500$, and in the assimilation scheme, we defined four iterations with equal inflation factors at all iteration steps. The total number of production data is $N_d = 5{,}552$ and the total number of model parameters is $N_m = 192{,}335$. Kalman gain matrix was inverted using the 'subspace inversion' method describe above, retaining 99% of the energy of the singular values (EVENSEN, 2004; EMERICK, 2016). For the case with distance-based localization, the critical length used was equal to 2000 meters, following other works that applied distance-based localization to UNISIM-I-H benchmark (SILVA et al., 2017; EMERICK, 2018; RANAZZI; SAMPAIO, 2019a).

Figure 12 – First layer grid top of UNISIM-I-H benchmark (wells projected).



Source: Ranazzi and Sampaio (2019a).

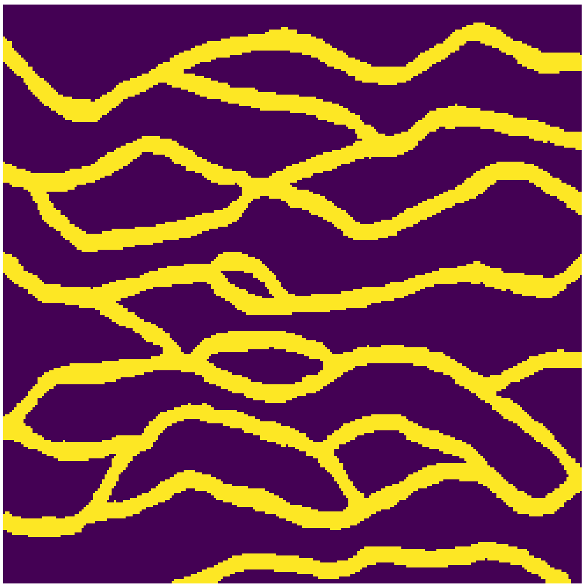
4.2 Parameterization

4.2.1 Reservoir classifier and Fréchet Reservoir Distance

To construct the dataset used in the Reservoir Classifier network, we considered two distinct cases, to make the classifier more able to generalize the feature extraction that will be used in the computation of the Fréchet Distance. Firstly, we considered realizations representing property values, where the variables arise from a continuous distribution. The second type of reservoir considered realizations representing facies values, where the variables are originated from an integer distribution representing the facies indicator.

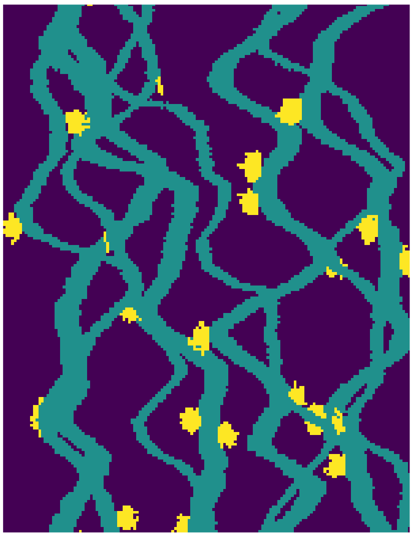
To build the training dataset containing facies, we used different training images found in the literature: the two-facies channelized training image from Caers and Zhang (2004); the “Stanford 5” three facies training image and the four facies training image (REMY; BOUCHER; WU, 2009, cap. 8); the four facies training image from Remy et al. (2009, cap. 8). Figure 13, Figure 14, Figure 15 show the Two facies channelized, “Stanford 5” three facies, and Four facies training images, respectively. For each training image, random realizations were generated with the Single Normal Equation Simulation Algorithm (STRÉBELLE, 2000, 2002), by using the Stanford Geostatistical Modeling Software - SGeMS (REMY; BOUCHER; WU, 2009), varying the angle rotation and affinity ratios in x and y directions. Additionally, a straightforward algorithm was developed to generate realizations consisting of property values. For a single realization, a Gaussian random field was generated, then, an exponential random field was generated for each facies using specific mean, variance, correlation length (or range), anisotropy, and angle parameters. Finally, the random fields were combined by truncating the Gaussian field based on a given facies proportion.

Figure 13 – Two facies channelized 250×250 training image.



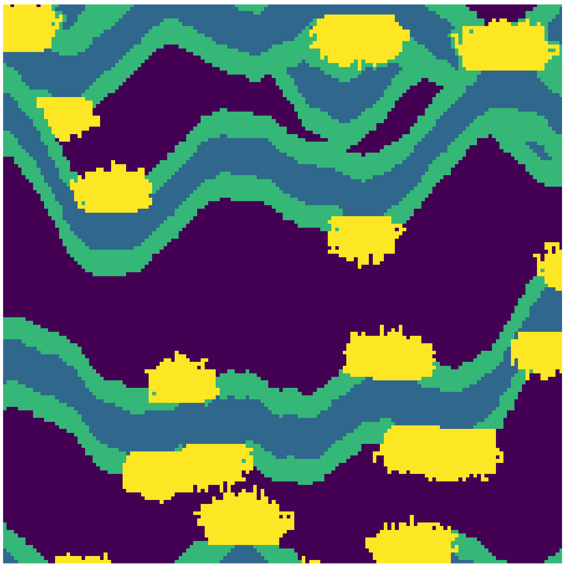
Source: Caers and Zhang (2004).

Figure 14 – “Stanford 5” three facies 195×150 training image.



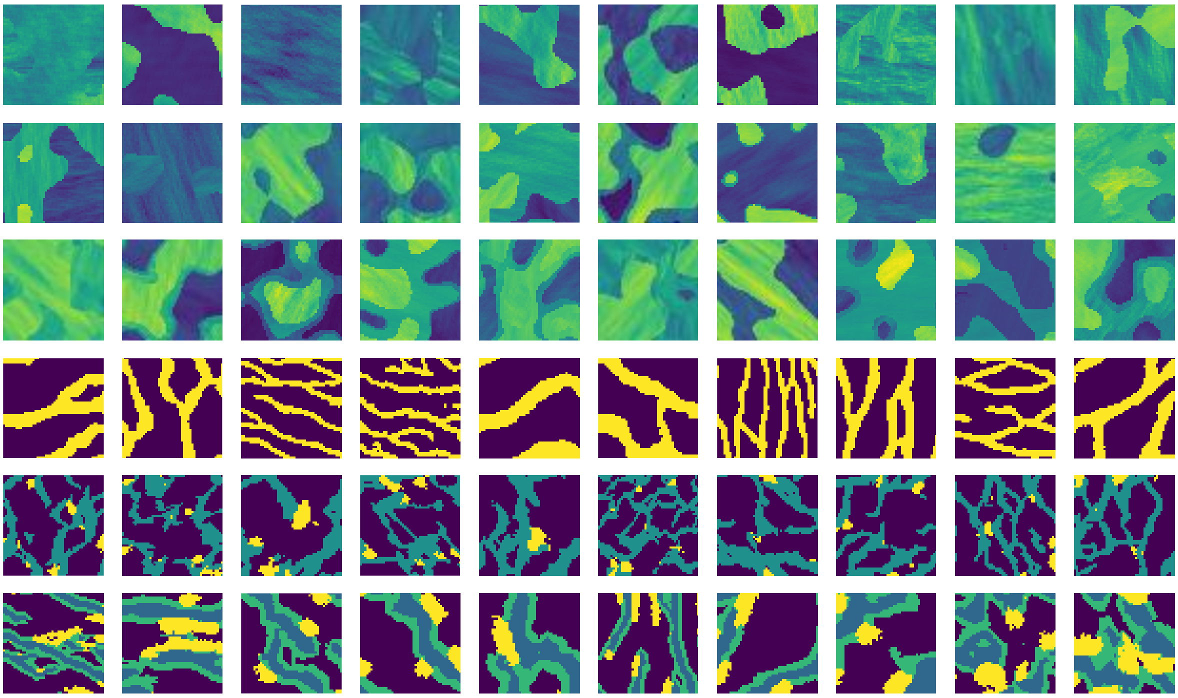
Source: REMY; BOUCHER; WU (2009).

Figure 15 – Four facies $150 \ \times 150$ training image.



Source: REMY; BOUCHER; WU (2009).

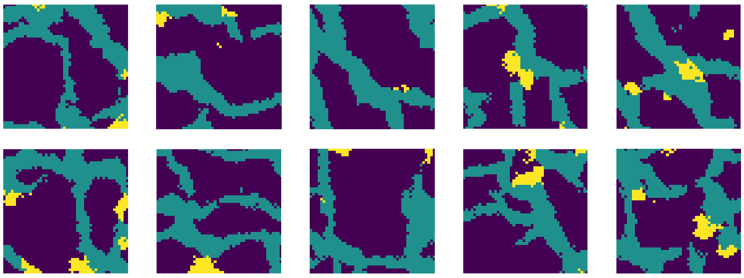
Figure 16 – Examples of realizations used during the Reservoir Classifier Network training. Each row represents a different reservoir category, containing continuous (first three rows) and integer (last three rows) values, representing properties and facies realizations, respectively.



4.2.2 GAN experiments

We evaluated our method in the "Stanford 5" three-facies from (REMY; BOUCHER; WU, 2009). The dataset contains a total of 80,000 realizations, Figure 17 shows some random realizations from the training dataset.

Figure 17 - Random realizations from the training dataset.



The generative adversarial network, containing the generator and the discriminator was build following the architecture of Table 3. In this work, all models have a base filter number equal to $n_f = 32$. Although more complex networks can give better results (with the cost of required computational power and available dataset size), we focused on the improvements of the regularization and augmentation applications. Here, upsampling is performed through the combination of resize-convolution layers, to avoid checkerboard artifacts (ODENA; DUMOULIN; OLAH, 2016). In the generator, all layers (except the last one) are followed by a batch normalization layer as common practice during GAN experiments (YANG; JIN; XU, 2020). This is also valid for the discriminator only with the case without regularization, where the batch normalization is also skipped after the first convolution. Batch normalization changes the optimization problem, in such way that discriminator regularization will be infeasible since it compute gradients with respect to sample inputs (GULRAJANI et al., 2017). Thus, we follow another works with discriminator regularization application and do not use any batch normalization when it is applied (GULRAJANI et al., 2017; KARRAS et al., 2019; KARRAS; LAINE; AILA, 2018). Table 4 shows the GAN discriminator and generator architectures in detail.

All models were optimized with Adam optimization algorithm (KINGMA; BA, 2014) with a learning rate of $l_r = 0.0001$ for both generator and discriminator, as well as $\beta_1 = 0.5$ and $\beta_2 = 0.9$. The batch size for all experiments was set to 32, and all networks were trained with 150,000 iterations.

Table 4 – Generative Adversarial Network generator and discriminator architectures. Here, nf refers to the number of base filters used, n_i and n_j the reservoir size in $i-$ and $j-$directions, respectively.

Layer	Options	Obs.
Generator		
Input	Shape = (512)	Input latent vector
Fully-connected	Units = $((n_i/8) \times (n_j/8) \times 8 \times nf)$	
Reshape	Shape = $(n_i/8, n_j/8, 8 \times nf)$	
Convolution 2D	Filters = $8 \times nf$, size = $(3,3)$, Shape = $(n_i/8, n_j/8, 8 \times nf)$	
Upsampling block	Filters = $4 \times nf$, size = $(3,3)$, Shape = $(n_i/4, n_j/4, 4 \times nf)$	
Upsampling block	Filters = $2 \times nf$, size = $(3,3)$, Shape = $(n_i/2, n_j/2, 2 \times nf)$	
Upsampling block	Filters = $1 \times nf$, size = $(3,3)$, Shape = $(n_i/1, n_j/1, 1 \times nf)$	
Convolution 2D	Filters = 1, size = $(5,5)$, Shape = $(n_i, n_j, 1)$	Activation = *tanh*
Discriminator		
Input	Shape = $(n_i, n_j, 1)$	Input image
Convolution 2D	Filters = $1 \times nf$, size = $(5,5)$, Shape = $(n_i, n_j, 1)$	
Downsampling block	Filters = $2 \times nf$, size = $(3,3)$, Shape = $(n_i/2, n_j/2, 2 \times nf)$	
Downsampling block	Filters = $4 \times nf$, size = $(3,3)$, Shape = $(n_i/4, n_j/4, 4 \times nf)$	
Downsampling block	Filters = $8 \times nf$, size = $(3,3)$, Shape = $(n_i/8, n_j/8, 8 \times nf)$	
Fully-connected	Units = (1)	Activation = linear

4.3 Two-step data assimilation applied in the benchmark UNISIM-II-H

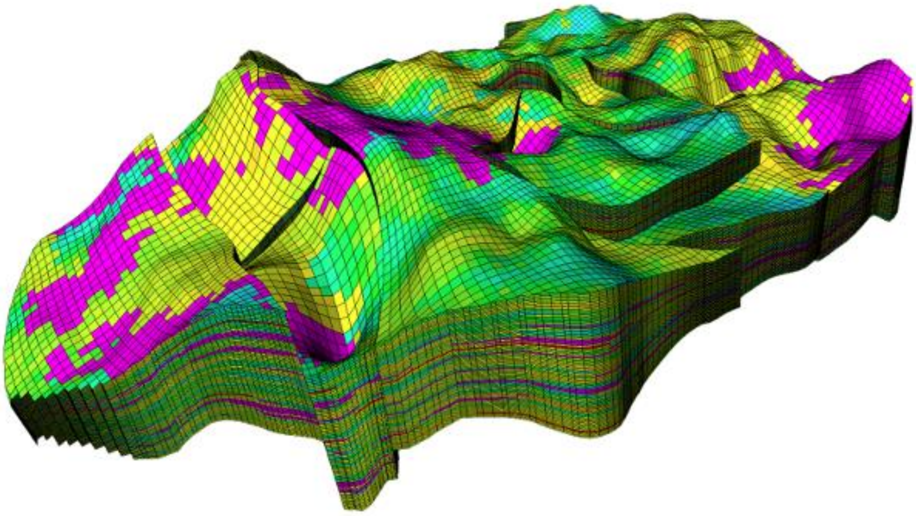
The final case study involves the data assimilation of the large-scale three-dimensional carbonate reservoir UNISIM-II (CORREIA et al., 2015; MASCHIO; SANTOS; CORREIA, 2018), which is a significant example of reservoir with carbonate features. This reservoir was created using data from Brazilian pre-salt reservoirs and the Ghawar giant field, and it is represented by a black-oil model with $46 \times 69 \times 30$ gridblocks (around 65,000 active cells) with dual-porosity dual-permeability (DPDP) modelling. The benchmark contains a significant number of uncertain parameters, with many exhibiting non-Gaussian behavior. To enable data assimilation using an ensemble-based method, it is necessary to apply a parameterization technique. Table 5 illustrates the uncertain field properties of the UNISIM-II-H benchmark, including those that will require parameterization.

Table 5 – UNISIM-II-H field properties. Here, NTG denotes Net-to-gross ratio.

Property		Parameterization
Porosity	Matrix	No
	Fracture	Yes
Permeability	Matrix	Yes
	Fracture	Yes
NTG		Yes
Fracture Spacing		Yes

Following Correia et al. (2015) and Maschio and Schiozer (2019), the rock type is defined based on a cutoff procedure with respect to the matrix permeability, where values higher than 800 mD are assigned as Super-K, following a different relative permeability curve (CORREIA; SCHIOZER, 2018). Figure 18 illustrates the matrix permeability of the three-dimensional reservoir.

Figure 18 – UNISIM-II benchmark (matrix permeability).



Source: Correia et al. (2015).

This reservoir contains 11 producer and 9 injector wells, with 9 years of monthly measurements for all wells. The measurement errors used to build the covariance error measurements matrix was defined as 10% of the measurement value (with a minimum value equal to 1). The total number of significant measurements (excluding data of closed wells) is $N_d = 2{,}822$. The total number of parameters is $N_m = 151{,}578$.

The assimilation was performed using the ES-MDA with four iterations with constant inflation factors ($N_i = \alpha_i = 4$), and ensemble size equal to $N = 50$ (due to the available computational resources). In addition, adaptive localization was employed using the pseudo-optimal localization with random-shuffle and Gaspari and Cohn tapering (POL-GC) for rock compressibility and relative permeability curves parameters. The threshold value for non-local properties was determined by shuffling the ensemble N_r times.

In our approach for projecting real samples onto the latent space, we found data downsampling the real sample reduces the inverse problem without losing performance. Thus, during the construction of the $\hat{\mathbf{d}}_{\mathrm{r}}$ vector in (53) real samples $\mathbf{m}$ are downsampled by a factor of $1/4$ before the flattening by simple resizing the input image. Additionally, we constructed the covariance matrix of the measurement errors considering only non-correlated measurements (diagonal matrix), with values of 0.01 for image data and 0.05 for feature data. This choice was motivated by the observation that the image data occupy a space between $\boldsymbol{\mathcal{G}}(\mathbf{z}_j) \in [-1, 1]$, while the feature data lie in a space between $\boldsymbol{c}_1\left(\boldsymbol{\mathcal{G}}(\mathbf{z}_j)\right) \in [0, \infty]$, as the features are obtained through rectified linear units (ReLU) activation functions in the Reservoir Classifier Network.

CHAPTER 5 - RESULTS AND DISCUSSIONS

5.1 Localization

5.1.1 Simple model: ¼ five-spot

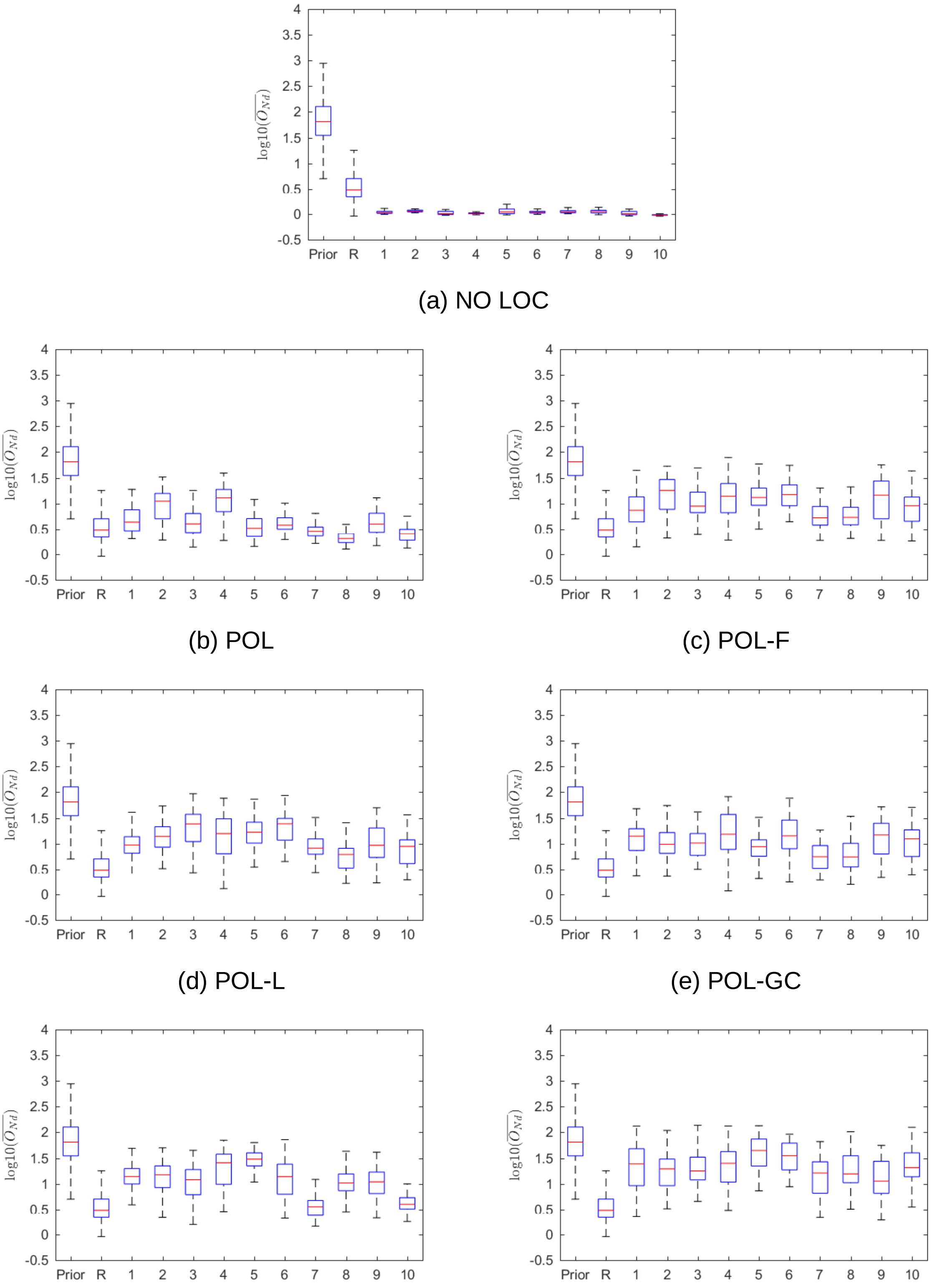
Table 6 shows the values of objective functions in the ¼ five-spot case study. Analyzing the results, it is possible to verify that the case without localization results in data overfitting with $\overline{O_{N_d}}$ (30) values than the reference case for both ensemble sizes, large model changes with high values for $\overline{O_{N_m}}$ (32) and ensemble collapse with ANV (33) almost zero. Similar to what was found by Lacerda et al. (2019), the pseudo-optimal localization results in an improvement of data assimilation performance, reducing ensemble collapse and data overfitting with both higher ANV and $\overline{O_{N_d}}$ values. Furthermore, inspecting the performance of the four methods proposed in this work, one can verify that all these proposed methods achieved improved assimilation results in terms of $\overline{O_{N_m}}$ and ANV, when comparing them to the corresponding results in the reference case. Even with somewhat higher $\overline{O_{N_d}}$ values, all posterior ensembles obtained by these four methods exhibited reasonably good data match, close to that of the reference case. It is important to note that the improvements mentioned above have been observed even with a relatively small ensemble size of 50 members. Figure 19 and Figure 20 show the boxplots of 10 ensembles for all methods analyzed in this work.

Table 6 – Results in the ¼ five-spot model. Here, $\overline{O_{N_d}}$ is the averaged normalized data-mismatch defined in equation (30), $\overline{O_{N_m}}$ is the averaged model-mismatch defined in equation (32) and ANV is the averaged normalized variance defined in equation (33). The results are reported in the form of mean ± standard deviation over 10 ensembles of each experiment, except for the reference case ($N = 5000$).

		$\overline{O_{N_d}}$	$\overline{O_{N_m}}$	ANV
	$N = 5000$	8.9935	0.6272	0.7450
$N = 50$	Prior	392.63 ± 148.32	-	-
	NO LOC	01.0827 ± 0.0590	1.3754 ± 0.1252	0.0590 ± 0.0080
	POL	05.7002 ± 3.7025	1.1721 ± 0.1297	0.3871 ± 0.0686
	POL-F	14.1178 ± 4.7794	0.6710 ± 0.0502	0.5947 ± 0.0401
	POL-L	16.4024 ± 6.8490	0.7249 ± 0.0669	0.6000 ± 0.0485
	POL-GC	14.2850 ± 4.8839	0.6937 ± 0.0541	0.5998 ± 0.0451
	POL-EXP	16.0327 ± 8.1830	0.9642 ± 0.1117	0.5722 ± 0.0770
	CL	29.9590 ± 9.2037	0.2890 ± 0.0543	0.8750 ± 0.0434
$N = 200$	Prior	401.16 ± 92.50	-	-
	NO LOC	02.6787 ± 0.3506	0.9716 ± 0.0278	0.3665 ± 0.0112
	POL	04.6470 ± 1.1783	0.8602 ± 0.0222	0.5889 ± 0.0282
	POL-F	11.8339 ± 3.4949	0.5465 ± 0.0221	0.7268 ± 0.0257
	POL-L	10.5968 ± 2.5628	0.5588 ± 0.0222	0.7212 ± 0.0235
	POL-GC	11.2693 ± 2.6187	0.5510 ± 0.0286	0.7236 ± 0.0240
	POL-EXP	10.5702 ± 4.0380	0.6600 ± 0.0288	0.7013 ± 0.0299
	CL	13.7150 ± 2.3787	0.3956 ± 0.0270	0.7365 ± 0.0237

It is possible to figure out the possible causes of the improvements of the methods proposed here by analyzing the localization matrices produced by these methods. For illustration, we generated localization matrices using the first n members of the reference case at the first iteration. Figure 21 and Figure 22 show the localization matrices using the first 50 members on Days 31 and 456, respectively; whereas Figure 23 and Figure 24 present the localization matrices using the first 100 ensemble members on Days 31 and 456, respectively. In addition, we generated localization matrices for both ensemble sizes using the covariances obtained from the reference ensemble with 5,000 members. Analyzing these figures, it is evident that the pseudo-optimal localization method can handle the time-lapse effect of observations, since it generates different tapering values adaptive to different simulation times. Furthermore, the proposed methods also produce different tapering values adaptive to different production data types (not reported here for conciseness). For comparison, one can see that the localization matrices generated by POL are highly affected by the sampling errors, deviating from the reference localization matrix estimated from the covariances of the reference case. In contrast, the introduction of the penalty term into equation (37) was able to reduce the effect of spurious correlations, even for the relatively small ensemble size with 50 members.

Figure 19 - Boxplots of log-data-mismatch values for posterior ensembles obtained in all experiments with $N = 50$. Here, the horizontal axis represents the index of relevant ensembles, with the labels ‘Prior’ and ‘R’ referring to the prior and posterior ensembles in the reference case ($N = 5000$).



(a) NO LOC

(b) POL

(c) POL-F

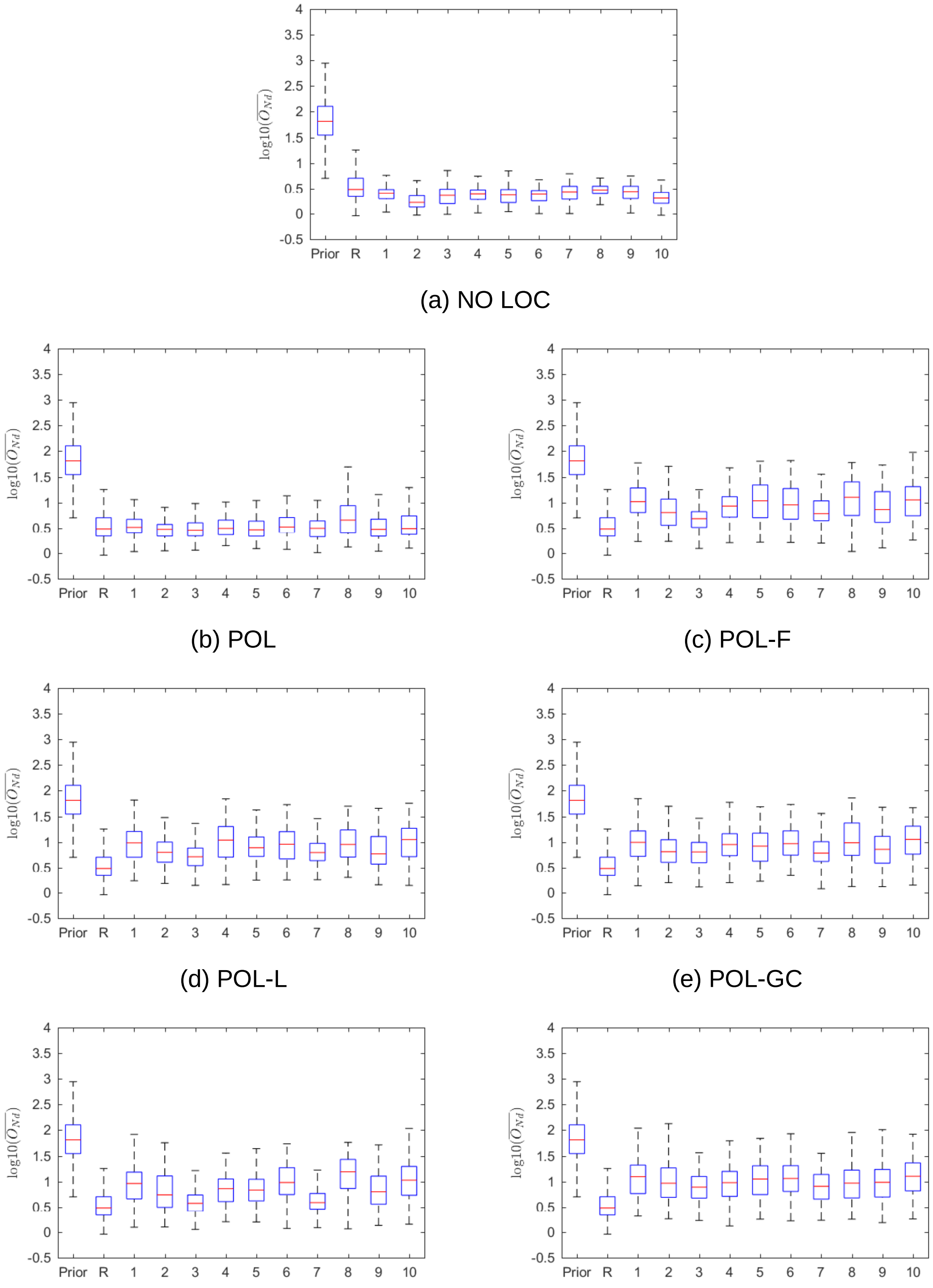
(d) POL-L

(e) POL-GC

(f) POL-EXP

(g) CL

Figure 20 - As in Figure 19, but with the ensemble size changed to $N = 200$ in the experiments.



(a) NO LOC

(b) POL

(c) POL-F

(d) POL-L

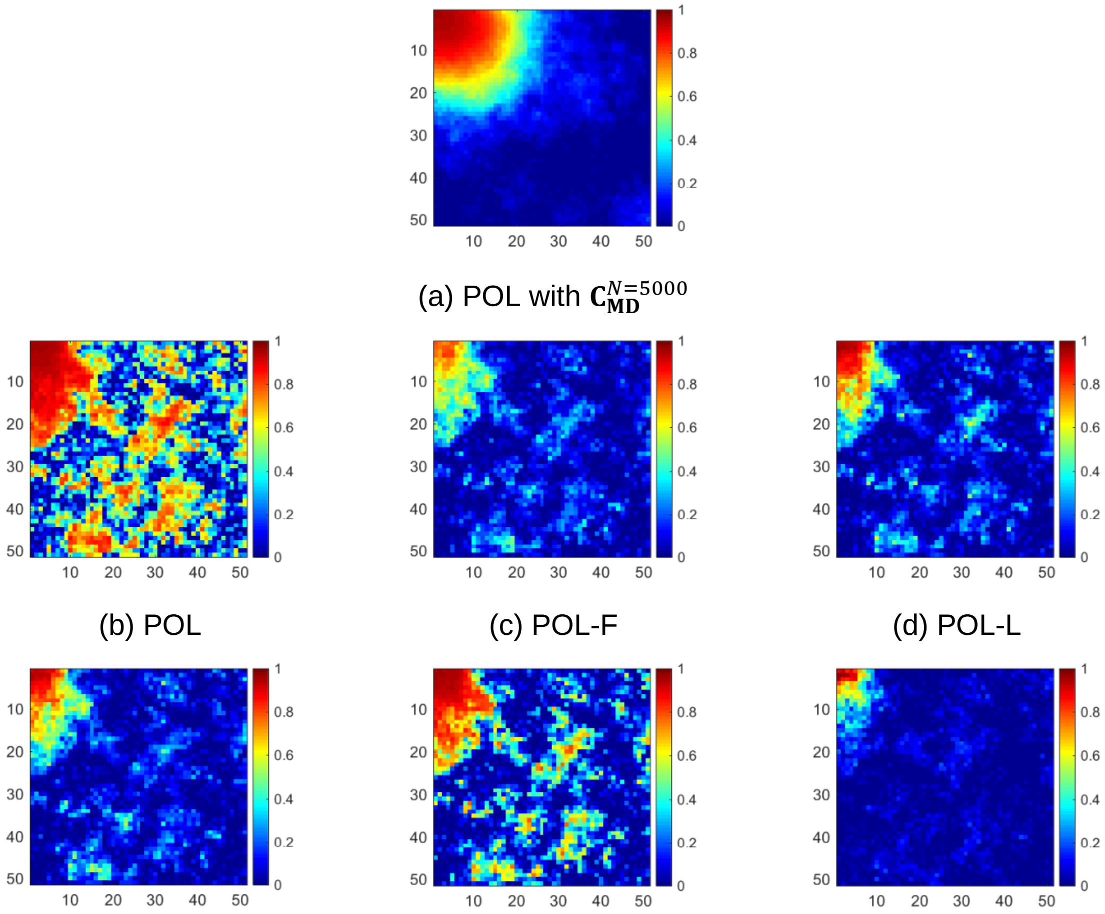
(e) POL-GC

(f) POL-EXP

(g) CL

Analyzing the localization matrices generated with the ensemble size $N = 50$ (Figure 21), one can also observe that the penalty factor resulted in certain extreme tapering in the localization matrix. Our explanation of this behavior is that almost all covariances obtained using this ensemble size are found within the noise level. Therefore, it is reasonable to take these small tapering values to stay on the ‘safe side’ (LUO; BHAKTA, 2020).

Figure 21 - Localization matrix of $\log(k)$ for oil rate on Day 31, generated with $N = 50$.



(a) POL with $\mathbf{C}_{\mathbf{MD}}^{N=5000}$

(b) POL

(c) POL-F

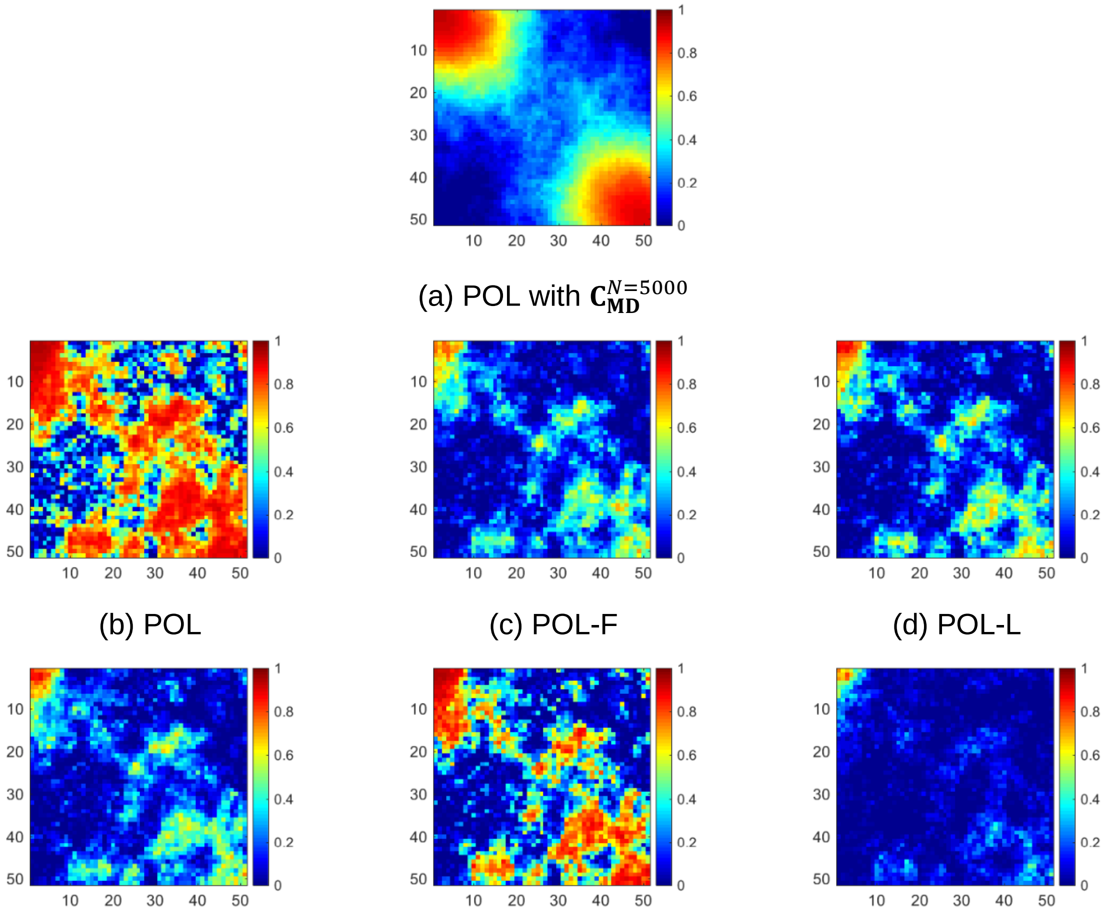
(d) POL-L

(e) POL-GC

(f) POL-EXP

(g) CL

Figure 22 - Localization matrix of $\log(k)$ for oil rate on Day 456, generated with $N = 50$.



(a) POL with $\mathbf{C}_{\mathbf{MD}}^{N=5000}$

(b) POL

(c) POL-F

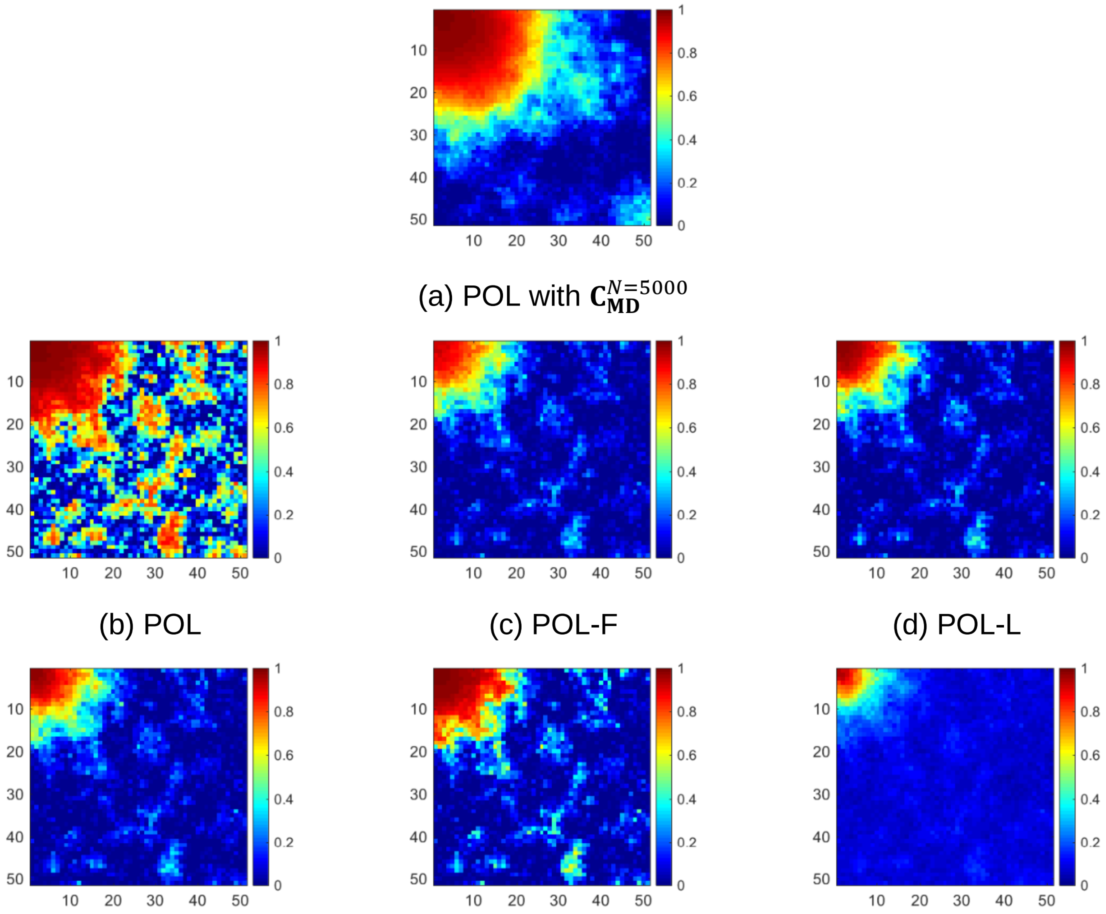
(d) POL-L

(e) POL-GC

(f) POL-EXP

(g) CL

Figure 23 - Localization matrix of $\log(k)$ for oil rate on Day 31, generated with $N = 200$.



(a) POL with $\mathbf{C}_{\mathbf{MD}}^{N=5000}$

(b) POL

(c) POL-F

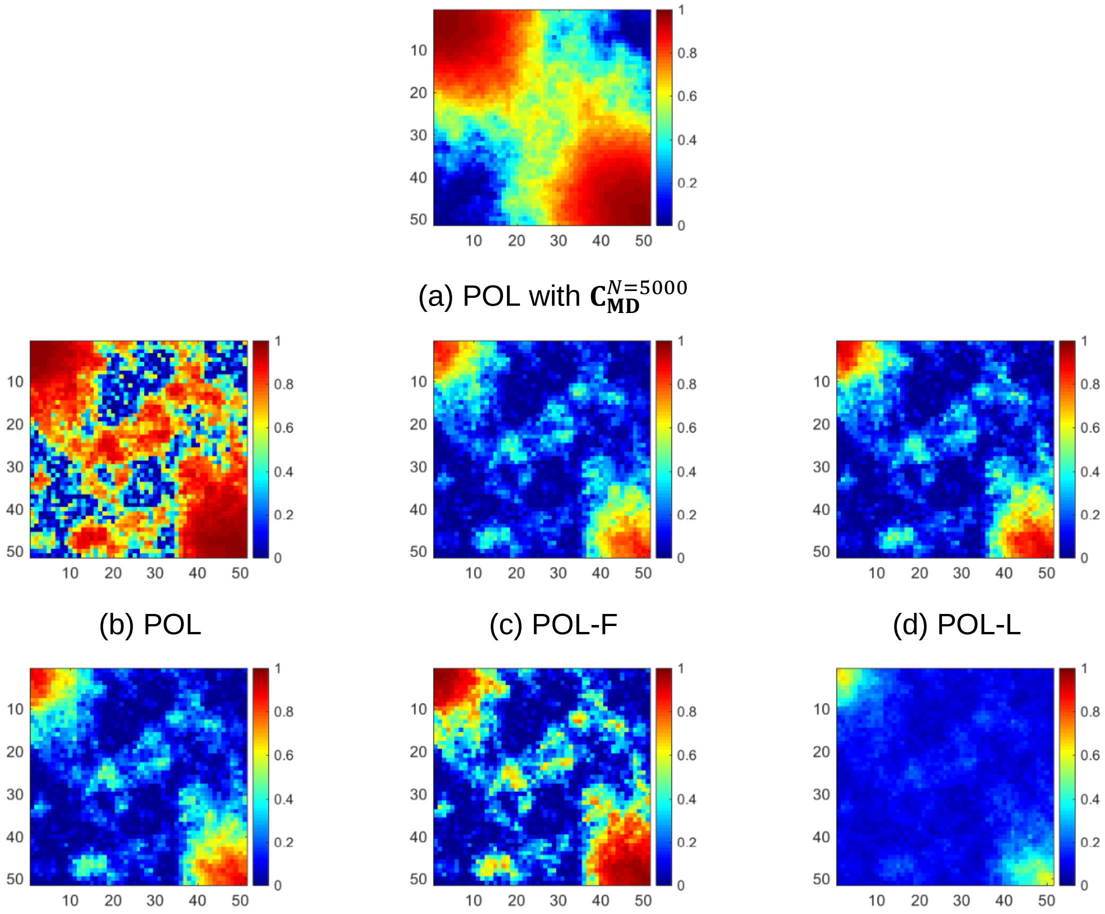
(d) POL-L

(e) POL-GC

(f) POL-EXP

(g) CL

Figure 24 - Localization matrix of $\log(k)$ for oil rate on Day 456, generated with $N = 200$.



(a) POL with $\mathbf{C}_{\mathbf{MD}}^{N=5000}$

(b) POL

(c) POL-F

(d) POL-L

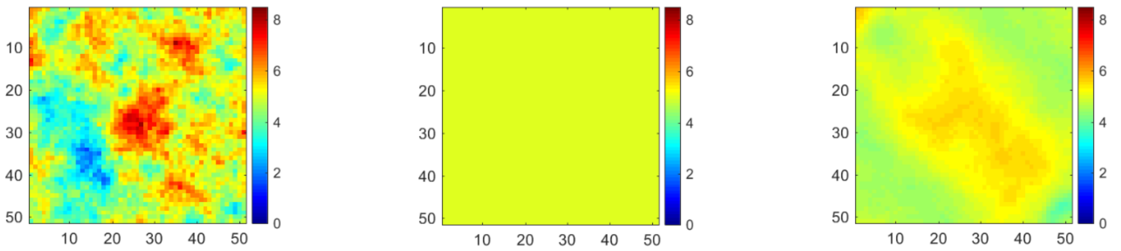
(e) POL-GC

(f) POL-EXP

(g) CL

Figure 25, Figure 26 and Figure 27 show the posterior $\log(k)$ ensemble mean of the ensemble obtained in the first experiment (called the first ensemble hereafter). There, one can see that the cases without localization and with pseudo-optimal localization resulted in rough estimates posteriors (extreme values) induced by spurious correlations. However, the presence of the penalty factor was able to help reduce this effect, even for the cases with an ensemble size equals to 50.

Figure 25 – Reference field, prior mean and posterior ensemble mean of $\log(k)$ for $N = 5000$.

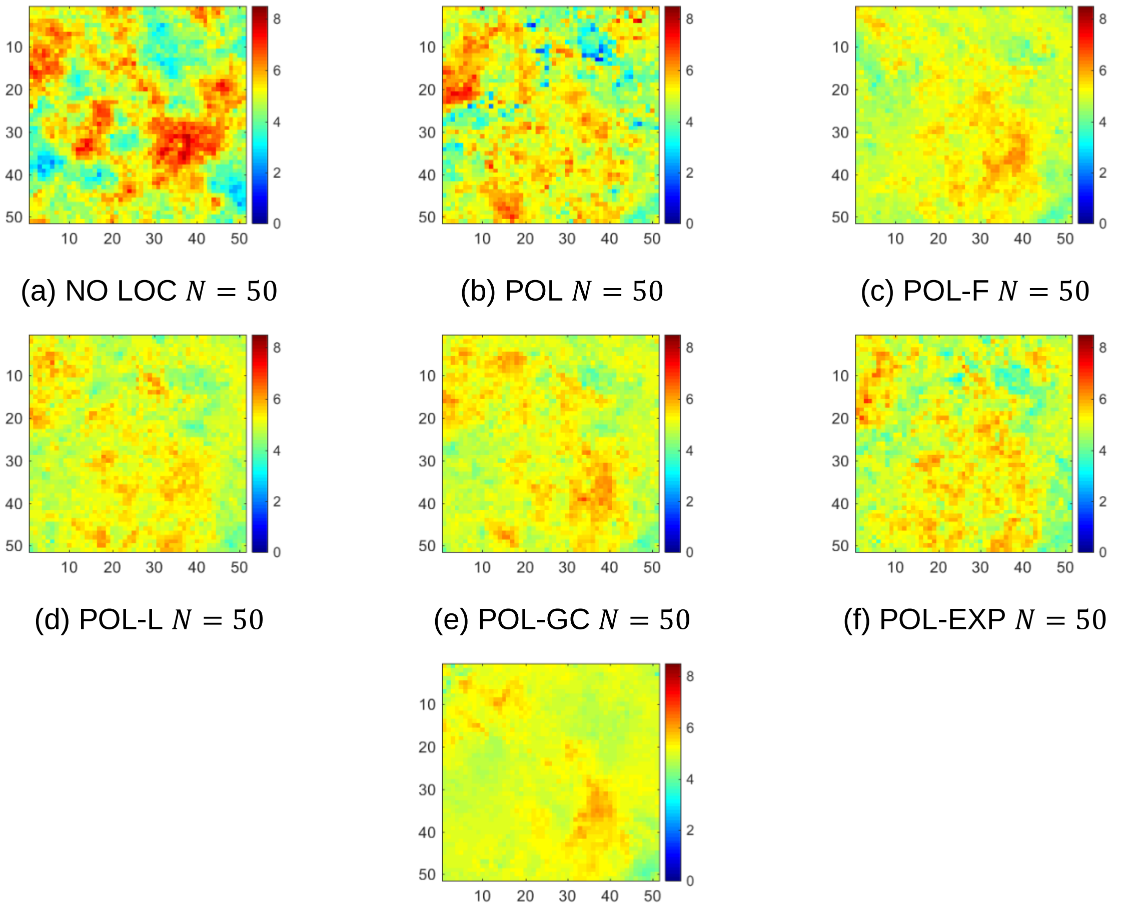


(a) Reference

(b) Prior

(c) $N = 5000$

Figure 26 - Posterior ensemble mean of $\log(k)$ of the first ensemble for $N = 50$.



(a) NO LOC $N = 50$

(b) POL $N = 50$

(c) POL-F $N = 50$

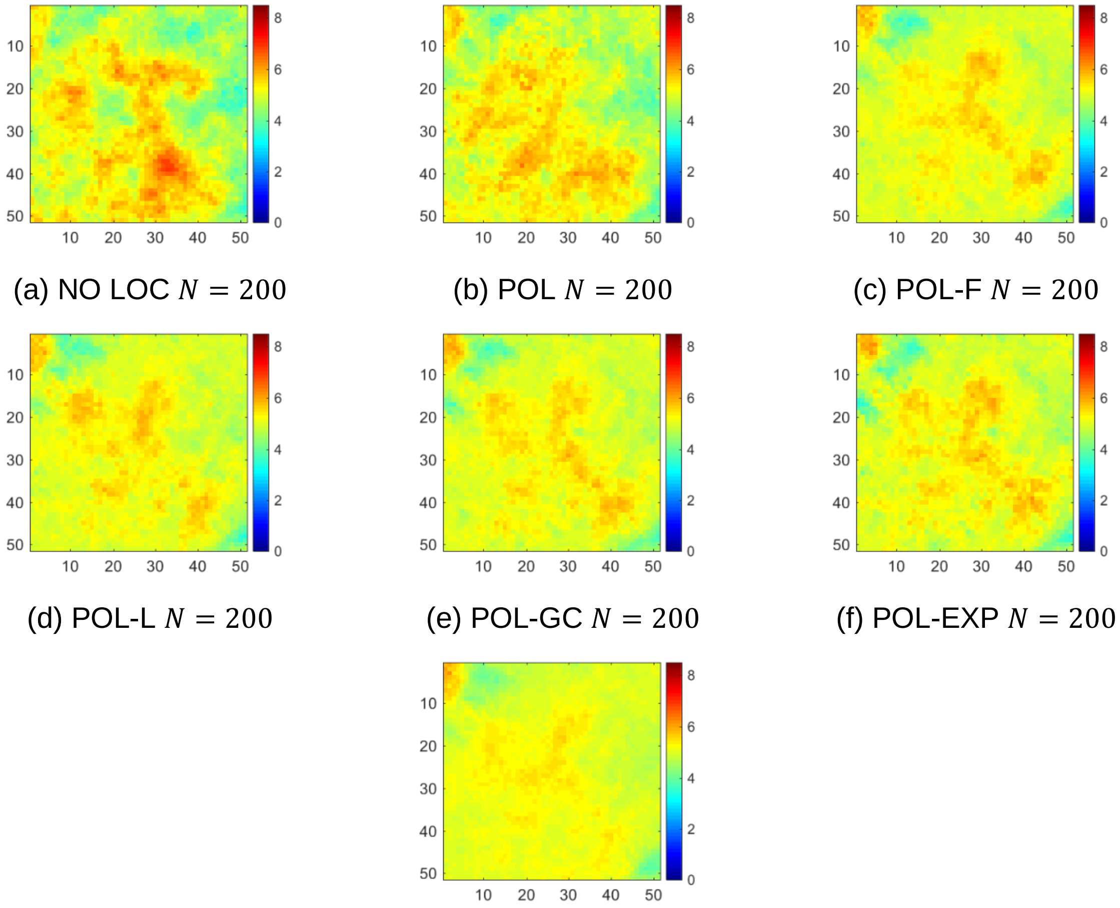
(d) POL-L $N = 50$

(e) POL-GC $N = 50$

(f) POL-EXP $N = 50$

(g) CL $N = 50$

Figure 27 - Posterior ensemble mean of $\log(k)$ of the first ensemble for $N = 200$.



(a) NO LOC $N = 200$

(b) POL $N = 200$

(c) POL-F $N = 200$

(d) POL-L $N = 200$

(e) POL-GC $N = 200$

(f) POL-EXP $N = 200$

(g) CL $N = 200$

Figure 28, Figure 29 and Figure 30 show the normalized variance of $\log(k)$ of the first ensemble in all the cases. For both ensemble sizes, the case without localization has a substantial variance reduction, or even ensemble collapse at $N = 50$. Furthermore, it is possible to verify that all localization methods induce higher posterior variances, whereas the variances resulting from the standard pseudo-optimal localization tend to be lower than the others. Finally, one can observe that the presence of the penalty term helped reduce the difference between the normalized variance of the reference case and those of the corresponding posterior ensembles, even with the ensemble size equals to 50.

Figure 28 - Normalized variance of $\log(k)$ of the first ensemble for $N = 5000$.

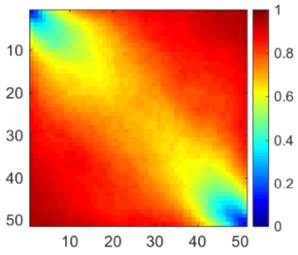
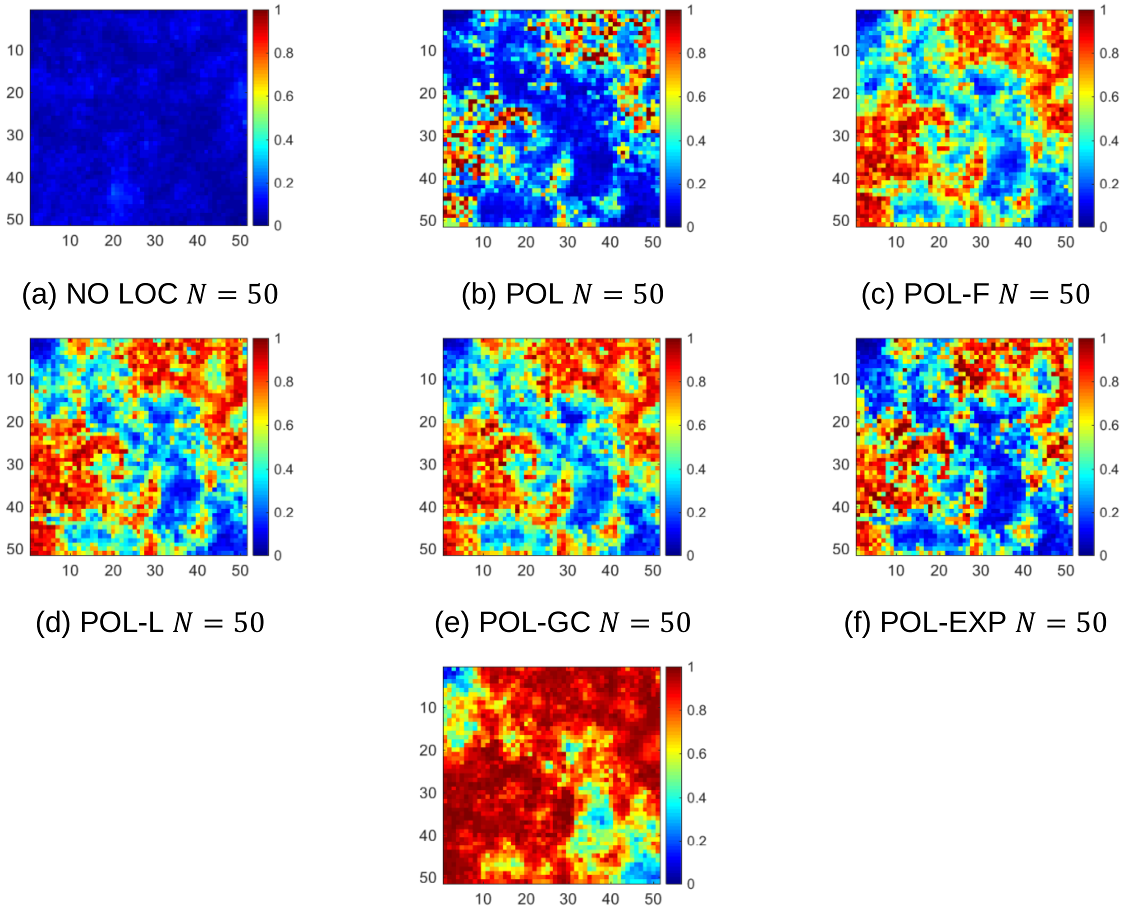


Figure 29 - Normalized variance of $\log(k)$ of the first ensemble for $N = 50$.



(a) NO LOC $N = 50$

(b) POL $N = 50$

(c) POL-F $N = 50$

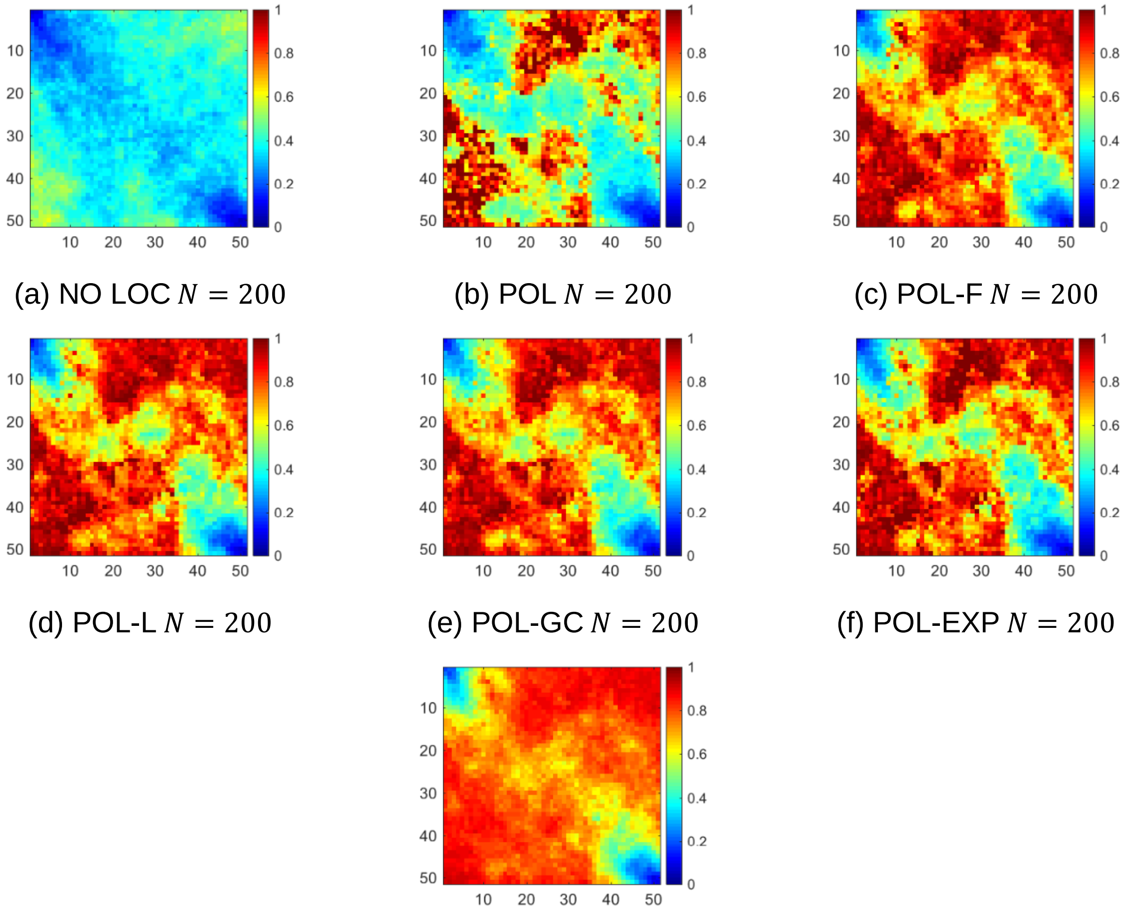
(d) POL-L $N = 50$

(e) POL-GC $N = 50$

(f) POL-EXP $N = 50$

(g) CL $N = 50$

Figure 30 - Normalized variance of $\log(k)$ of the first ensemble for $N = 200$.



(a) NO LOC $N = 200$

(b) POL $N = 200$

(c) POL-F $N = 200$

(d) POL-L $N = 200$

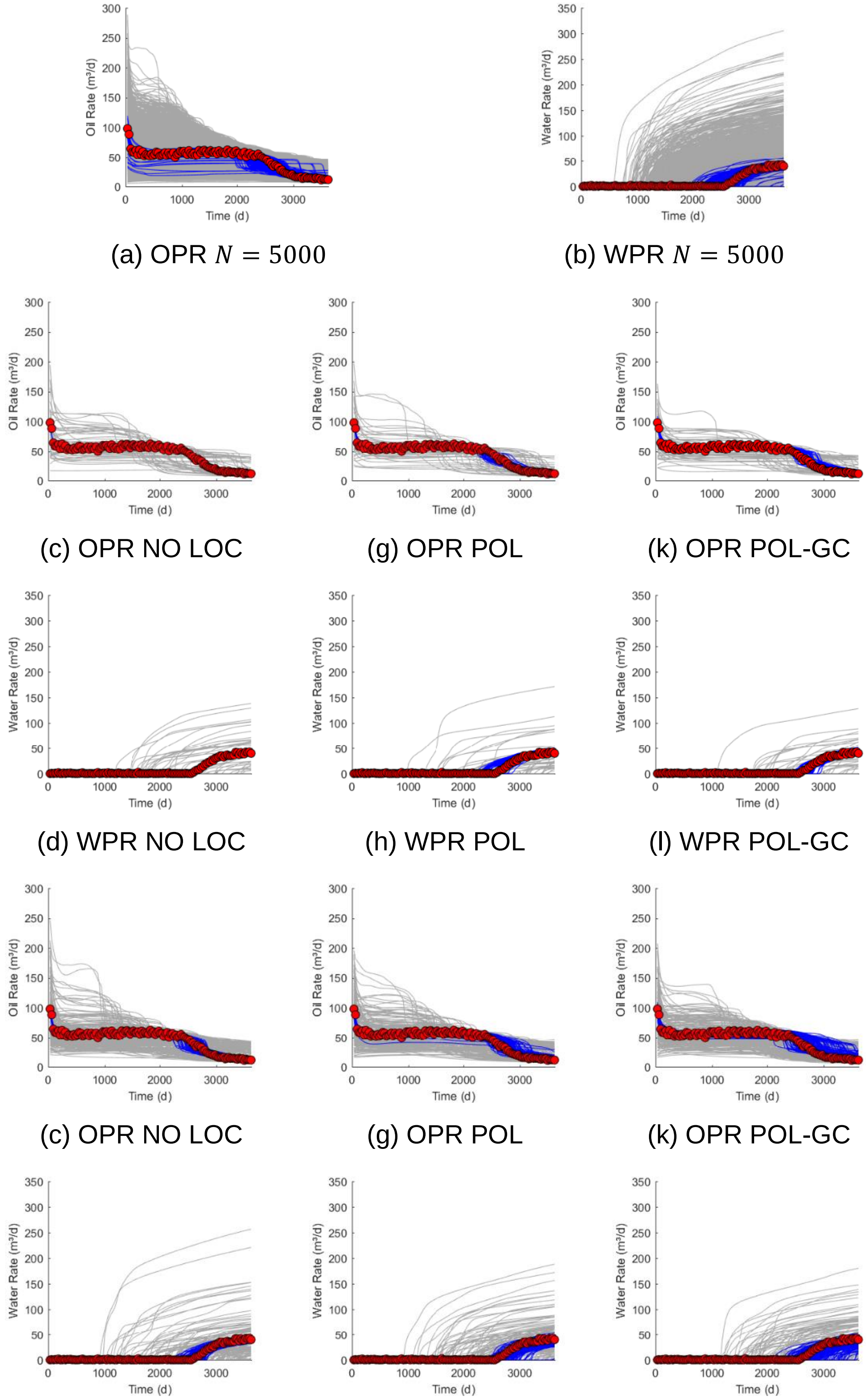
(e) POL-GC $N = 200$

(f) POL-EXP $N = 200$

(g) CL $N = 200$

Analyzing the time series of production data (Figure 31), it is possible to confirm that ensemble collapse seems to take place in the case without localization at $N = 50$. In addition, all cases with localization resulted in improvements in terms of ensemble spread.

Figure 31 - Time series of production data from the producer well. OPR refers to oil production rate and WPR refers to water production rate. Here, the gray lines represent the prior ensemble, blue lines represent the posterior ensemble, and the red dots represent the measurements.



(a) OPR $N = 5000$

(b) WPR $N = 5000$

$N = 50$

(c) OPR NO LOC

(g) OPR POL

(k) OPR POL-GC

(d) WPR NO LOC

(h) WPR POL

(l) WPR POL-GC

$N = 200$

(c) OPR NO LOC

(g) OPR POL

(k) OPR POL-GC

(d) WPR NO LOC

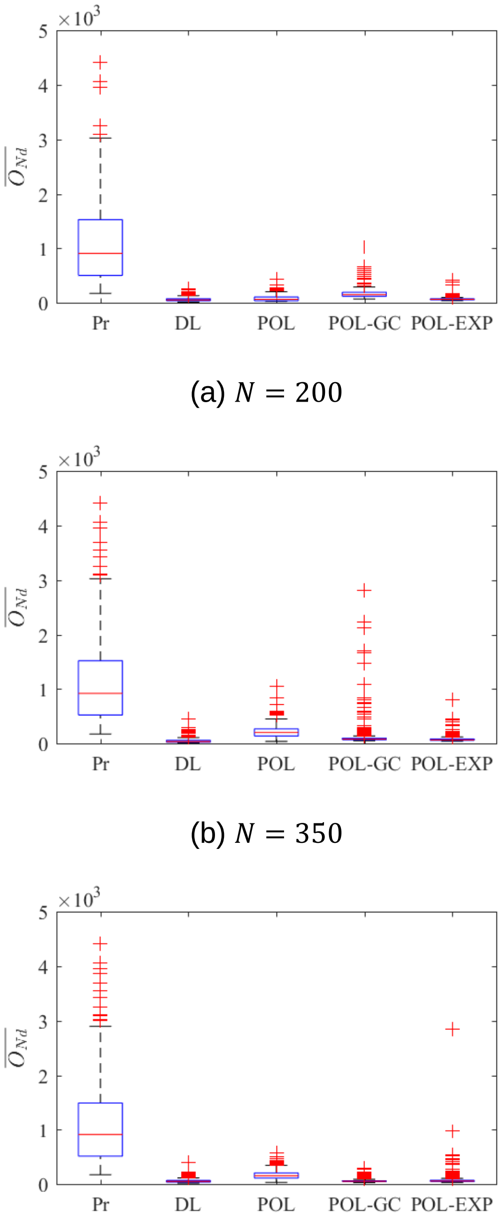
(h) WPR POL

(l) WPR POL-GC

5.1.2 UNISIM-I-H

The prior and posterior data mismatch obtained with different ensemble sizes and localization schemes are shown in Figure 32, where it is possible to verify that in all cases data mismatch were reduced significantly.

Figure 32 - Boxplots of data mismatch for the UNISIM-I-H case study. Here, the label 'Pr' refers to the prior ensemble, 'DL' to distance-based localization, 'POL' to pseudo-optimal localization, and 'POL-GC' and 'POL-EXP' to the pseudo-optimal localization with Gaspari and Cohn penalty factor and exponential penalty factor, respectively.



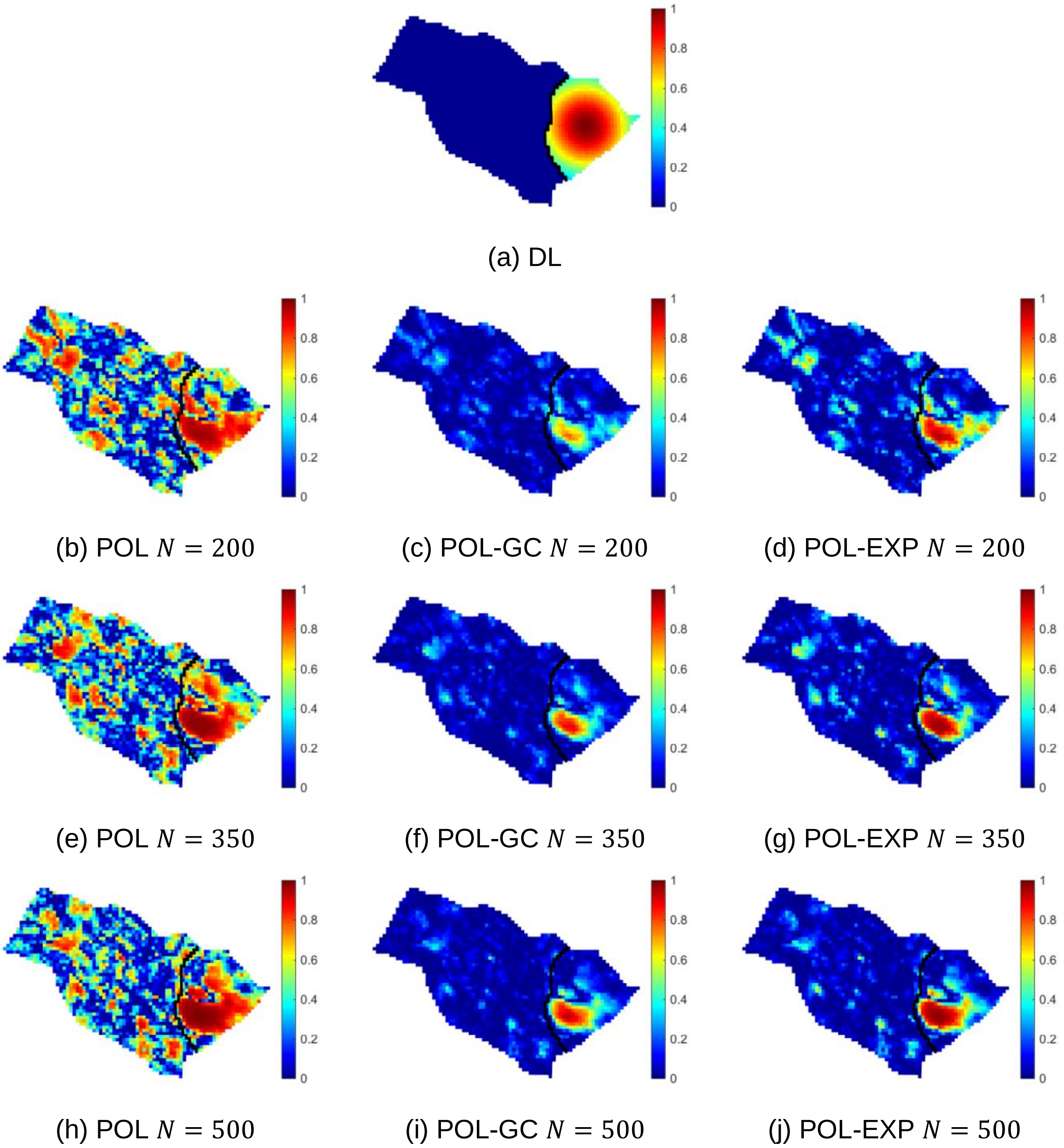
(a) $N = 200$

(b) $N = 350$

(c) $N = 500$

Figure 33 shows the maps of tapering coefficients obtained by different localization schemes, for log-permeability in x-direction on Layer 9, with respect to the producer well PROD023A water production rate at the last simulation period (Day 4018). Here, for distance-based localization (DL), the localization matrix was made to contain nonzero values only in the same zone of the data. One can observe that pseudo-optimal localization in its standard form is not able to properly suppress spurious correlations in the UNISIM-I-H case, even with the largest tested ensemble size ($N = 500$). Furthermore, the addition of the penalty factor resulted in smoother tapering values and spurious correlations reduction for all localization schemes where the penalty factor is introduced to equation (16).

Figure 33 - Map of tapering coefficients for $\log(k_x)$ of Layer 9, PROD023A water production rate, on Day 4018, for the UNISIM-I-H case study.



(a) DL

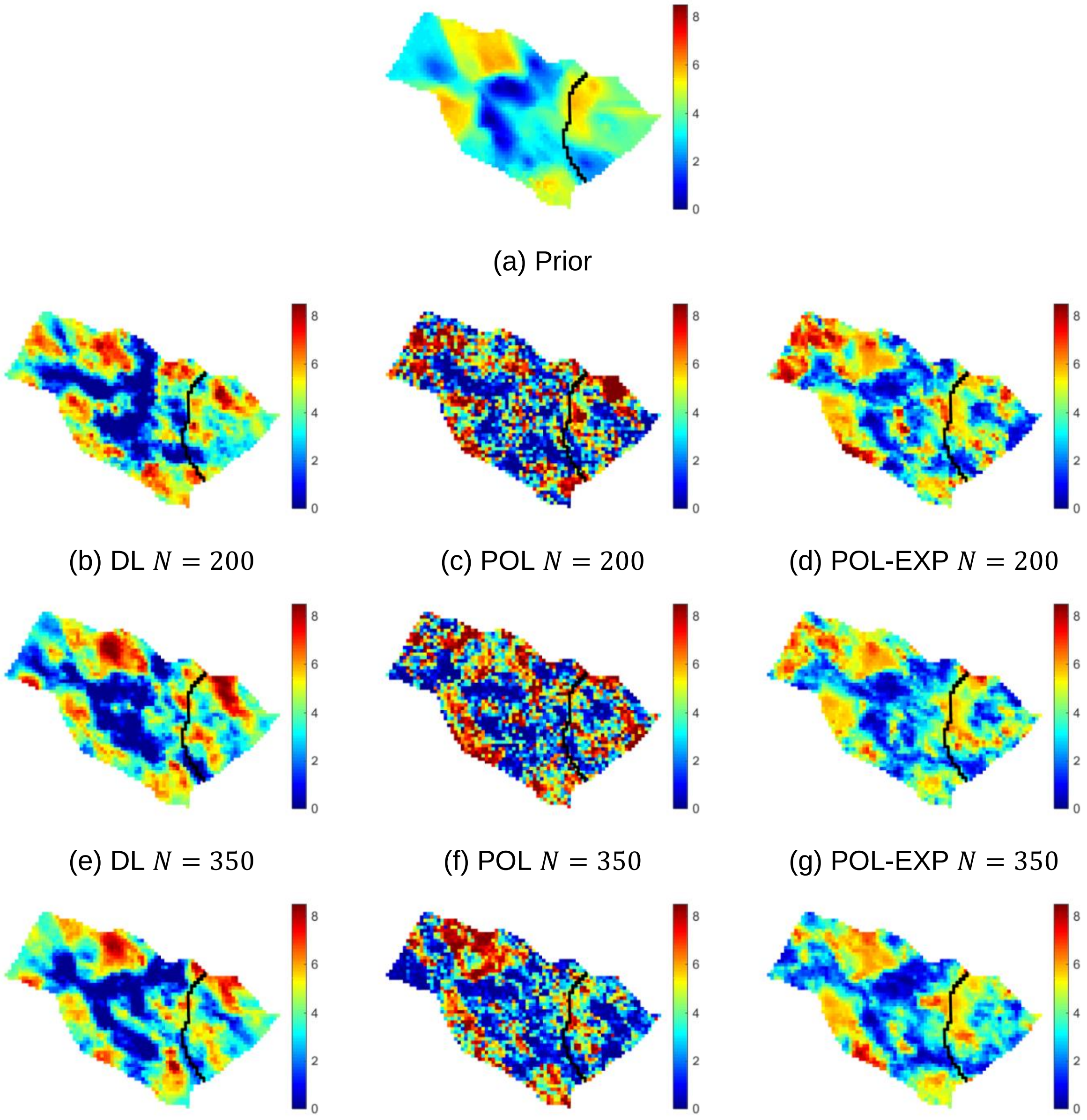
(b) POL $N = 200$ (c) POL-GC $N = 200$ (d) POL-EXP $N = 200$

(e) POL $N = 350$ (f) POL-GC $N = 350$ (g) POL-EXP $N = 350$

(h) POL $N = 500$ (i) POL-GC $N = 500$ (j) POL-EXP $N = 500$

In Figure 34, the prior and posterior ensemble means of the log-permeability in the x-direction (on Layer 9) are shown. Here, one can see how history matching with the standard pseudo-optimal localization resulted in extreme values and how the inclusion of the penalty factor mitigates this effect, making the posterior ensemble mean closer to the prior mean, even in comparison to history matching results with distance-based localization.

Figure 34 - Prior and Posterior ensemble mean of $\ln(k_x)$ on Layer 9 of the UNISIM-I-H case study.



(a) Prior

(b) DL $N = 200$

(c) POL $N = 200$

(d) POL-EXP $N = 200$

(e) DL $N = 350$

(f) POL $N = 350$

(g) POL-EXP $N = 350$

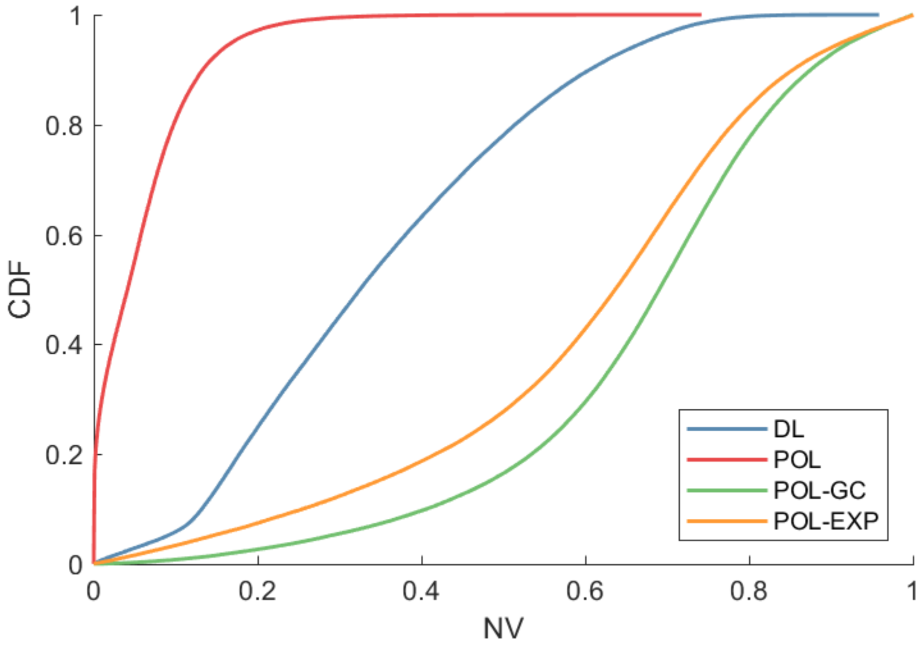
(h) DL $N = 500$

(i) POL $N = 500$

(j) POL-EXP $N = 500$

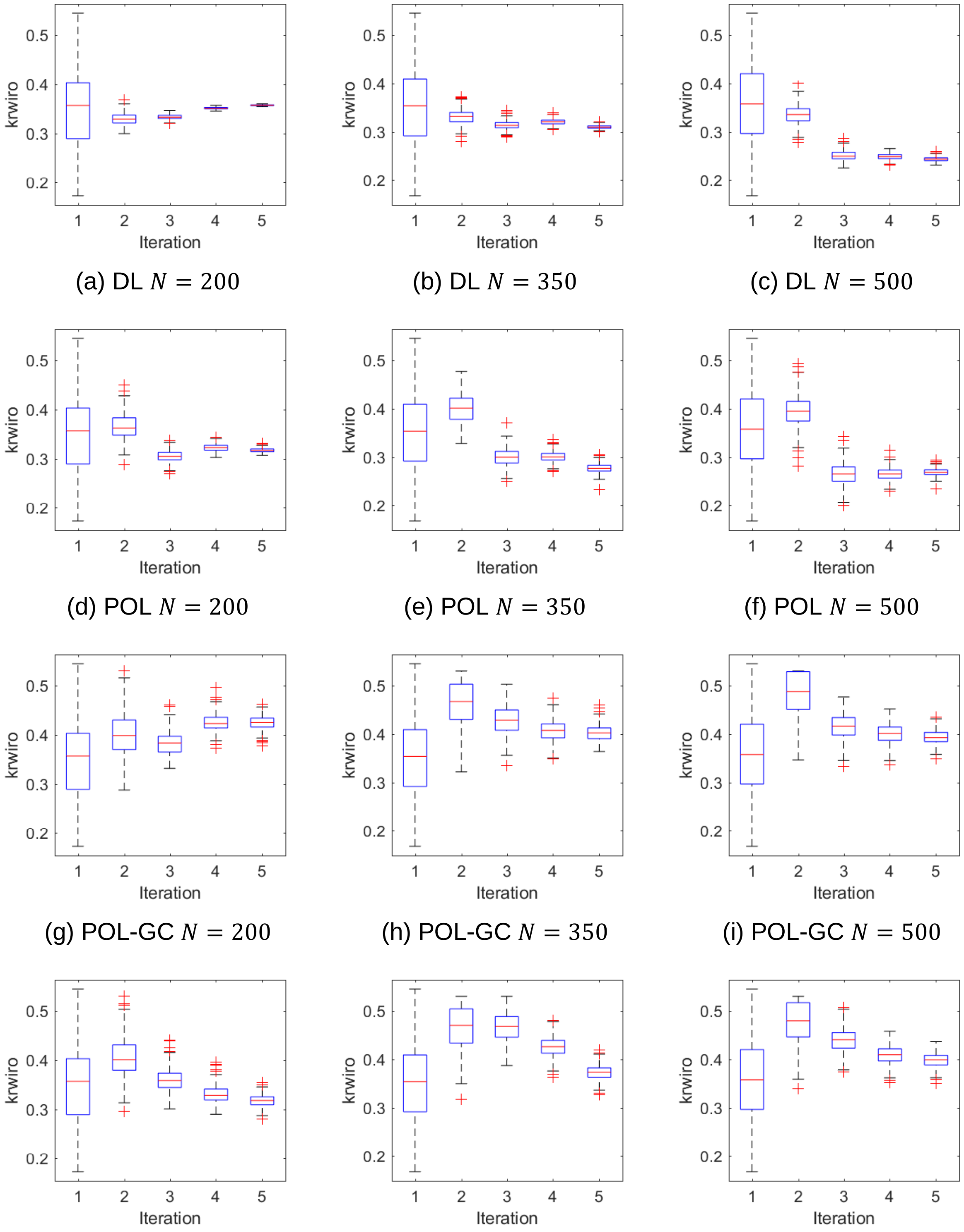
Inspecting the cumulative distribution function (CDF) of normalized variances of all reservoir uncertain parameters (Figure 35), one can see a sign of ensemble collapse (with a concentration of relatively small normalized variances) when the standard pseudo-optimal localization is applied, and the improvement when the penalty factor is introduced, with the POL-EXP results showing a slightly higher variance reduction.

Figure 35 - Cumulative distribution function (CDF) of normalized variance for the UNISIM-I-H case study.



The normalized variances of non-local and local parameters are analyzed separately in the sequence. Figure 36 shows the boxplots of maximum water relative permeability at each iteration step, and Figure 37 presents the normalized variance of the log-permeability in x-direction on Layer 9. When distance-based localization is applied only to local parameters, it results in an ensemble collapse in the estimated non-local parameters, independently on the ensemble size. Relatively small ensemble spread is also spotted in the standard pseudo-optimal localization, mainly due to the retained spurious correlations. In contrast, excessive uncertainty reduction was mitigated when the penalty factor is applied to the pseudo-optimal localization.

Figure 36 - Boxplots of maximum water relative permeability (krwiro) at all iteration steps of the UNISIM-I-H case study.



(a) DL $N = 200$
(b) DL $N = 350$
(c) DL $N = 500$

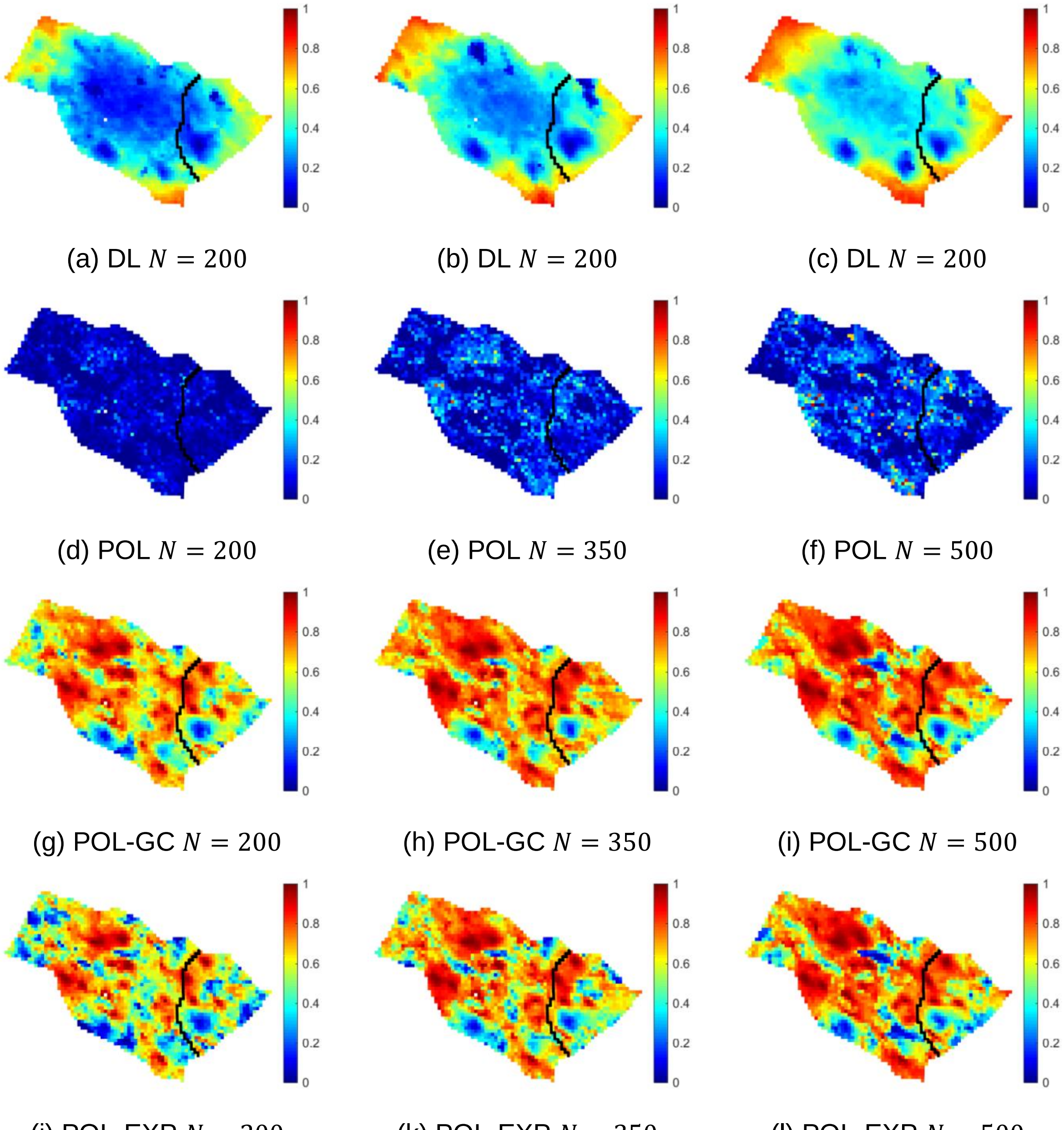
(d) POL $N = 200$
(e) POL $N = 350$
(f) POL $N = 500$

(g) POL-GC $N = 200$
(h) POL-GC $N = 350$
(i) POL-GC $N = 500$

(j) POL-EXP $N = 200$
(k) POL-EXP $N = 350$
(l) POL-EXP $N = 500$

Analyzing the normalized variance of $\log(k_x)$ on Layer 9, one can see that distance-based localization reduces the posterior variance in almost the entire region around the wells. Furthermore, the standard pseudo-optimal localization induces a sign of ensemble collapse. As such, one can conclude that the application of the penalty factor resulted in better variance preservation. In addition to data mismatch, the history matching performance can also be analyzed by having an examination on the time series of production data, with respect to the prior and posterior ensembles. Figure 38 shows the production data from the producer well NA1A, for the case with the ensemble size equals to 500. There, one can see slightly worse data match for the case with the penalty factor applied to the pseudo-optimal localization. On the other hand, the introduction of the penalty factor is clearly beneficial for preserving the ensemble varieties.

Figure 37 - Normalized variance of $\ln(k_x)$ on Layer 9 of the UNISIM-I-H case study.



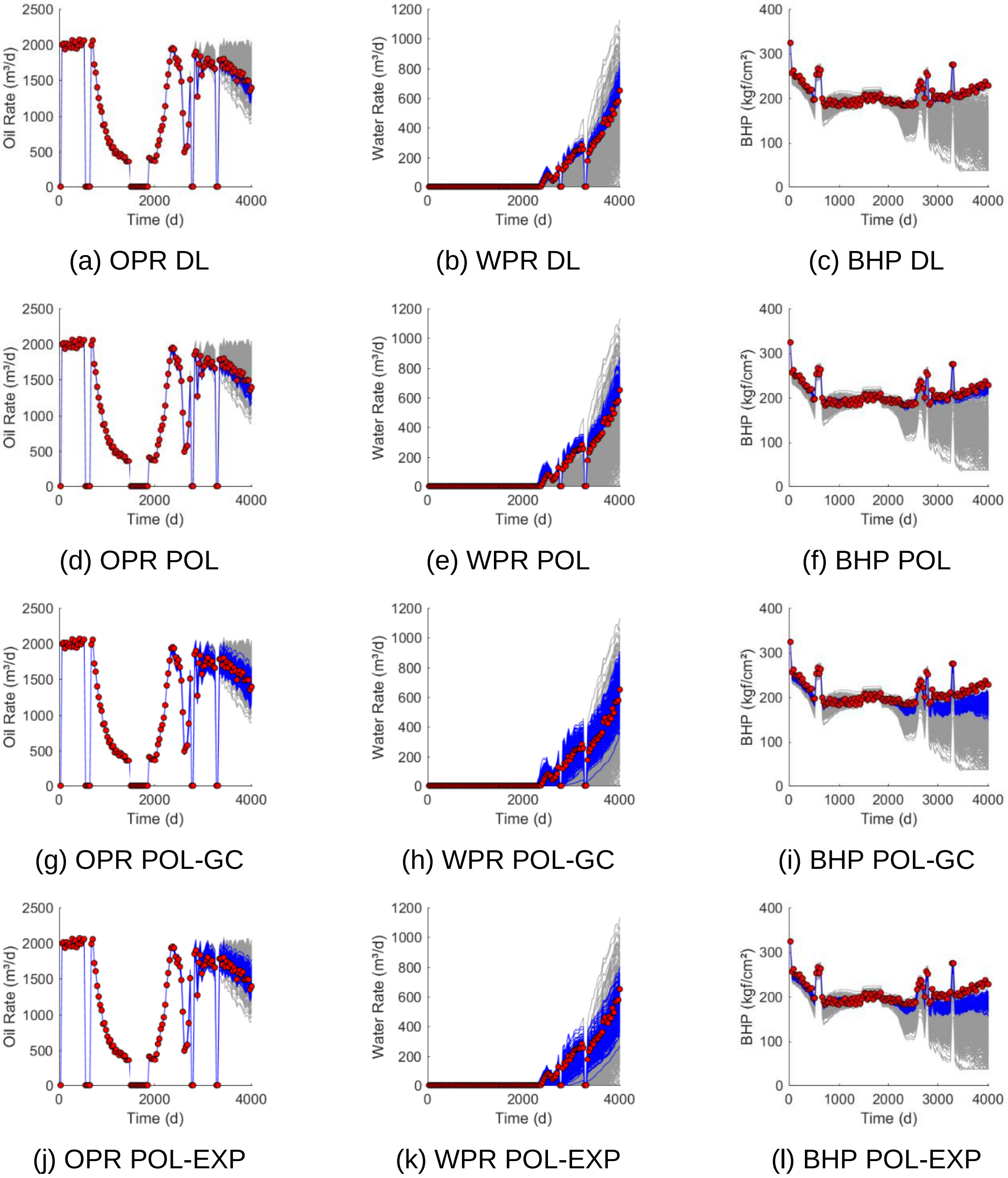
(a) DL $N = 200$ (b) DL $N = 200$ (c) DL $N = 200$

(d) POL $N = 200$ (e) POL $N = 350$ (f) POL $N = 500$

(g) POL-GC $N = 200$ (h) POL-GC $N = 350$ (i) POL-GC $N = 500$

(j) POL-EXP $N = 200$ (k) POL-EXP $N = 350$ (l) POL-EXP $N = 500$

Figure 38 - Time series of production data from the producer well NA1A, for the case with the ensemble size equals to 500. Here, OPR stands for oil production rate, WPR for water production rate, and BHP for bottom-hole pressure.



(a) OPR DL (b) WPR DL (c) BHP DL

(d) OPR POL (e) WPR POL (f) BHP POL

(g) OPR POL-GC (h) WPR POL-GC (i) BHP POL-GC

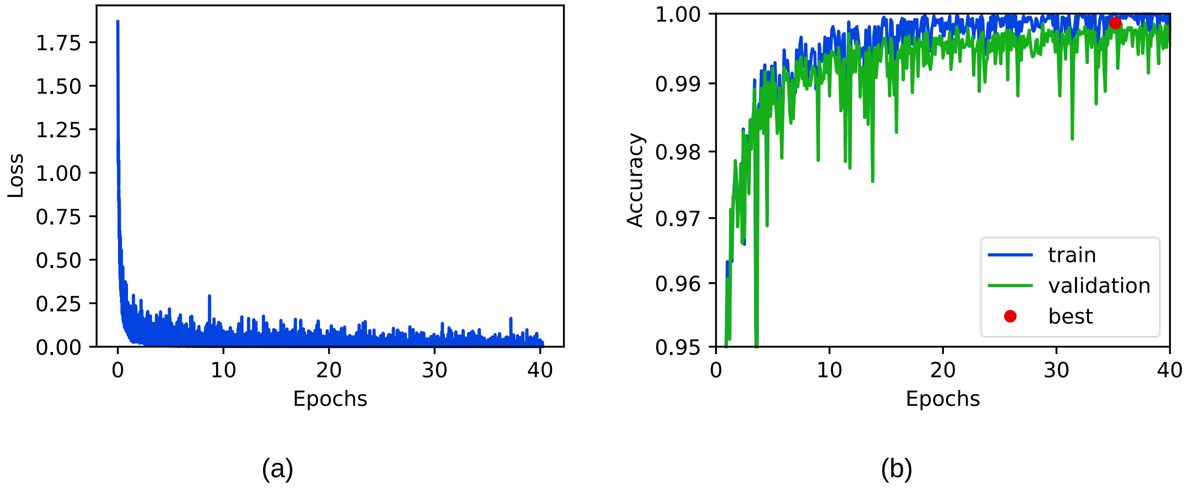
(j) OPR POL-EXP (k) WPR POL-EXP (l) BHP POL-EXP

5.2 Parameterization

5.2.1 Reservoir classifier and Fréchet Reservoir Distance results

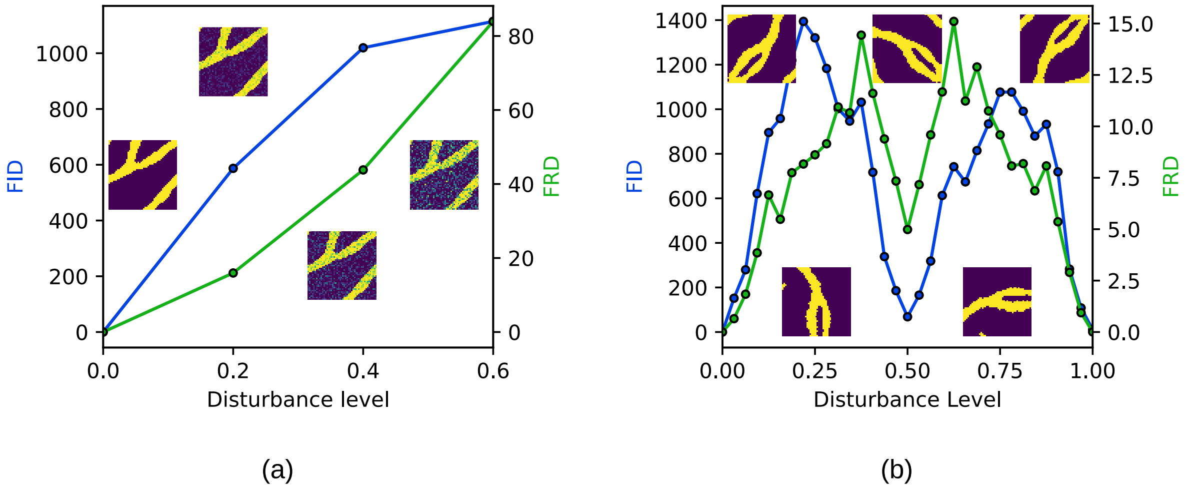
The Reservoir Classifier was trained with 40 epochs, being the best accuracy obtained in the validation set around 0.9987 at the end of the training process. Figure 39 shows the loss function values during the training (a), and the accuracy measured using the training and validation dataset over the iterations (b).

Figure 39 – Loss (a) and accuracy (b) for training and validation dataset obtained after Reservoir Classifier training. In (b), the red dot represents the best accuracy obtained in the validation dataset.



Subsequently, an evaluation was conducted to assess the performance of the metrics while employing the Inception and Reservoir classifier networks under the influence of distortions with different disturbance levels. To avoid biased metric results, we used an ensemble sufficiently large ($N = 20{,}000$) and computed the metric between the original ensemble and the ensemble after introducing two types of distortions (Figure 40): Gaussian noise (a) and image rotation with a fixed angle (b). Based on the results, it is possible to confirm that the Reservoir Classifier network yielded improved results. Specifically, for certain rotated images, the FID results tend to 0. This implies that the Inception-v3 network was unable to extract features that differentiate between the two datasets.

Figure 40 – FID and FRD for different disturbance levels. Left image represents the metric computed after the addition of a Gaussian random noise, while the right image represents the metric computed after an image rotation. In the Gaussian noise the disturbance level represents the noise variance, while the disturbance level represents the angle rotation varying between 0 and π.



We also compared both metrics in terms of their evolution over the GAN training for two datasets. The first dataset consists of the two facies channelized reservoir built using properties that differ from those used during the Reservoir Classifier training. The second dataset comprises randomly cropped sections of the matrix permeability of the UNISIM-II-H benchmark, which was not used in the Reservoir Classifier training. The purpose of using these two datasets was to compare the metrics using datasets considering facies and property distributions. In terms of GAN training, both cases were training and its metrics evaluated with a dataset sufficiently large ($N = 80{,}000$) to prevent discriminator overfitting and biased metric computation.

Analyzing the FID and FRD results (Figure 41 and Figure 42), it can be observed that both metrics yielded similar results. It is worth noting that the best FID value for the two-facies channelized case was achieved approximately at the mid-point of the training, due to the high variance of the FID metric. This behavior is significant since the dataset contains images with integer values, which were not included in the Inception-v3 network's training dataset, the ImageNet dataset (DENG et al., 2009).

Figure 41 - FID and FRD evolution over iterations for NSGAN application in the 2 facies channelized dataset. The upper rows contain two random realizations for different iteration steps (represented by the vertical dashed lines).

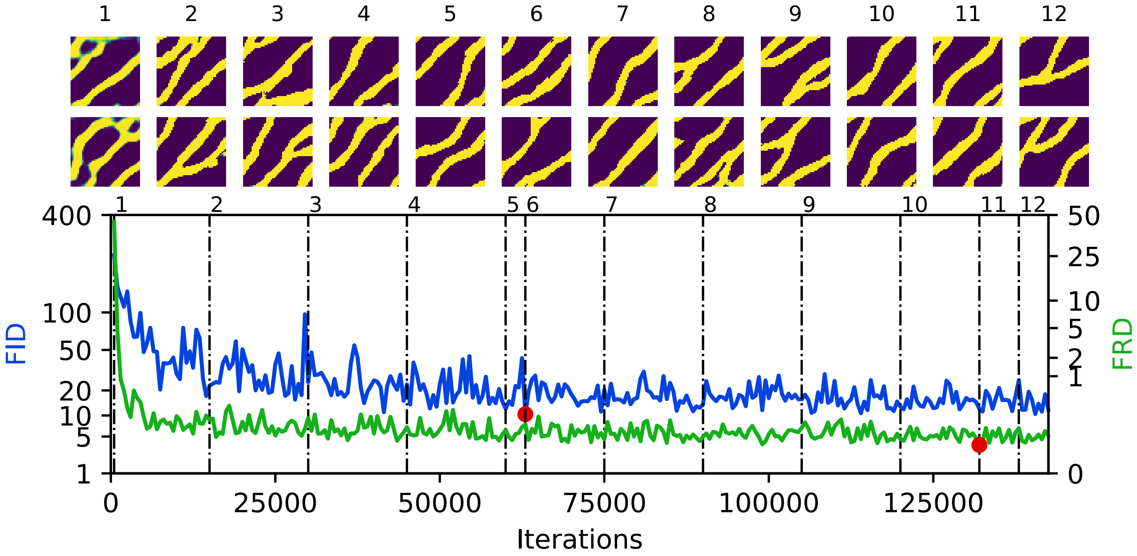
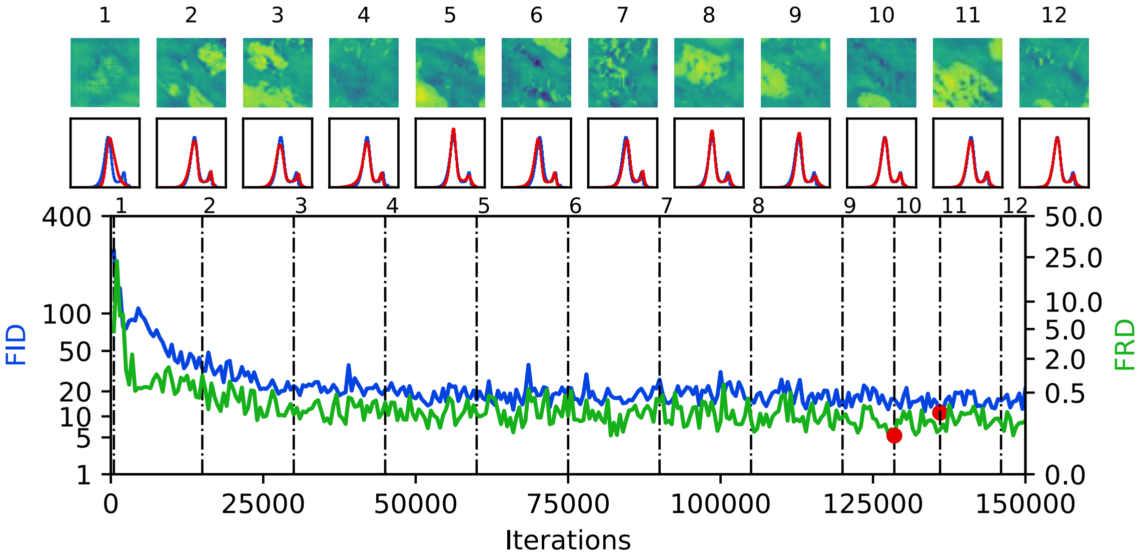


Figure 42 – FID and FRD evolution over iterations for NSGAN application in the UNISIM-II-cropped dataset. The upper rows contain a random realization and its corresponding histogram for different iteration steps (represented by the vertical dashed lines). Regarding the histograms, the blue line represents the histogram of the training dataset, and the red line represent the histogram of the generated dataset.



5.2.2 GAN regularization results

Initially, we evaluated the impact of discriminator regularization on GAN training performance by training multiple networks with different regularization weights $\gamma_{R_1} = [0.002, 0.005, 0.01, \ldots, 100, 200, 500]$, with a dataset size of 80,000 samples. For each weight value, we conducted three network runs. Figure 43 shows the best obtained FRD after a given number of iterations (50,000; 100,000; and 150;000) during the GAN training, and Figure 44 shows the FRD over the iterations for different regularization weights. Based on the results, it can be concluded that the optimum weight differs significantly from the recommended by Karras et al. (2020). This difference may be attributed to the fact that in simpler cases (single-channel petroleum reservoir realizations instead of more complex faces images), the discriminator has a tendency to learn the generated distribution at a much faster rate. Consequently, the discriminator loss function approaches zero, requiring higher values of the regularization weight.

Figure 43 – Best FRD obtained after a given number of iterations as function of the γ of the R_1 regularization. The lines represent the median, while the spread represents the maximum and minimum obtained value for each value of γ. The red square area represents the range recommended by Karras et al. (2020) to choose the regularization weight.

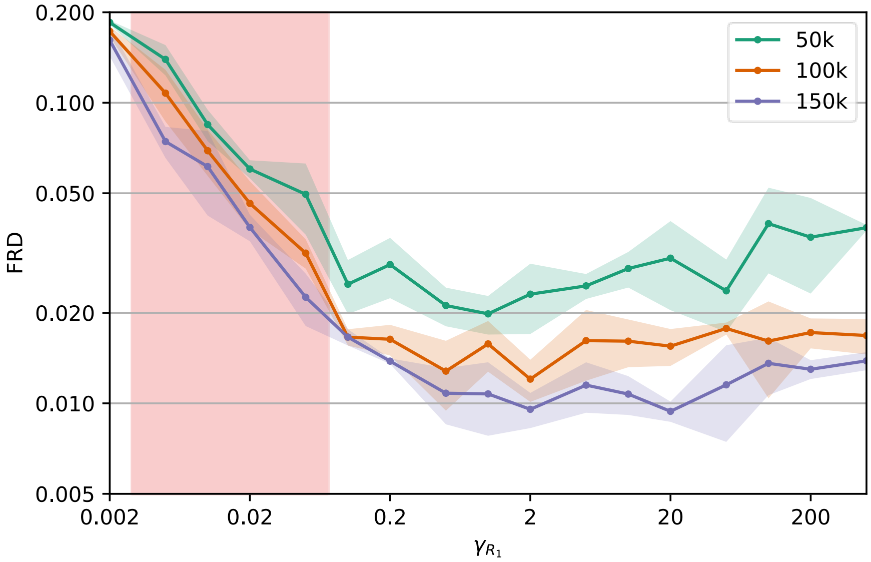
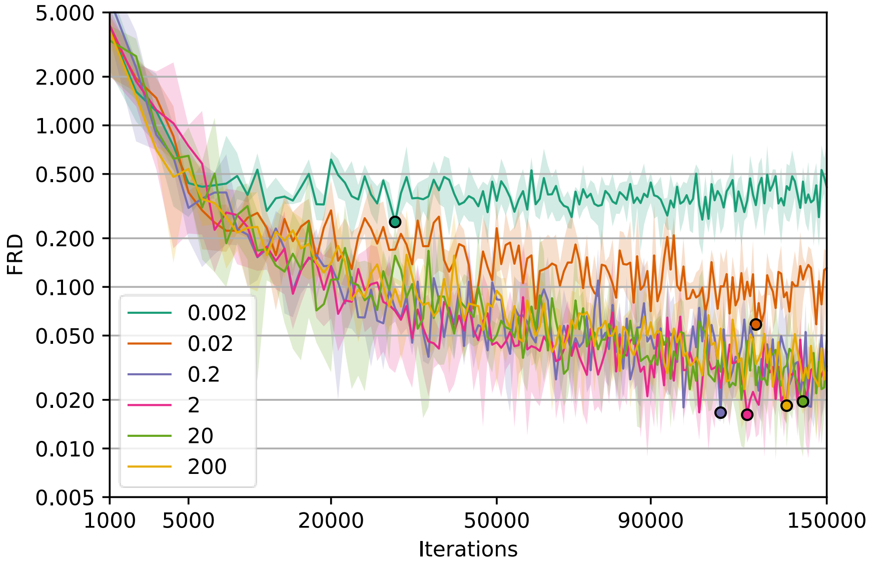


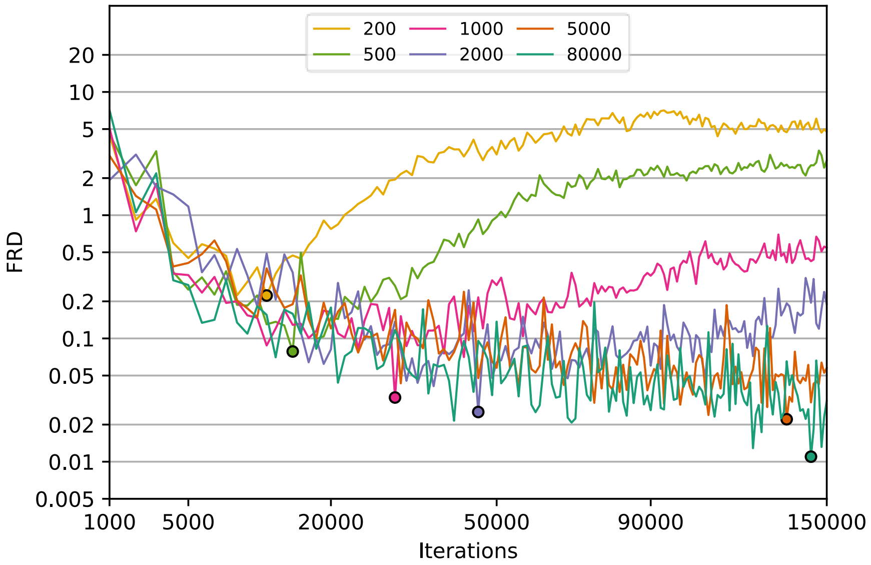
Figure 44 – FRD over iterations for different regularization weights. The lines represent the median, while the spread represent the maximum and minimum. The dots represent the best median value obtained for each case.



5.2.3 Adaptive Discriminator Augmentation Results

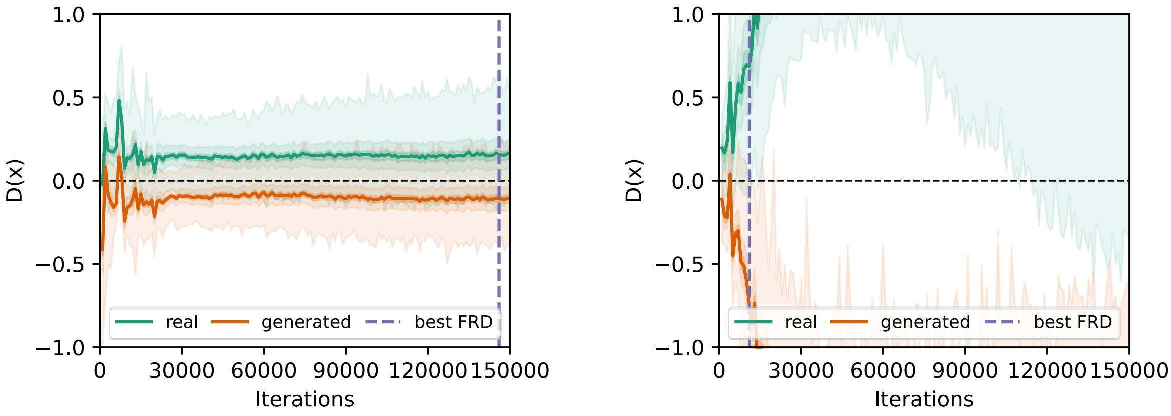
Before investigating the effect of adaptive augmentation on GAN performance, we conducted a prior investigation on the influence of dataset sample size. Figure 45 depicts the FRD across iterations for different dataset sizes. From the figure, it is apparent that even ensemble sizes considered as large to data assimilation studies (such as the case with 1,000 samples) are insufficient to train a generative adversarial network with reasonable performance.

Figure 45 – FRD over iterations for different dataset sizes. Here, the dots represent the best FRD obtained.



According to the hypothesis from Karras et al. (2020), the quality of the generator begin to deteriorate when the discriminator overfits. This overfitting can be measured by the discriminator outputs across the training. In Figure 46 the discriminator outputs for the original and generated images are displayed for two distinct cases with different dataset sizes ($N = 200$ and $N = 80{,}000$). Here, we can observe that the aforementioned hypothesis is still valid. In the case with the larger dataset, the discriminator outputs remain close to zero with both distributions overlapping, and the FRD consistently decreases throughout the entire training process. On the other hand, in the case where overfitting occurs, the discriminator outputs move away from zero, and the overlap between the real and generated distributions ends, coinciding with the points where the FRD starts to deteriorate.

Figure 46 – Discriminator outputs for different dataset sizes N. Here, the solid lines represent the mean and the spread represent the percentiles. The purple dashed line represents the iteration where the best FRD was obtained.

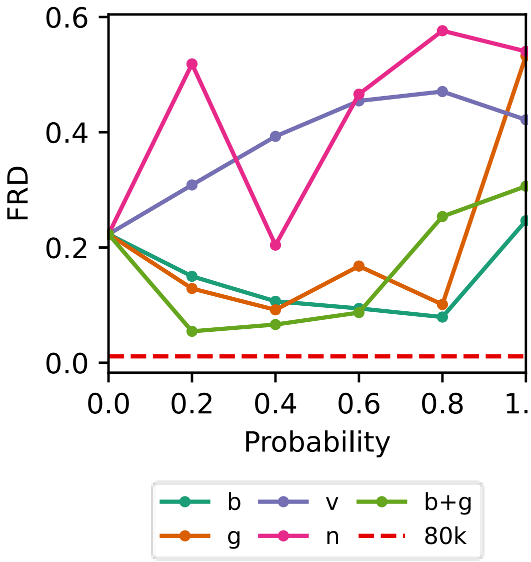


(a) $N = 80{,}000$

(b) $N = 200$

Next, we conducted an investigation into the effect of different augmentation categories on the performance of GANs in the context of petroleum reservoirs (Figure 47). To evaluate the efficacy of each category, we trained a separate network with a fixed augmentation probability $p = [0.2, 0.4, 0.6, 0.8, 1.0]$ for each category. Our analysis revealed that only blitting and geometric augmentations had a positive impact on the GAN training. Specifically, the GANs trained using color and noise augmentations resulted in higher FRD values compared to those without augmentations (Figure 47, $p = 0$). Moreover, we found that the cumulative augmentation of blitting and geometric transformations resulted in even lower FRD values. Furthermore, one interesting finding was that even with lower augmentation probability values, a significant improvement in the FRD was obtained, resulting in values that are closer to the reference case with 80,000 samples used during training. In addition, the leaking behavior is also observed when higher augmentation probabilities are applied, presenting higher FRD values. This suggests that the use of augmentation techniques can compensate the lack of training data, which is especially valuable in cases where generating a large prior dataset is costly or impractical. These findings are consistent with those made by Karras et al. (2020), who also applied the adaptive data augmentation to train a generative adversarial network with the FFHQ dataset (KARRAS; LAINE; AILA, 2018). Thus, it is possible to state that not all augmentations are effective in improving the GAN performance, beyond the petroleum reservoir context. Finally, it is possible to verify that even with lower augmentation probability values, good improvement in the FRD was obtained, achieving values closer to the reference case with 80,000 samples used during the training.

Figure 47 – FRD for different augmentation probabilities and categories. Here, ‘b’ represents blitting augmentations, ‘g’ represents geometric augmentations, ‘v’ represents value augmentations, ‘n’ represents Gaussian noise augmentation, and ‘b+g’ represents the cumulative case with blitting and geometric augmentations combined. The red dashed line represents the case using a dataset size equal to 80,000.



In order to assess the effectiveness of the heuristics presented in Equation (48), we conducted an experiment where the GAN performance was evaluated for different values of r_{target}. To carry out this experiment, we utilized the augmentation categories of blitting and geometric transformations, which were shown to be effective in the previous results. Multiple GANs were trained with r_{target} values ranging from 0.4 to 0.975, as well as the case with $r_{target} = 1$ which can be interpreted as the cases without any augmentation. The results of the experiment are displayed in Figure 48, where it can be observed that the adaptive augmentation method leads to lower FRD values compared to the best obtained FRD with fixed augmentation probability, indicating an improvement in the training performance. Additionally, the proposed heuristics using the PD controller was able to keep the r_t values on track with the r_{target} values throughout the entire training process (as shown in Figure 49). As demonstrated in Figure 50, there were no significant variations in the augmentation probability for all cases, indicating that the PD controller was effective in control the augmentation probability properly. Furthermore, it is worth noting that the discriminator outputs remain close to zero when the adaptive augmentation is applied, as shown in Figure 51, similarly as the case with $N = 80{,}000$ even with smaller dataset sizes. Consequently, the FRD decreases throughout the entire training process, indicating the effectiveness of the method in improving the GAN performance. Additionally, the figure demonstrates that increasing the augmentation probability also results in a slower convergence speed. This is evidenced by the fact that the best FRD is achieved at later iterations as the r_{target} becomes smaller. However, it is important to note that further analysis is required to make a final conclusion regarding the relationship between augmentation probability and convergence speed.

From Figure 52, we can conclude that the networks trained without adaptive augmentation resulted in samples that were visually distinguishable from the real ones and the case with a dataset equal to $N = 80{,}000$, where the presence of some facies with blurred aspect are visible. Moreover, the cases with the adaptive discriminator augmentation resulted in samples that were more similar with the training and reference datasets ($N = 80{,}000$),compared to those generated by networks without any augmentation.

Figure 48 – FRD for different r_{target}. Here, ‘best’ represents the lower FRD value obtained with a constant augmentation probability (Figure 47), and ‘80k’ represents the case with dataset size equal to 80,000.

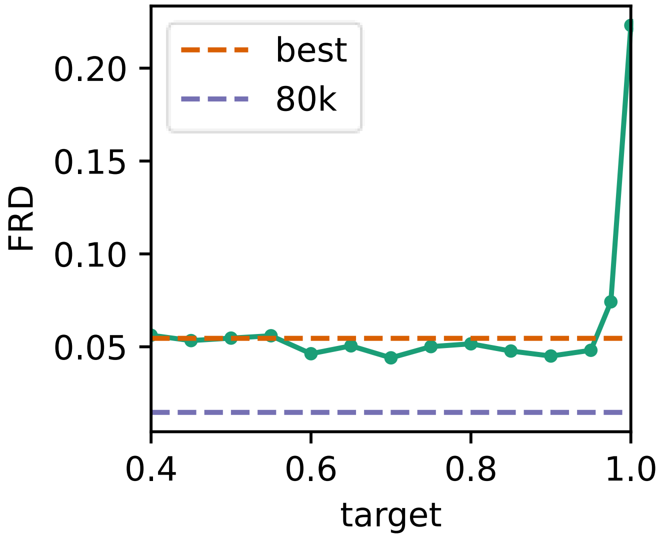


Figure 49 - r_t over iterations for different r_{target} (also represented here by the dashed lines).

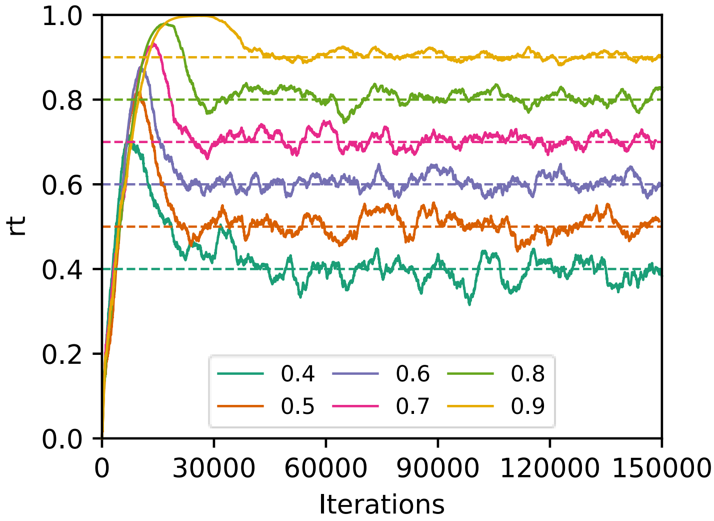


Figure 50 – Augmentation probability over iterations for different r_{target}.

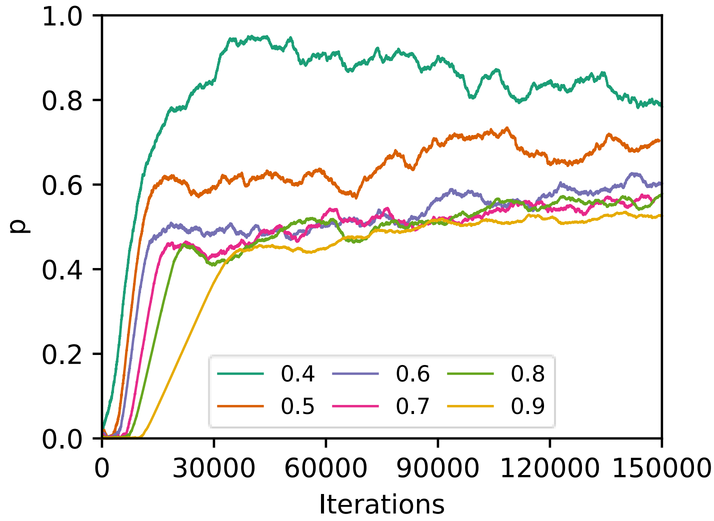
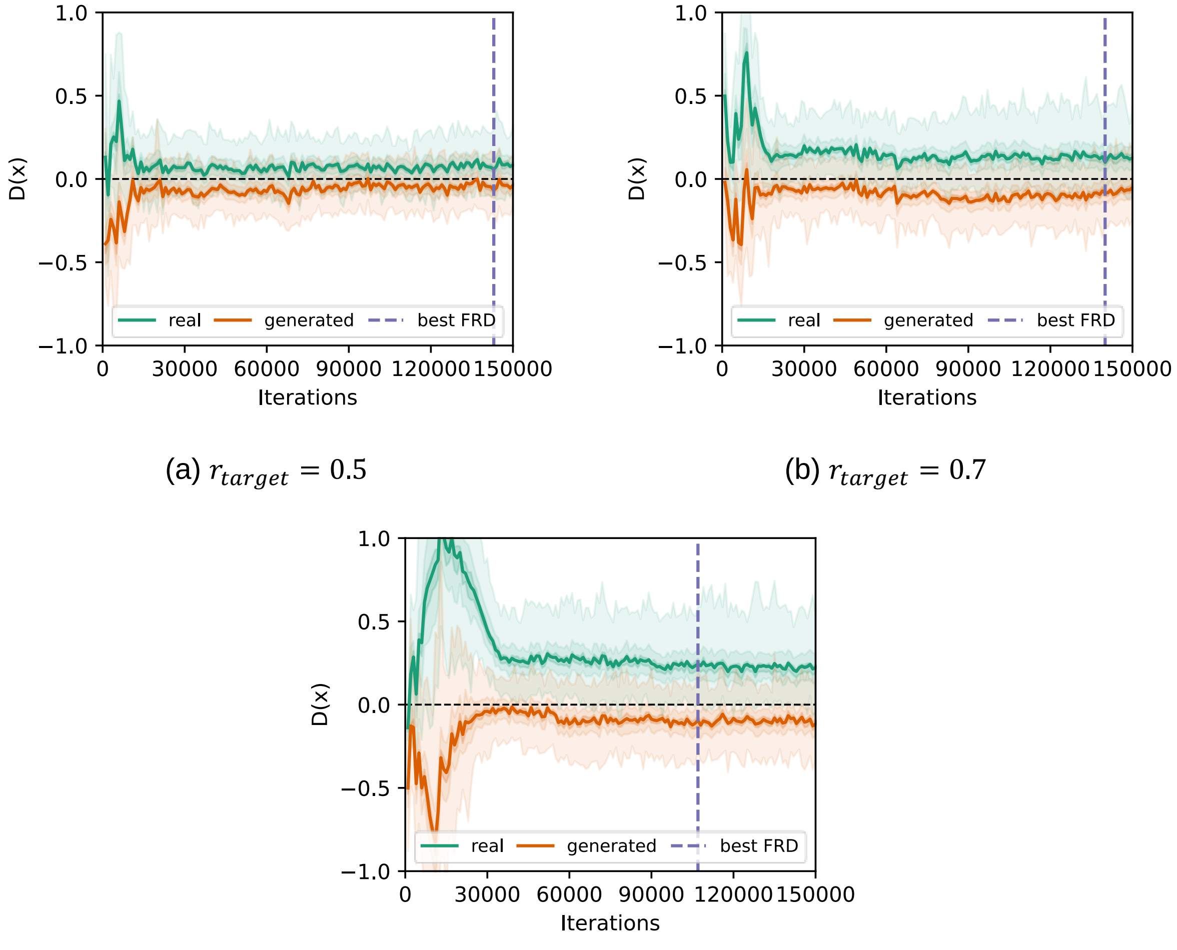


Figure 51 - Discriminator outputs for different r_{target}. For image colors description, readers are referred to Figure 46.

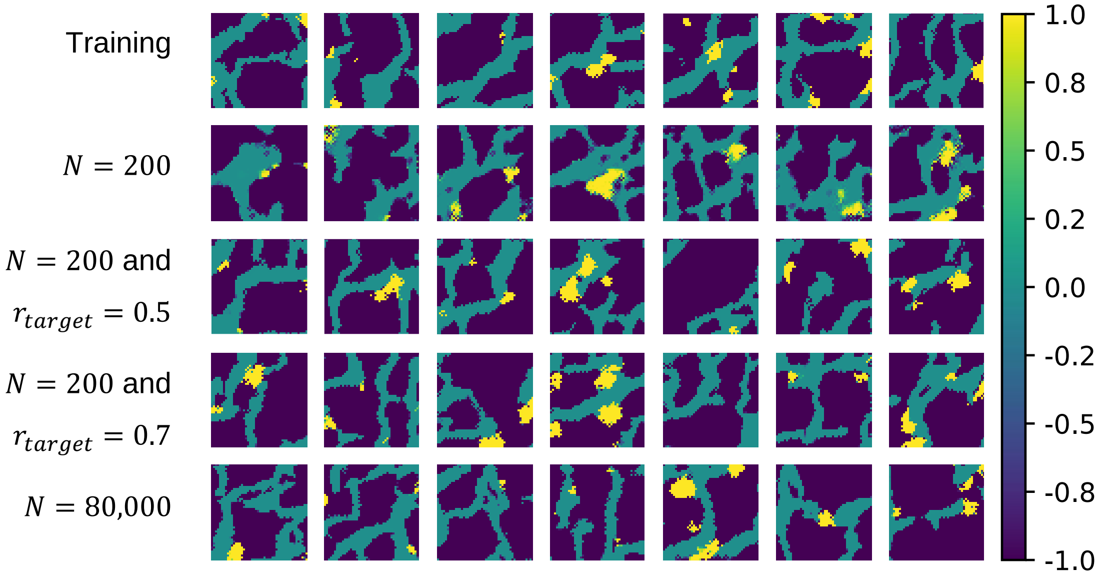


(a) $r_{target} = 0.5$

(b) $r_{target} = 0.7$

(c) $r_{target} = 0.9$

Figure 52 – Examples of realizations. Each row represents random realizations of a given dataset. 1) training; 2) $N = 200$ without data augmentation; 3) $N = 200$ with adaptive data augmentation and $r_{target} = 0.5$; 4) $N = 200$ with adaptive data augmentation and $r_{target} = 0.7$; 5) $N = 80{,}000$.



5.3 UNISIM-II

5.3.1 Generator and projection results

After the training of all networks, one for each field parameter, consistent results were observed with respect to the Fréchet Reservoir Distance (FRD), as shown in Table 7. The networks with higher FRD values can be explained by the low variability in the prior ensemble, which have the same effect than a lower dataset training size. However, the implementation of adaptive discriminator augmentation yielded satisfactory results for all properties, indicating that this approach effectively mitigates the impact of the low variability on network training. An effective way to determine whether a generator can successfully generalize the training distribution rather than simply memorize the image samples is by interpolating between two latent vectors and observing if the resulting generated samples exhibit a smooth transition between the bounds with a diverse range of outputs (BERTHELOT; SCHUMM; METZ, 2017). Here, we interpolate between two latent vectors by using the spherical linear interpolation (SLERP) instead of the traditional linear interpolation, following the approach described by White (2016). This technique involves interpolating the latent vectors along a great circle arc, rather than a straight line, resulting in a more continuous transition between the two points. Figure 54 depicts the interpolation between latent samples from some properties. The results show that all interpolations are smooth and continuous, even between two vastly different bounds, as can be seen in the net-to-gross ratio interpolation example between an image with almost zero values and an image containing higher property values. This result is particularly important in the context of data assimilation since it demonstrates the ability of the ill-posed inverse problem to search the solution in a space that is not constrained by the training dataset samples or the GAN generator itself. This behavior also can serve as a useful indicator of mode collapse in the GAN context, where the generator fails to learn the entire distribution (THANH-TUNG; TRAN, 2018; WIATRAK; ALBRECHT; NYSTROM, 2019; IGLESIAS; TALAVERA; DÍAZ-ÁLVAREZ, 2022). In extreme cases of mode collapse, the generator produces identical images, independently of the latent input. Thus, besides the FRD which measures the difference between the distributions (using the features from a classifier network), this obtained diversity in the interpolated samples indicates the generator ability to generalize and capture the variability of the training data.

Figure 53 depicts various field properties of the training dataset and the generated by the networks. The results demonstrated that the generated dataset effectively captures the essential features of the training dataset for all properties, both in terms of generating images that are visually similar to those in the original dataset and in preserving the statistical characteristics of the real dataset. These findings highlight the robustness and efficacy of the generative models in reproducing complex field properties.

Table 7 – Fréchet reservoir distance obtained for each parameter after the GAN training.

Parameter			FRD
Fracture porosity			0.0199
Permeability	Matrix	x-direction	0.0107
		z-direction	0.0082
	Fracture	x-direction	0.0162
		z-direction	0.0591
Net-to-gross ratio			0.0522
Fracture Spacing		x-direction	0.0272
		z-direction	0.0222

An effective way to determine whether a generator can successfully generalize the training distribution rather than simply memorize the image samples is by interpolating between two latent vectors and observing if the resulting generated samples exhibit a smooth transition between the bounds with a diverse range of outputs (BERTHELOT; SCHUMM; METZ, 2017). Here, we interpolate between two latent vectors by using the spherical linear interpolation (SLERP) instead of the traditional linear interpolation, following the approach described by White (2016). This technique involves interpolating the latent vectors along a great circle arc, rather than a straight line, resulting in a more continuous transition between the two points. Figure 54 depicts the interpolation between latent samples from some properties. The results show that all interpolations are smooth and continuous, even between two vastly different bounds, as can be seen in the net-to-gross ratio interpolation example between an image with almost zero values and an image containing higher property values. This result is particularly important in the context of data assimilation since it demonstrates the ability of the ill-posed inverse problem to search the solution in a space that is not constrained by the training dataset samples or the GAN generator itself. This behavior also can serve as a useful indicator of mode collapse in the GAN context, where the generator fails to learn the entire distribution (THANH-TUNG; TRAN, 2018; WIATRAK; ALBRECHT; NYSTROM, 2019; IGLESIAS; TALAVERA; DÍAZ-ÁLVAREZ, 2022). In extreme cases of mode collapse, the generator produces identical images, independently of the latent input. Thus, besides the FRD which measures the difference between the distributions (using the features from a classifier network), this obtained diversity in the interpolated samples indicates the generator ability to generalize and capture the variability of the training data.

Figure 53 – UNISIM-II-H fracture porosity (a), matrix permeability (i-direction) (b), net-to-gross ratio (c), and fracture spacing (z-direction) (d) for the training (upper row) and generated (bottom row) datasets. Here, left column shows the mean over the dataset, the mid column displays the standard deviation, and the right column shows a random realization for both datasets. Here, all properties were normalized between -1 and 1.

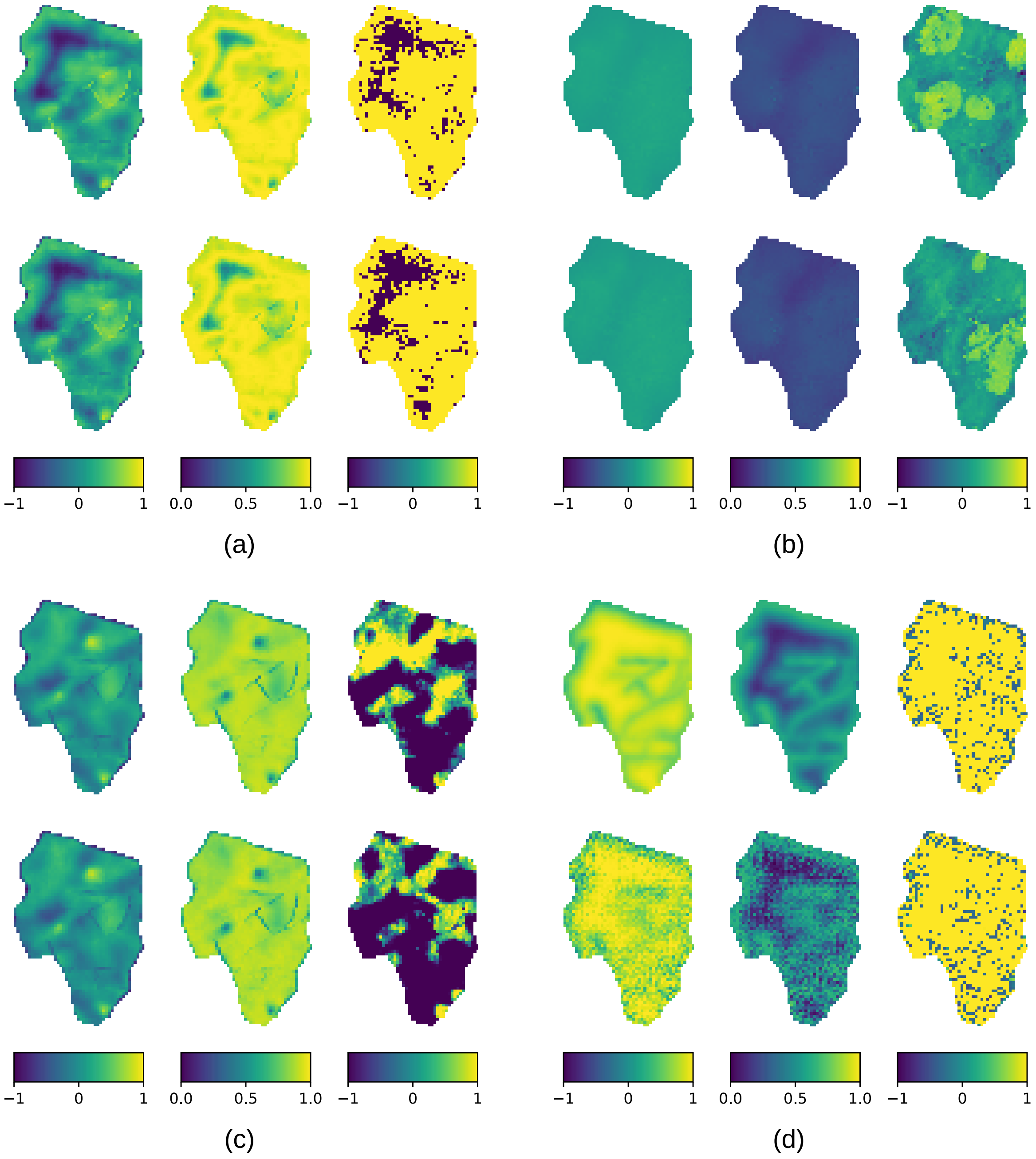
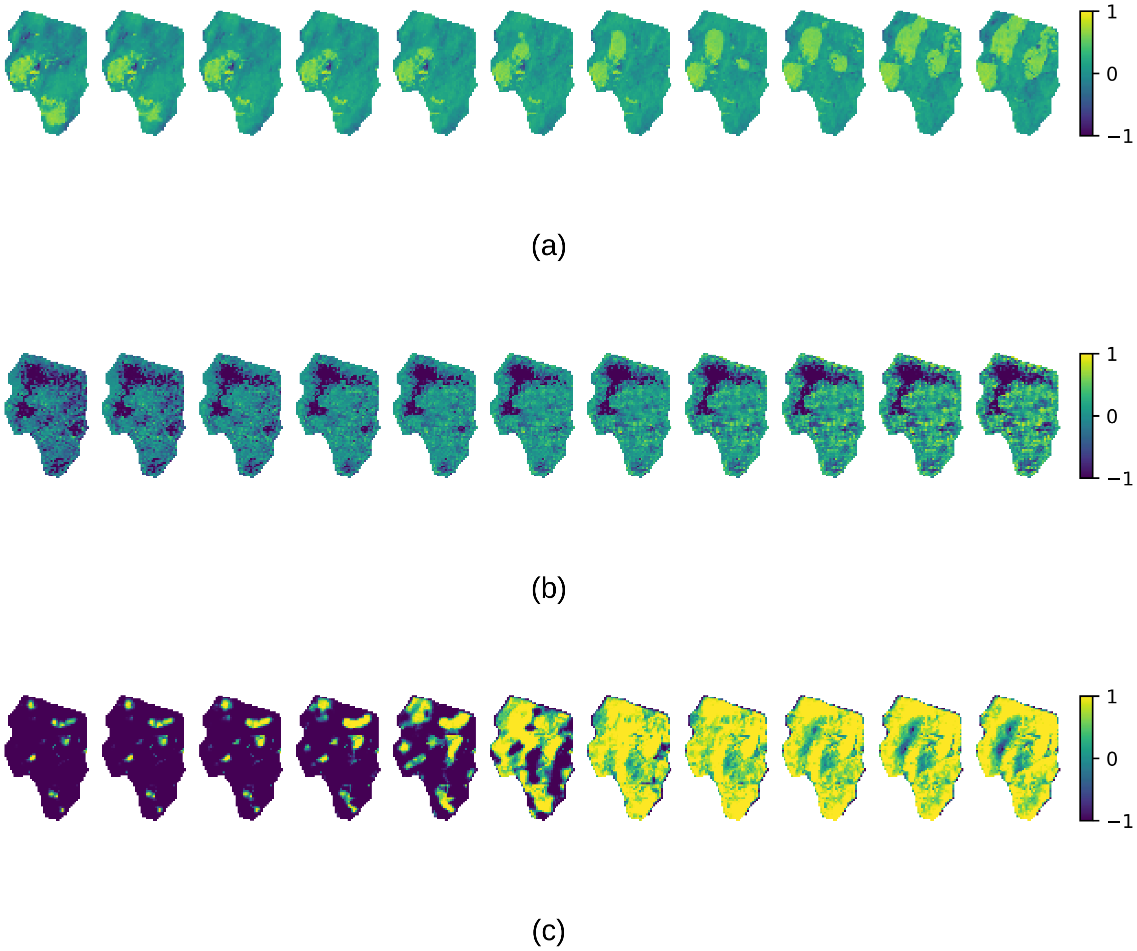


Figure 54 – Interpolations of two generated images (represented here by the images at the corners). Here, upper row shows the matrix permeability (i-direction) (a), mid row shows the fracture permeability (z-direction) (b), and bottom row shows the net-to-gross ratio (c). Here, all properties were normalized between -1 and 1.



(a)

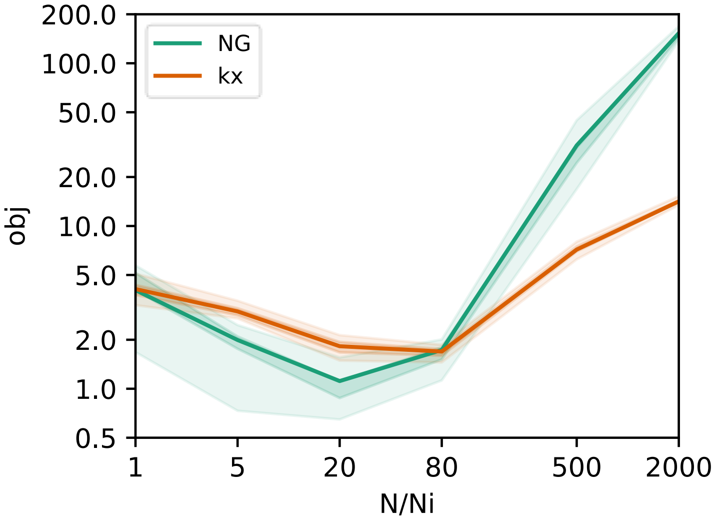
(b)

(c)

5.3.2 ES-MDA latent space projection

Firstly, we conducted an investigation to determine the ensemble size and the number of iterations for projecting the real samples onto the latent space. Starting by fixing the total number of latents passed through the generator by 2000, we projected a fixed number of 15 images onto the latent space with different N/N_i ratio. Figure 55 presents the statistics of the mean objective function defined in (29), for all projected samples of two different reservoir properties. Here, in higher N/N_i values, the main source of errors is the dominance of nonlinearities of the forward model in the update, while in lower N/N_i values, the main source of errors is spurious correlations due to the small ensemble size. Therefore, a tradeoff between these two sources with the optimal objective function results can be observed when $20 < N/N_i < 80$. In the subsequently experiments, we decided for a high ensemble size to minimize the effect of spurious correlations. Consequently, all next experiments were performed with $N = 400$ and $N_i = 5$.

Figure 55 – Objective function percentiles (transparent lines) and mean (solid line) for a fixed total number of forward runs equal to 2,000 runs, for matrix permeability (kx) and net-to-gross ratio (NG). Here, N/N_i represents the ratio between the ensemble size N and the total number of iterations N_i.



After applying the iterative ensemble smoother, the method successfully identified a latent representation that adequately represents the real sample for all prior ensemble and for all reservoir properties. This can be observed in Figure 56, which shows the boxplots of the mean objective function for all 15,000 samples over the iterations. To save computational time, the last forward step with the updated latent ensemble was not performed. Thus, horizontal axis shows only the 5 update steps.

Figure 56 – Boxplot of the mean objective function for all projected images. (a) matrix permeability, (b) net-to-gross ratio.

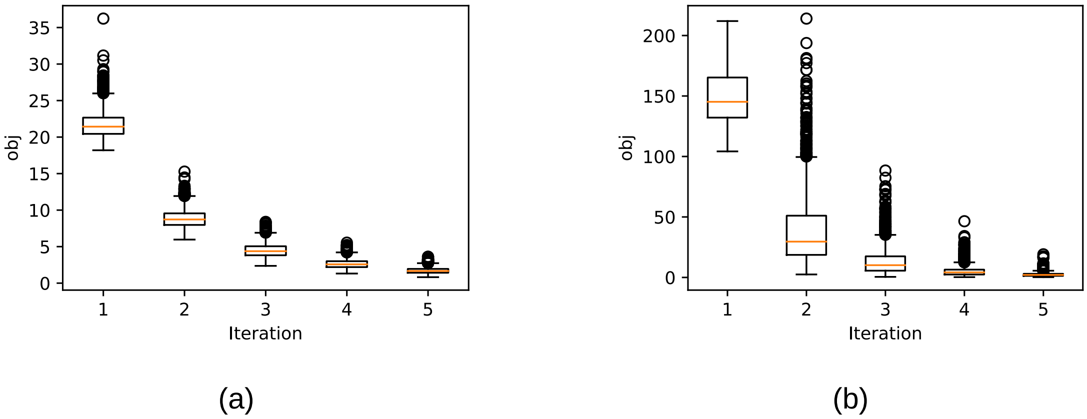


Figure 57 and Figure 58 compare two randomly selected realizations with the generator outputs of their respective projections, while Figure 59 and Figure 60 show some examples of the ensemble variogram maps in both horizontal and vertical directions for both the original and the projected ensembles. The images demonstrate that the original and projected ensembles are almost indistinguishable, indicating that the projection method is effective in preserving the original spatial structure of the initial ensemble.

Figure 57 – Vertical layer 12 (columns 1, 3) and layer 35 j-sec (columns 2, 4) matrix permeability for two real random realizations (first row) and the samples generated from its respective latent projections (bottom row). Here, all properties were normalized between -1 and 1.

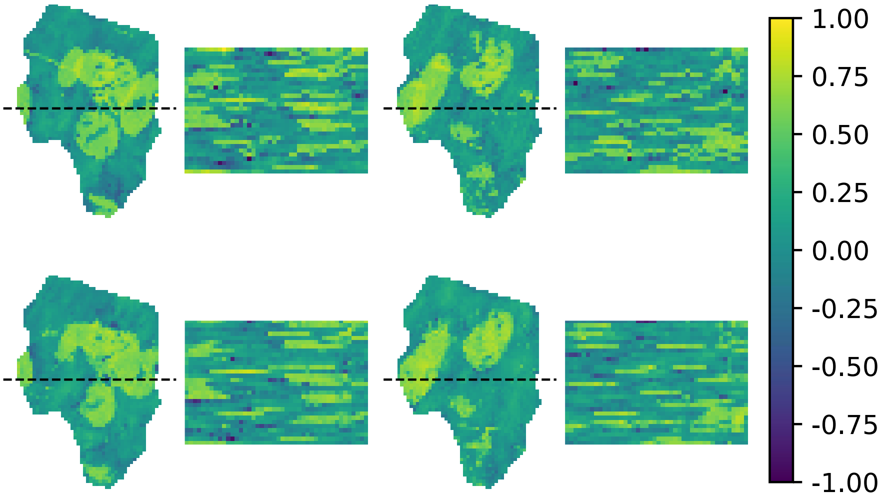


Figure 58 – Vertical layer 12 (columns 1, 3) and layer 23 i-sec (columns 2, 4) net-to-gross ratio for two real random realizations (first row) and the samples generated from its respective latent projections (bottom row). Here, all properties were normalized between -1 and 1.

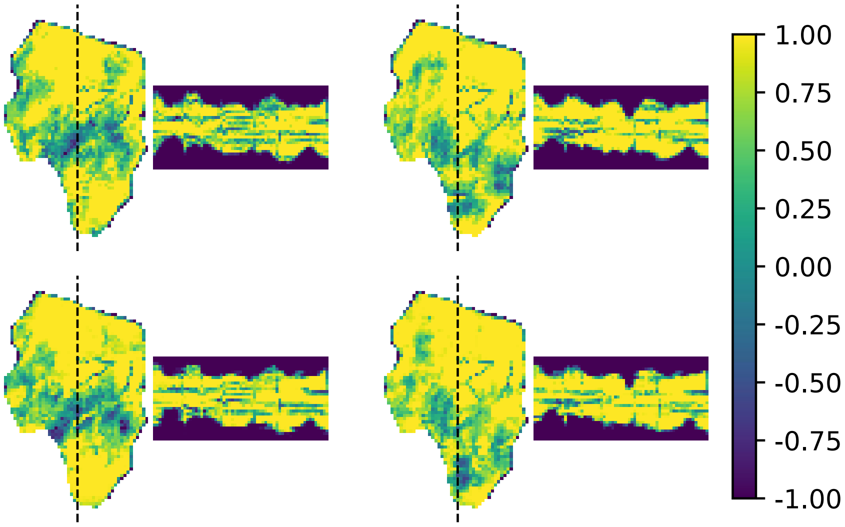


Figure 59 - Vertical layer 12 (left column) and layer 23 i-sec (right column) net-to-gross ratio variogram maps for the real dataset (upper row) and the ensemble generated from the projected latent vector (bottom row).Reference points to compute the variogram map is $[25, 25]$ for the left column and $[15, 15]$ for the right column.

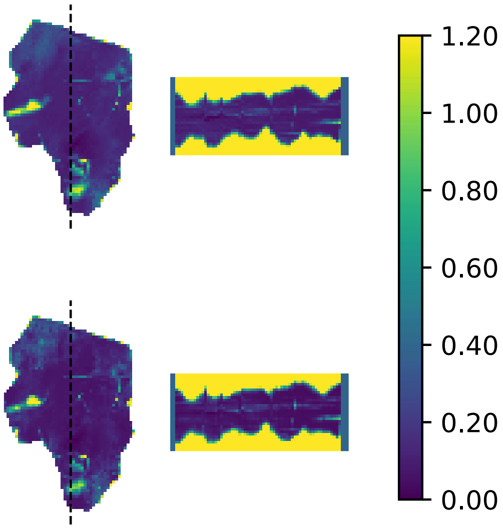
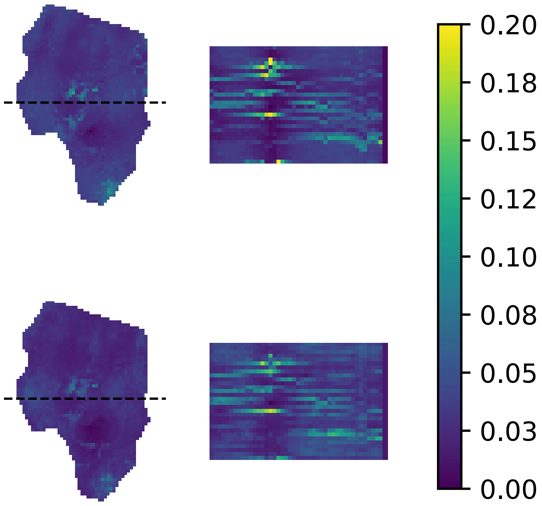


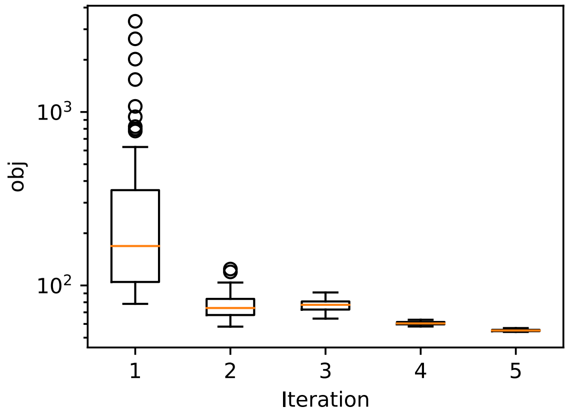
Figure 60 – Vertical layer 12 (left column) and layer 35 j-sec (right column) matrix permeability variogram maps for the real dataset (upper row) and the ensemble generated from the projected latent vector (bottom row). Reference points to compute the variogram map is $[25, 25]$ for the left column and $[15, 15]$ for the right column.



5.3.3 Data assimilation results

The results of the data assimilation performed using an ensemble size of only 50 members (considered too small) yielded satisfactory results, indicating the effectiveness of the proposed methodology. Figure 61 shows the boxplot of the normalized objective function across the iterations (including the last forward step), where a significant reduction in the objective function values is observed.

Figure 61 – Boxplot of O_{N_d} across the iterations.



The data assimilation process led to an ensemble collapse of the posterior ensemble, as evidenced by Figure 62 through Figure 65 (this can be attributed to the small ensemble size of 50 members used in the assimilation). However, despite this behavior, an interesting observation was made regarding the proposed parameterization method. Specifically, the posterior ensemble members demonstrated geological features that were consistent with the prior ensemble, despite the occurrence of ensemble collapse. This is noteworthy, as ensemble collapse typically leads to the posterior ensemble displaying extreme parameter values. Some examples of members of the posterior ensemble are shown in Figure 66 through Figure 70. Furthermore, the methodology of considering each vertical reservoir layer separately during the parameterization, in addition to the initial ensemble projection, proved effective in maintaining satisfactory vertical correlation after the assimilation process, as demonstrated by the posterior ensemble results. The effect of the adaptive localization can be observed by comparing the prior and posterior relative permeability curves, which were generated from non-local uncertain parameters. As shown in Figure 71, the POL-GC adaptive localization method was effective in preventing ensemble collapse in non-local parameters.

Figure 62 – Fracture porosity standard deviation of prior (upper row) and posterior (bottom row) ensembles.

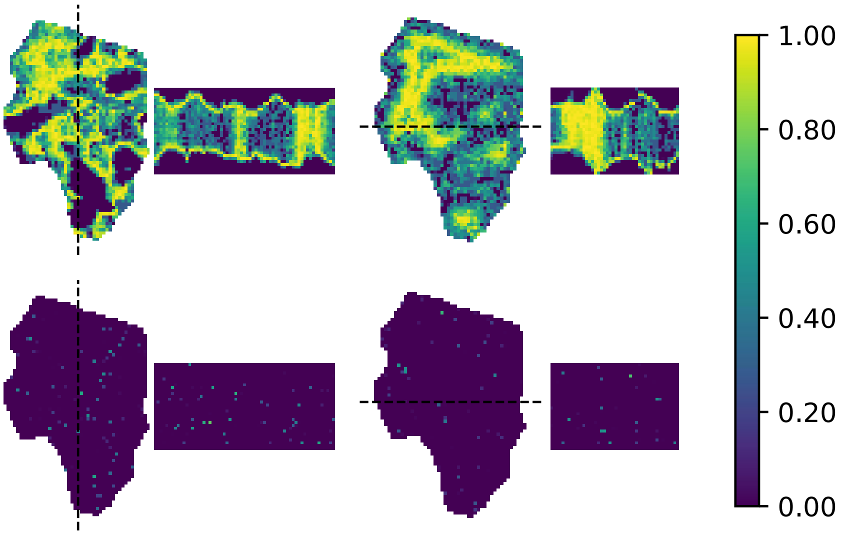


Figure 63 – Matrix permeability (i-direction) standard deviation of prior (upper row) and posterior (bottom row) ensembles.

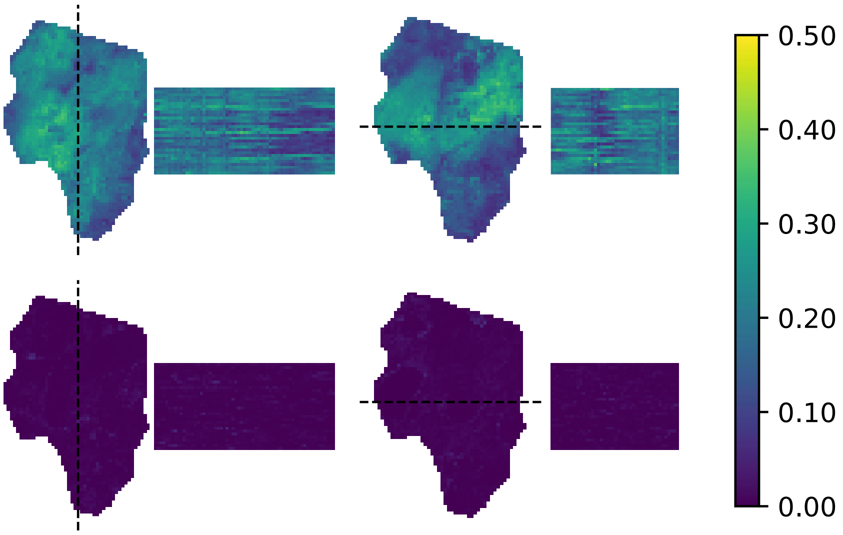


Figure 64 - Fracture permeability (i-direction) standard deviation of prior (upper row) and posterior (bottom row) ensembles.

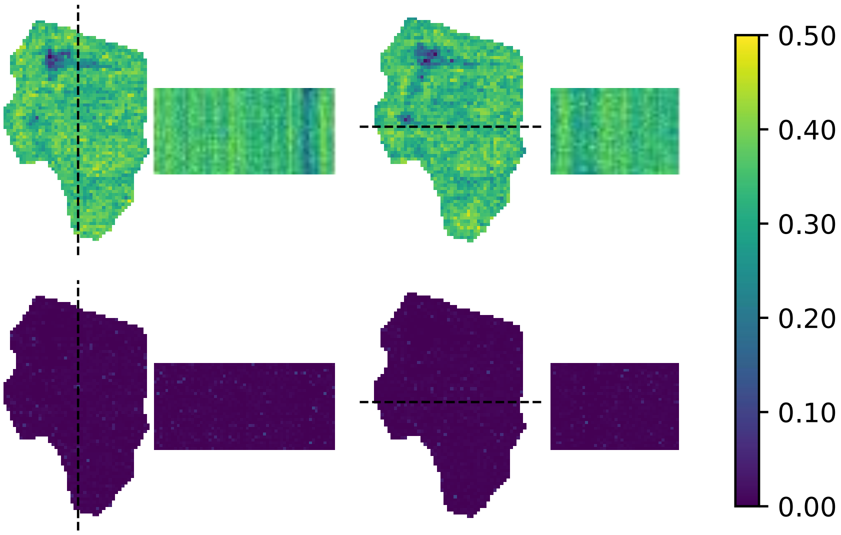


Figure 65 – Fracutre spacing (i-direction) standard deviation of prior (upper row) and posterior (bottom row) ensembles.

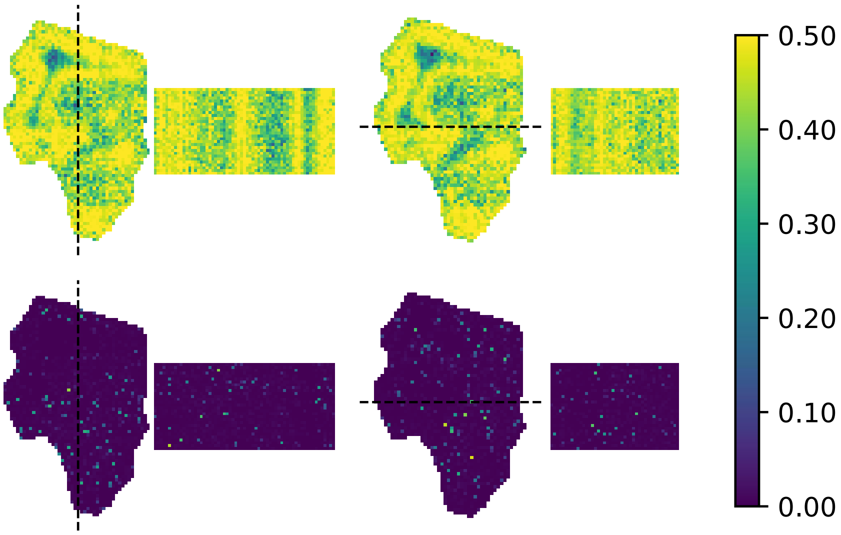


Figure 66 – Fracture Porosity first ensemble member prior (upper row) and posterior (bottom row) the data assimilation. First column represents vertical layer 5, second column layer 23 i-sec, third column vertical layer 12, and forth column layer 35 j-sec. The dashed lines in vertical layers represent the cross sections. Here, all properties were normalized between -1 and 1.

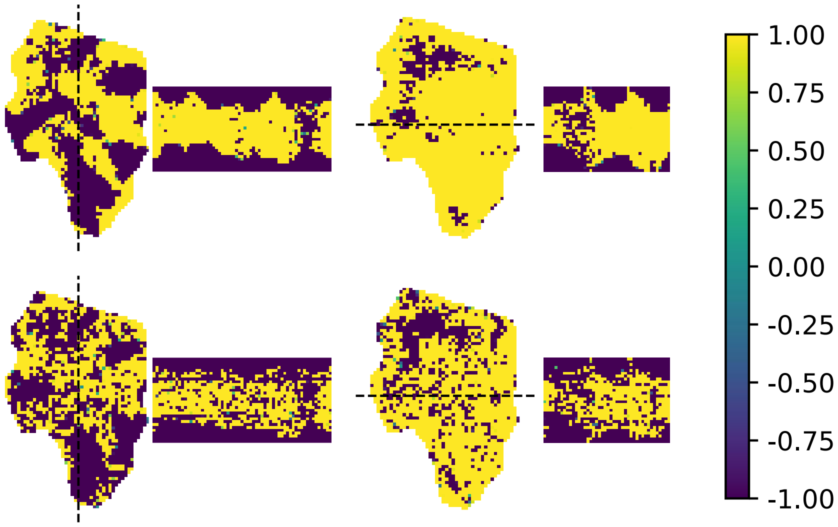


Figure 67 – Matrix permeability (i-direction) first ensemble member prior (upper row) and posterior (bottom row) the data assimilation. First column represents vertical layer 5, second column layer 23 i-sec, third column vertical layer 12, and forth column layer 35 j-sec. The dashed lines in vertical layers represent the cross sections. Here, all properties were normalized between -1 and 1.

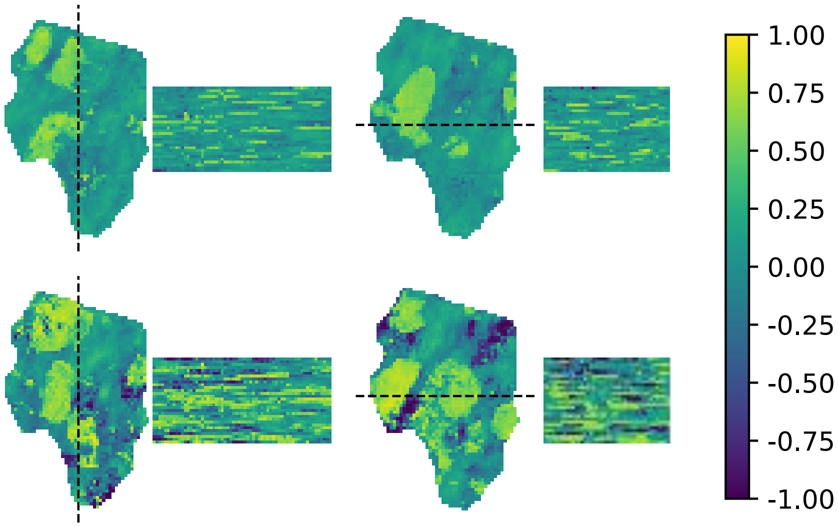


Figure 68 - Fracture permeability (i-direction) first ensemble member prior (upper row) and posterior (bottom row) the data assimilation. First column represents vertical layer 5, second column layer 23 i-sec, third column vertical layer 12, and forth column layer 35 j-sec. The dashed lines in vertical layers represent the cross sections. Here, all properties were normalized between -1 and 1.

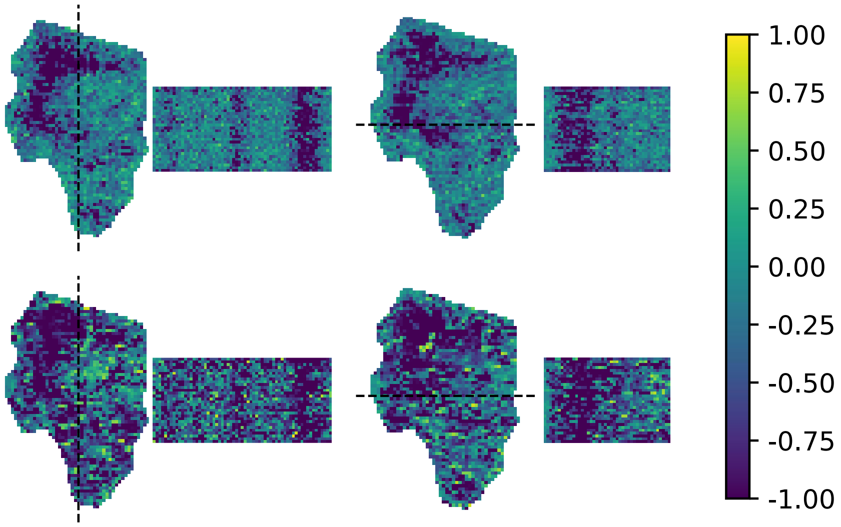


Figure 69 – Net-to-gross ratio first ensemble member prior (upper row) and posterior (bottom row) the data assimilation. First column represents vertical layer 5, second column layer 23 i-sec, third column vertical layer 12, and forth column layer 35 j-sec. The dashed lines in vertical layers represent the cross sections. Here, all properties were normalized between -1 and 1.

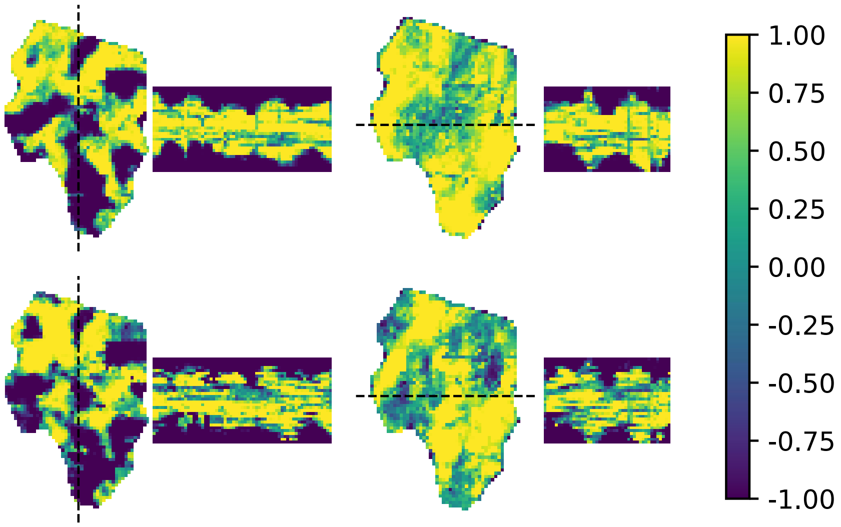


Figure 70 – Fracture spacing (i-direction) first ensemble member prior (upper row) and posterior (bottom row) the data assimilation. First column represents vertical layer 5, second column layer 23 i-sec, third column vertical layer 12, and forth column layer 35 j-sec. The dashed lines in vertical layers represent the cross sections. Here, all properties were normalized between -1 and 1.

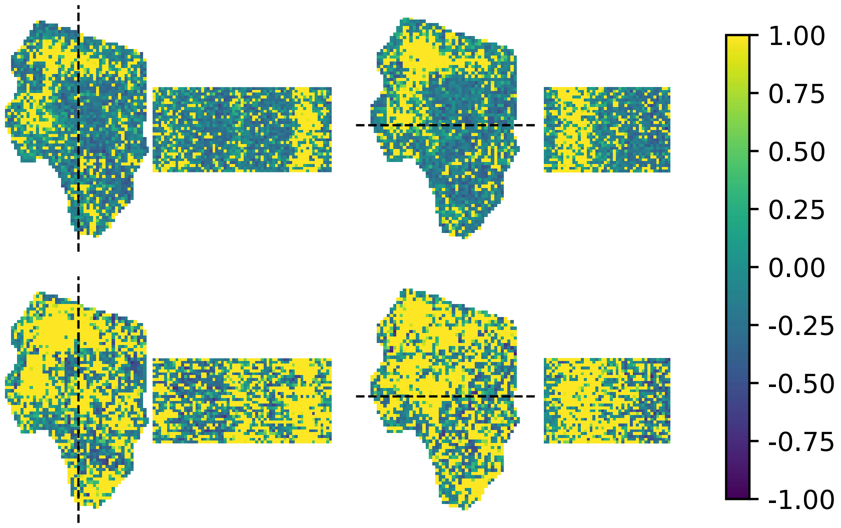
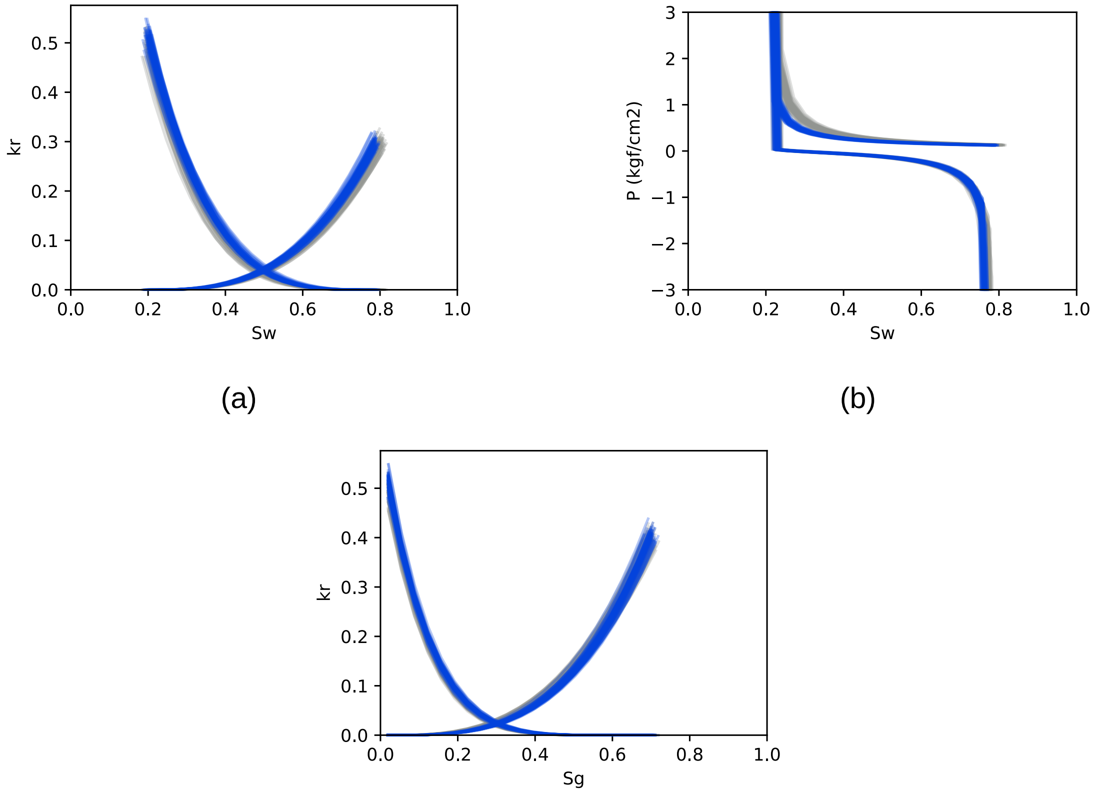


Figure 71 – Relative permeability and capillary pressure curves. (a) Water relative permeability curves, (b) Capillary curves, (c) Gas relative permeability curves. Gray lines represent the prior ensemble, while blue lines represent the posterior ensemble.



(a)

(b)

(c)

Despite achieving a significant reduction in data mismatch, the time series figures (Figure 72 and Figure 73) suggest that the observed production data cannot be fully matched in some wells. This discrepancy can be attributed to several factors. Firstly, the producer well Prod-6 is a critical well to match production data, and its behavior poses challenges in ensemble-based methods. The prior ensemble fails to adequately cover the measurements associated with the particular well. Maschio and Schiozer (2019) extensively discuss the influence of Prod-6 on data assimilation outcomes. Another possible explanation is that parameterization increases the complexity and nonlinearity of the forward problem, represented now as $\mathbf{d} = g(\mathcal{G}(\mathbf{z}))$, where an increase in $\mathbf{z}$ may not necessarily lead to an increase in $\mathbf{d}$, representing a non-monotonic behavior. Evensen (2018) provides a detailed discussion of the behavior of ensemble-based methods in such models. Although, an increase number of iterations steps can help address this issue, the time-consuming forwards in complex reservoirs can be prohibitive. Regarding the adaptive localization, compute this only with the prior ensemble might amplify the errors, especially in small ensemble sizes where the sample covariances are found in the threshold level for almost all uncertain parameters, resulting in excessive damping. However, as can be observed in Figure 74, an excessive damping in the latent ensemble does not mean that the parameter ensemble will suffer from excessive damping. Both behaviors can be verified by interpolating a vector with constant elements $\tilde{z} = [v_1, v_2, \ldots, v_{N_z}]^{\mathrm{T}}$, maintaining a given fraction of the realization unchanged (equal to minus one), simulating the localization process. Figure 74 shows the extrapolation between two $\tilde{z}$ vectors using as the extreme values of -1 and 1 as bounds. Each row represents this interpolation only between a fraction of the latent vector, with the rest kept unchanged. The image shows that Super-k appear in the same region of the realization at the latent limits, and tapering a higher number of parameters does not guarantee small changes in the original domain.

Figure 72 – Time series of production data. OPR refers to oil production rate (first column), WPR refers to water production rate (second column), BHP refers to well bottom-hole pressure (third column). Each row represents a single well: 1) Prod-2, 2) Prod-4, 3) Prod-6, 4) Prod-8, 5) Wildcat. Here, the gray lines represent the prior ensemble, blue lines represent the posterior ensemble, and the red dots represent the measurements.

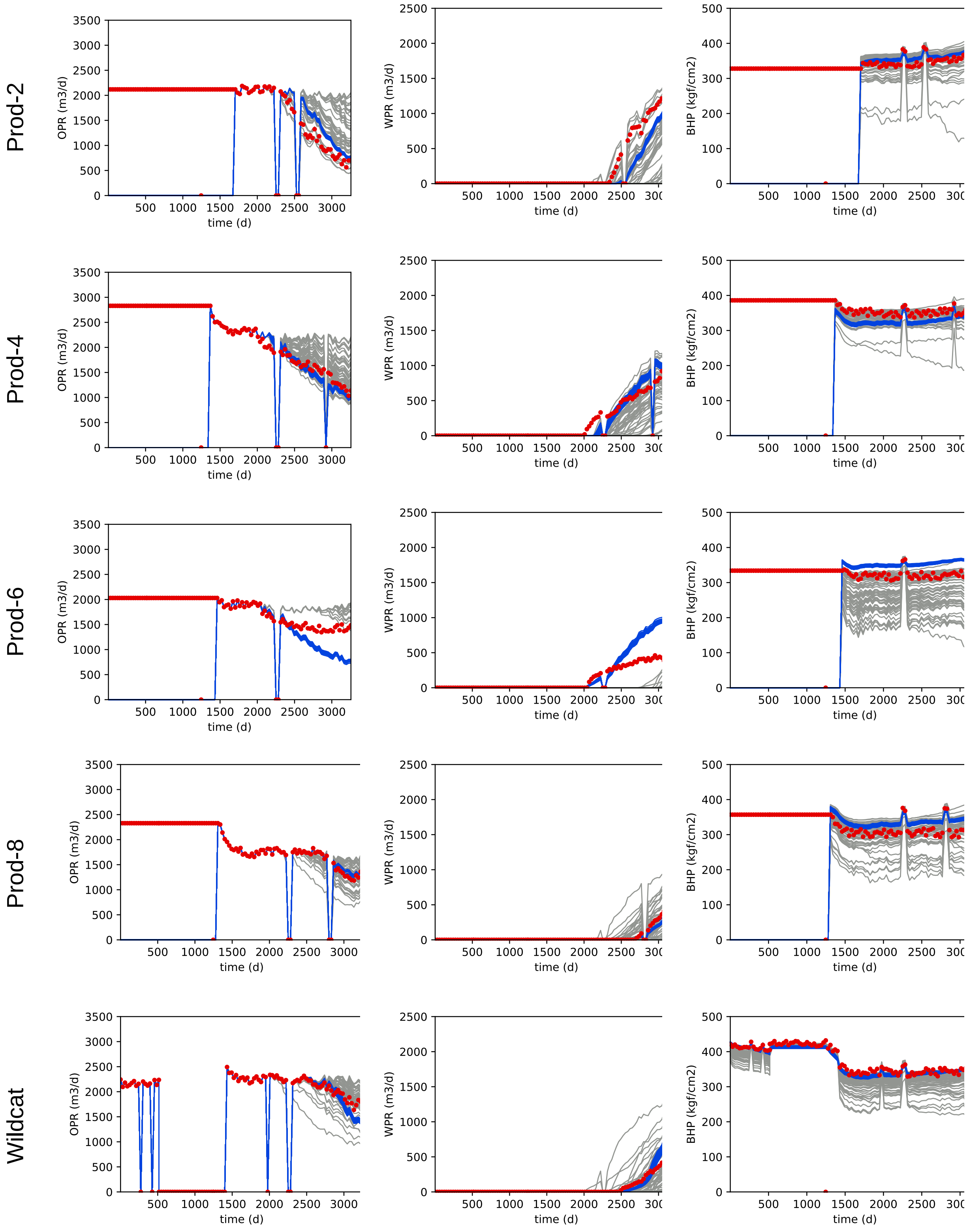


Figure 73 – Time series of injection data. WIR refers to water injection rate (first column), BHP refers to well bottom-hole pressure (second column). Each row represents a single well: 1) Inj-1, 2) Inj-3, 3) Inj-5, 4) Inj-7, 5) Inj-9. Here, the gray lines represent the prior ensemble, blue lines represent the posterior ensemble, and the red dots represent the measurements.

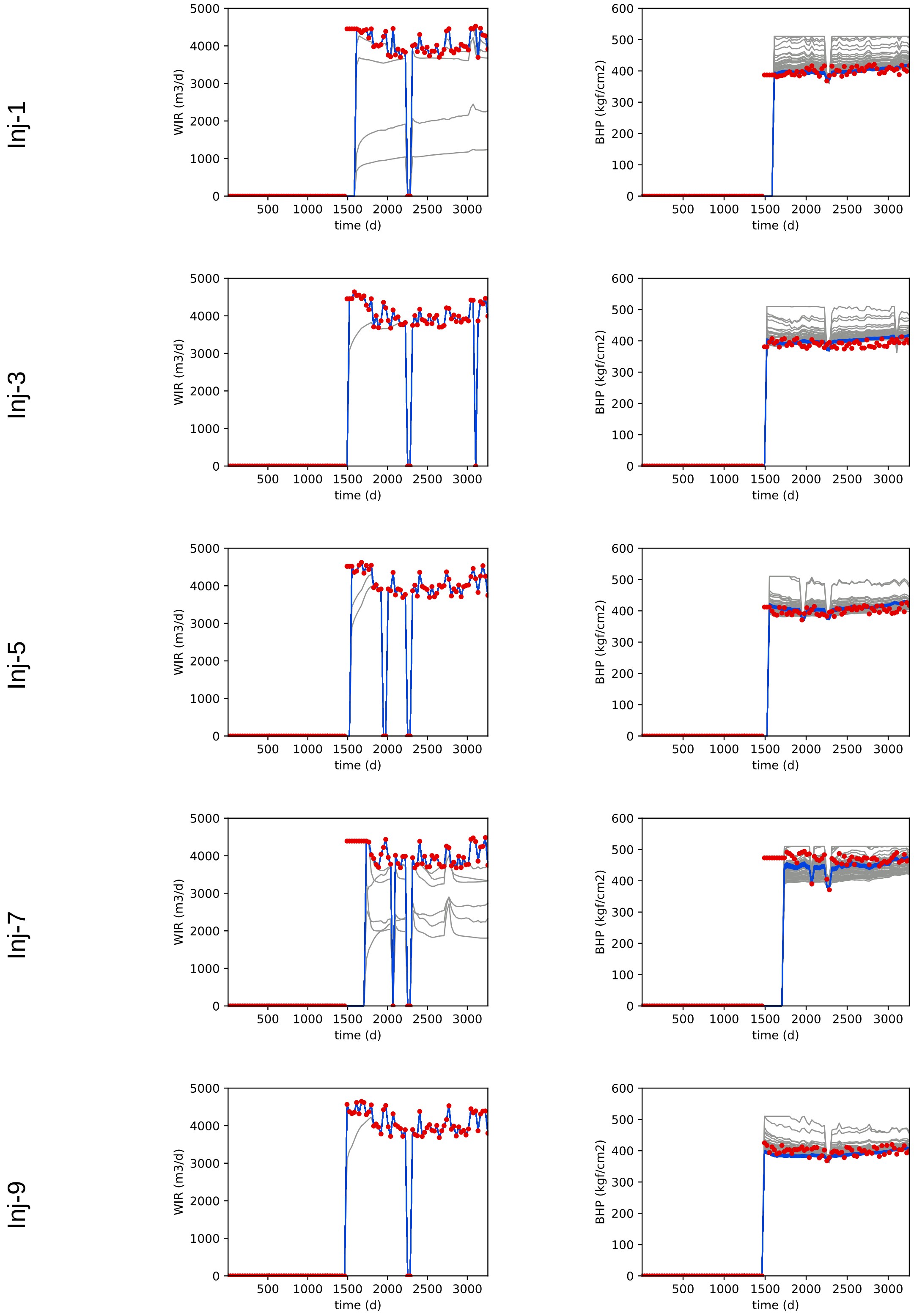
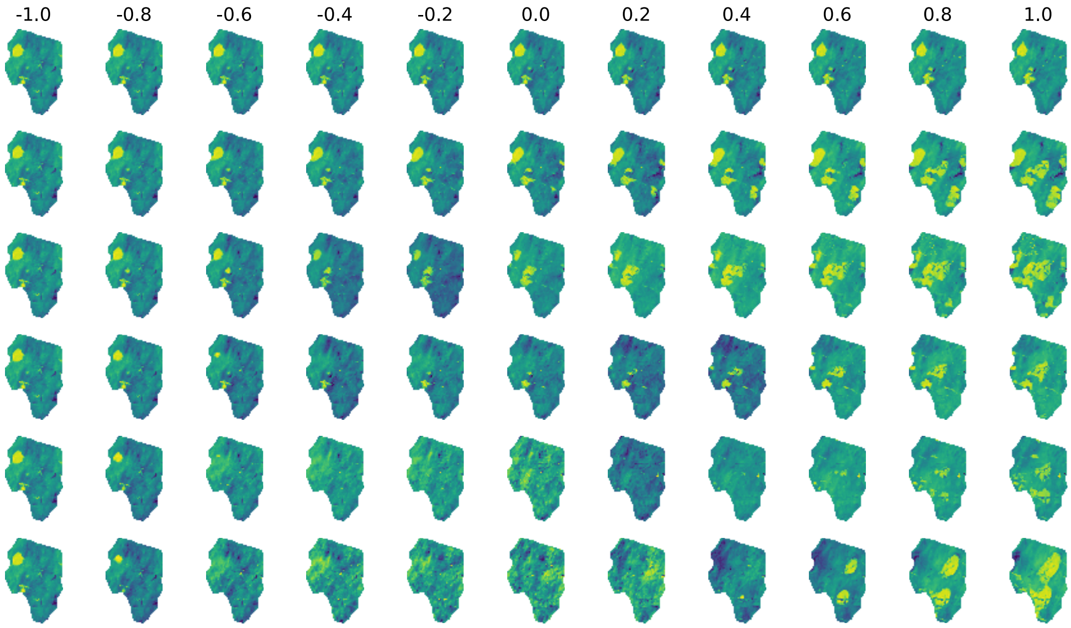


Figure 74 - Interpolation with constant values for the latent vector. Each row represents a fraction k of the parameters kept unchanged during the interpolation: 1) $k = 97\%$, 2) $k = 80\%$, 3) $k = 60\%$, 4) $k = 40\%$, 5) $k = 20\%$, 6) $k = 0\%$. Here, all properties were normalized between -1 and 1.



CHAPTER 6 - CONCLUSIONS

6.1 Localization

One of the main issues of the iterative ensemble smoother is the loss of variance and filter divergence due to sampling errors caused by the limited ensemble size. The main method in petroleum literature to reduce this effect is the application of distance-based Kalman gain localization. However, several issues regarding this kind of methodology arise, making it suboptimal or impractical in many cases. In this present work, we proposed an alternative adaptive pseudo-optimal localization method with the addition of a penalty factor similar to the noise level applied in the correlation-based localization method. Following the ideas of Lacerda et al. (2019), which demonstrated that a weaker correlation suffers more strongly the effects of the limited ensemble size, we proposed four distinct options to compute this penalty factor, based on the sample correlation between each parameter and data. Also, we proposed an alternative method to compute the threshold values. This might be useful to determine the threshold values for non-local parameters of parameters that might not be grouped. Moreover, the application of the penalty rule in the pseudo-optimal localization method resulted in a fully automatic localization method that does not require any prior tuning.

We applied the proposed method to two case studies: a 2D ¼ five-spot, and the large-scale UNISIM-I-H benchmark. The results showed a significantly higher posterior ensemble variance (with relatively small poorer data-matchings) to the distance-based and the standard pseudo-optimal methods. It is important to point out that these results were obtained even without the optimal definition of the penalty rule computation. Its development will be the subject of further investigation.

6.2 Parameterization

Parameterization of petroleum reservoir is a challenging task that has been the subject of intense research in recent years. One promising approach is the use of generative adversarial networks (GAN). While this approach has shown great potential, there are still several points that need to be improved upon. For instance, there is a need to develop reliable methods for measuring the quality of the trained generator in the reservoir context, evaluate the performance of GAN-based parameterization methods in other case studies beyond the well-studied two-facies channelized reservoir, and the limitations posed by the limited dataset size during the GAN training. Addressing these challenges will be important for advancing the field and making GAN-based parameterization a more widely applicable technique.

We applied a robust and stable GAN by using the R1 regularization technique, which has some advantages over the Wassertein GAN with gradient penalty (WGAN-GP). Unlike WGAN-GP, GAN-R1 adopts a zero-centered regularization approach and does not require training the generator and the discriminator at different rates (such as using different learning rates or training the discriminator with more iteration steps). The experimental results exhibited significant improvements compared to the standard GAN. Moreover, we have found that the recommended regularization weight in our case is much higher than previously suggested, likely due to the discriminator’s ease in distinguishing between real and generated samples.

In addition, we proposed a novel approach to measuring the difference between two ensembles of realizations. From previous works that replaced the Inception-v3 network in the computation of the Fréchet Inception Distance (FRD), we developed a Reservoir Classifier Network and leveraged the feature extraction component of this network to compute the Fréchet distance, thus creating a new metric, the Fréchet Reservoir Distance. The results demonstrated that our proposed metric is effective and significantly reduces the computation time required to evaluate the distance between two distributions.

Finally, our study applied the adaptive data augmentation (ADA) method in training a GAN using a limited ensemble size of 200 petroleum reservoir realizations. First, different augmentation categories were evaluated, including value augmentations designed specifically to this problem. The cumulative augmentation consisting of blitting and geometric transformations was found to yield the best FRD results. Moreover, results showed good consistency between the network trained with limited data plus the adaptive data augmentation and the reference case considering a dataset with 80,000 samples, highlighting the potential of ADA in addressing limited dataset size challenges in parameterization of petroleum reservoir samples.

6.3 Two-step data assimilation applied in the benchmark UNISIM-II-H

The application of data assimilation techniques in complex reservoirs models presents several challenges. In this study, we addressed these challenges by proposing a novel approach that combines the projection of original samples into the latent domain and the data assimilation into a two-step ES-MDA framework. The main goal was to improve the assimilation process and enhance the accuracy of the results in large-scale reservoir system. To handle three-dimensional reservoirs, we employ a mapping strategy that transforms each horizontal layer into the latent domain, resulting in a $N_z \times N_k$ matrix. Here, N_z represents the dimension of the latent domain, and N_k represents the vertical reservoir dimension. These layers are then updated using the analysis step of ensemble-based methods, and the updated ensemble is subsequently mapped back to the original three-dimensional domain. The application of adaptive data augmentation facilitated the training of distinct networks for each reservoir property, including those with limited variability such as fracture spacing. Furthermore, the successful projection of the initial ensemble into the latent space using ensemble-based methods was demonstrated, as the results indicated a high level of similarity between the ensemble projected from the initial latent space and the original ensemble.

The effectiveness of this method relies on preserving the vertical correlation when mapping the initial ensemble to the latent domain. Our results demonstrate a satisfactory reduction in data mismatch, with the data-mismatch values (30) decreasing from approximately 500 to 40. Regarding the non-local parameters (except latent parameters), the adaptive localization method was able to avoid ensemble collapse. However, the most notable outcome of our study is that despite using a small ensemble size with only 50 members that exhibits ensemble collapse in posterior ensemble, the generated posterior realizations maintained geological realism and vertical continuity. This finding underscores the robustness of our approach in capturing key geological features even under challenging conditions. The results highlight the potential of the method in improving the accuracy of reservoir characterization and provide a foundation for further advancements in the field of reservoir parameterization and data assimilation.

6.4 Broader discussion

In the field of data assimilation for petroleum reservoirs, there remains a significant disparity between the techniques that have been developed and validated for small reservoirs and their application in large-scale models with numerous uncertain properties. Additionally, the computational resources required to execute forward models of reservoirs are limited, which restricts the feasible ensemble size for conducting data assimilation. This study aims to bridge this gap by introducing innovative methodologies within the context of a large-scale carbonate reservoir. The outcome is a robust workflow that successfully performs data assimilation and yields reasonable results in the context of the geological realism preservation. It is desirable that the methodology will facilitates the application of ensemble-based data assimilation in other large-scale reservoir models, thus addressing a critical need in the field.

6.4.1 Contribution of this thesis

Here, we summarize the key contributions of this work:

- Novel adaptive localization method: we proposed a novel approach by combining the pseudo-optimal localization method with the random-shuffle method. This combination resulted in a new equation for computing tapering values, particularly when dealing with non-distance parameters or data in the update process.
- GAN framework: we assessed the effectiveness of a novel GAN framework utilizing the R1 regularization during the training. This regularization technique proved to be beneficial in enhancing the stability and performance of the GAN when training reservoir samples.
- Adaptive Data Augmentation: we introduced the application of adaptive data augmentation specifically tailored for reservoir realizations. This technique dynamically adjusts the augmentation power across the training process, resulting in improved training and generation of realistic samples.
- Novel discrepancy metric: we proposed a new metric to measure the discrepancy between two ensembles of samples. This metric, called the Fréchet Reservoir Distance, provides a reliable measure of similarity between

reservoir realizations, and facilitates the evaluation of GAN training performance, requiring much lower computation effort in comparison with the original method.

- Ensemble-based projection method: we introduced a novel method for projecting samples onto the latent space. This approach enables efficient and accurate mapping of reservoir properties from the original domain to the latent space.
- Two-Step ES-MDA: we developed a two-step Ensemble Smoother with Multiple Data Assimilation (ES-MDA) specifically designed for three-dimensional reservoirs. This approach makes possible the application of parameterization in complex reservoir systems.

REFERENCES

AANONSEN, Sigurd I.; NÆVDAL, Geir; OLIVER, Dean S.; REYNOLDS, Albert C.; VALLÈS, Brice. The Ensemble Kalman Filter in Reservoir Engineering - a Review. **SPE Journal**, v. 14, n. 03, p. 393–412, 2009. DOI: 10.2118/117274-PA.

ABDAL, Rameen; QIN, Yipeng; WONKA, Peter. **Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?**, arXiv preprint, 2019. DOI: 10.48550/ARXIV.1904.03189.

ALQAM, M. H.; NASR-EL-DIN, H. A.; LYNN, J. D. Treatment of Super-K Zones Using Gelling Polymers. *Em*: ALL DAYS 2001, Houston, Texas. **Anais**. *In*: SPE INTERNATIONAL SYMPOSIUM ON OILFIELD CHEMISTRY. Houston, Texas: SPE, 2001. p. SPE-64989-MS. DOI: 10.2118/64989-MS.

ANDERSON, Jeffrey L. An adaptive covariance inflation error correction algorithm for ensemble filters. **Tellus A: Dynamic Meteorology and Oceanography**, v. 59, n. 2, p. 210, 2007. DOI: 10.1111/j.1600-0870.2006.00216.x.

ANDERSON, Jeffrey L. Spatially and temporally varying adaptive covariance inflation for ensemble filters. **Tellus A**, v. 61, n. 1, p. 72–83, 2009. DOI: 10.1111/j.1600-0870.2008.00361.x.

ANDERSON, Jeffrey L. Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation. **Monthly Weather Review**, v. 140, n. 7, p. 2359–2371, 2012. DOI: 10.1175/MWR-D-11-00013.1.

ANDERSON, Jeffrey L. Reducing Correlation Sampling Error in Ensemble Kalman Filter Data Assimilation. **Monthly Weather Review**, v. 144, n. 3, p. 913–925, 2016. DOI: 10.1175/MWR-D-15-0052.1.

ANDERSON, Jeffrey L.; ANDERSON, Stephen L. A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. **Monthly Weather Review**, v. 127, n. 12, p. 2741–2758, 1999. DOI: 10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.

ARJOVSKY, Martin; CHINTALA, Soumith; BOTTOU, Léon. **Wasserstein GAN**. arXiv preprint, 2017. DOI: 10.48550/arXiv.1701.07875.

ARROYO-NEGRETE, Elkin; DEVEGOWDA, Deepak; DATTA-GUPTA, Akhil; CHOE, J. Streamline-Assisted Ensemble Kalman Filter for Rapid and Continuous Reservoir Model Updating. **SPE Reservoir Evaluation & Engineering**, v. 11, n. 06, p. 1046–1060, 2008. DOI: 10.2118/104255-PA.

BAHTIYAR, Hüseyin; SOYDANER, Derya; YÜKSEL, Esra. Application of multilayer perceptron with data augmentation in nuclear physics. **Applied Soft Computing**, v. 128, p. 109470, 2022. DOI: 10.1016/j.asoc.2022.109470.

BAO, Jichao; LI, Liangping; DAVIS, Arden. Variational Autoencoder or Generative Adversarial Networks? A Comparison of Two Deep Learning Methods for Flow and

Transport Data Assimilation. **Mathematical Geosciences**, v. 54, n. 6, p. 1017–1042, 2022. DOI: 10.1007/s11004-022-10003-3.

BERTHELOT, David; SCHUMM, Thomas; METZ, Luke. **BEGAN: Boundary Equilibrium Generative Adversarial Networks**. arXiv preprint, 2017. DOI: 10.48550/arXiv.1703.10717.

BIŃKOWSKI, Mikołaj; SUTHERLAND, Danica J.; ARBEL, Michael; GRETTON, Arthur. **Demystifying MMD GANs**. arXiv preprint, 2021. DOI: 10.48550/arXiv.1801.01401.

BISHOP, Craig H.; HODYSS, Daniel. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. **Quarterly Journal of the Royal Meteorological Society**, v. 133, n. 629, p. 2029–2044, 2007. DOI: 10.1002/qj.169.

BOND-TAYLOR, Sam; LEACH, Adam; LONG, Yang; WILLCOCKS, Chris G. Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 44, n. 11, p. 7327–7347, 2022. DOI: 10.1109/TPAMI.2021.3116668.

BORJI, Ali. **Pros and Cons of GAN Evaluation Measures**. arXiv preprint, 2018. DOI: 10.48550/arXiv.1802.03446.

BRESLAVICH, Ilya D.; SARKISOV, Gennadiy G.; MAKAROVA, Ekaterina S. Experience of MDA Ensemble Smoother Practice for Volga-Ural Oilfield. *Em*: DAY 2 TUE, OCTOBER 17, 2017, Moscow, Russia. **Anais**. *In*: SPE RUSSIAN PETROLEUM TECHNOLOGY CONFERENCE. Moscow, Russia: SPE, 2017. p. D023S007R004. DOI: 10.2118/187800-MS.

BRUHN, Carlos H. L.; PINTO, Antonio C. C.; JOHANN, Paulo R. S.; BRANCO, Celso C. M.; SALOMÃO, Marcelo C.; FREIRE, Ednilson B. Campos and Santos Basins: 40 Years of Reservoir Characterization and Management of Shallow- to Ultra-Deep Water, Post- and Pre-Salt Reservoirs - Historical Overview and Future Challenges. *Em*: DAY 1 TUE, OCTOBER 24, 2017, Rio de Janeiro, Brazil. **Anais**. *In*: OTC BRASIL. Rio de Janeiro, Brazil: OTC, 2017. p. D011S006R001. DOI: 10.4043/28159-MS.

CAERS, Jef; ZHANG, Tuanfeng. Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models. **AAPG memoir**, v. 80, p. 383–394, 2004.

CANCHUMUNI, Smith Arauco; EMERICK, Alexandre A.; PACHECO, Marco Aurelio. Integration of Ensemble Data Assimilation and Deep Learning for History Matching Facies Models. *Em*: DAY 1 TUE, OCTOBER 24, 2017 2017, Rio de Janeiro, Brazil. **Anais**. *In*: OTC BRASIL. Rio de Janeiro, Brazil: OTC, 2017. p. D011S006R005. DOI: 10.4043/28015-MS.

CANCHUMUNI, Smith W. A.; CASTRO, Jose D. B.; POTRATZ, Júlia; EMERICK, Alexandre A.; PACHECO, Marco Aurélio C. Recent developments combining ensemble smoother and deep generative networks for facies history matching.

Computational Geosciences, v. 25, n. 1, p. 433–466, 2021. DOI: 10.1007/s10596-020-10015-0.

CANCHUMUNI, Smith W. A.; EMERICK, Alexandre A.; PACHECO, Marco Aurélio C. History matching geological facies models based on ensemble smoother and deep generative models. **Journal of Petroleum Science and Engineering**, v. 177, p. 941–958, 2019. a. DOI: 10.1016/j.petrol.2019.02.037.

CANCHUMUNI, Smith W. A.; EMERICK, Alexandre A.; PACHECO, Marco Aurélio C. Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother. **Computers & Geosciences**, v. 128, p. 87–102, 2019. b. DOI: 10.1016/j.cageo.2019.04.006.

CHAN, Shing; ELSHEIKH, Ahmed H. **Parametrization and generation of geological models with generative adversarial networks**. arXiv preprint, 2017. DOI: 10.48550/arXiv.1708.01810.

CHAN, Shing; ELSHEIKH, Ahmed H. **Parametrization of stochastic inputs using generative adversarial networks with application in geology**. arXiv preprint, 2019. a. DOI: 10.48550/arXiv.1904.03677.

CHAN, Shing; ELSHEIKH, Ahmed H. Parametric generation of conditional geological realizations using generative neural networks. **Computational Geosciences**, v. 23, n. 5, p. 925–952, 2019. b. DOI: 10.1007/s10596-019-09850-7.

CHEN, Yan; OLIVER, Dean S. Cross-covariances and localization for EnKF in multiphase flow data assimilation. **Computational Geosciences**, v. 14, n. 4, p. 579–601, 2010. DOI: 10.1007/s10596-009-9174-6.

CHEN, Yan; OLIVER, Dean S. Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother. **Mathematical Geosciences**, v. 44, n. 1, p. 1–26, 2012. DOI: 10.1007/s11004-011-9376-z.

CHEN, Yan; OLIVER, Dean S. History Matching Of The Norne Full Field Model Using An Iterative Ensemble Smoother. *Em*: ALL DAYS 2013a, London, UK. **Anais***. In*: EAGE ANNUAL CONFERENCE & EXHIBITION INCORPORATING SPE EUROPEC. London, UK: SPE, 2013. p. SPE-164902-MS. DOI: 10.2118/164902-MS. Disponível em: https://onepetro.org/SPEEURO/proceedings/13EURO/All-13EURO/London,%20UK/177344. Acesso em: 16 mar. 2023.

CHEN, Yan; OLIVER, Dean S. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. **Computational Geosciences**, v. 17, n. 4, p. 689–703, 2013. b. DOI: 10.1007/s10596-013-9351-5.

CHEN, Yan; OLIVER, Dean 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, 2014. DOI: 10.2118/164902-PA.

CHEN, Yan; OLIVER, Dean S. Localization and regularization for iterative ensemble smoothers. **Computational Geosciences**, v. 21, n. 1, p. 13–30, 2017. DOI: 10.1007/s10596-016-9599-7.

COATS, K. H. Use and Misuse of Reservoir Simulation Models. **Journal of Petroleum Technology**, v. 21, n. 11, p. 1391–1398, 1969. DOI: 10.2118/2367-PA.

CORREIA, M.; HOHENDORFF, J.; GASPAR, A. T.; SCHIOZER, D. UNISIM-II-D: Benchmark Case Proposal Based on a Carbonate Reservoir. 2015. **Anais**. *In*: SPE LATIN AMERICAN AND CARIBBEAN PETROLEUM ENGINEERING CONFERENCE. Quito, Ecuador: SPE, 2015. p. D031S020R004. DOI: 10.2118/177140-MS.

CORREIA, M.; SCHIOZER, Denis José. Scaling Up Highly Permeable Thin Layers Into Flow Simulation. **SPE Reservoir Evaluation & Engineering**, v. 21, n. 02, p. 503–520, 2018. DOI: 10.2118/185875-PA.

DA ROCHA, Herson Oliveira; DA COSTA, Jéssica Lia Santos; CARRASQUILLA, Antonio Abel Gonzaléz; CARRASCO, Alfredo Moisés Vallejos. Permeability estimation and analysis of fracture networks using resistivity logs in an offshore Aptian carbonate reservoir pre-salt, in the Southeastern Santos Basin. **Journal of Applied Geophysics**, v. 184, p. 104241, 2021. DOI: 10.1016/j.jappgeo.2020.104241.

DAMIANI, Marcos Cabral. **Determinação de Padrões de Fluxo em Simulações de Reservatório de Petróleo Utilizando Traçadores**. 2007. Master of Science Thesis - Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 2007.

DENG, Jia; DONG, Wei; SOCHER, Richard; LI, Li-Jia; KAI LI; LI FEI-FEI. ImageNet: A large-scale hierarchical image database. 2009. **Anais**. *In*: 2009 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (CVPR WORKSHOPS). Miami, FL: IEEE, 2009. p. 248–255. DOI: 10.1109/CVPR.2009.5206848.

DEVEGOWDA, Deepak; ARROYO-NEGRETE, Elkin; DATTA-GUPTA, Akhil. Flow relevant covariance localization during dynamic data assimilation using EnKF. **Advances in Water Resources**, v. 33, n. 2, p. 129–145, 2010. DOI: 10.1016/j.advwatres.2009.10.001.

DONOHO, David L.; JOHNSTONE, Iain M. Ideal spatial adaptation by wavelet shrinkage. **Biometrika**, v. 81, n. 3, p. 425–455, 1994. DOI: 10.1093/biomet/81.3.425.

DONOHO, David L.; JOHNSTONE, Iain M. Adapting to Unknown Smoothness via Wavelet Shrinkage. **Journal of the American Statistical Association**, v. 90, n. 432, p. 1200–1224, 1995. DOI: 10.1080/01621459.1995.10476626.

EMERICK, Alexandre 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, 2016. DOI: 10.1016/j.petrol.2016.01.029.

EMERICK, Alexandre A. Investigation on Principal Component Analysis Parameterizations for History Matching Channelized Facies Models with Ensemble-Based Data Assimilation. **Mathematical Geosciences**, v. 49, n. 1, p. 85–120, 2017. DOI: 10.1007/s11004-016-9659-5.

EMERICK, Alexandre A. Deterministic ensemble smoother with multiple data assimilation as an alternative for history-matching seismic data. **Computational Geosciences**, v. 22, n. 5, p. 1175–1186, 2018. DOI: 10.1007/s10596-018-9745-5.

EMERICK, Alexandre A. Analysis of geometric selection of the data-error covariance inflation for ES-MDA. **Journal of Petroleum Science and Engineering**, v. 182, p. 106168, 2019. DOI: 10.1016/j.petrol.2019.06.032.

EMERICK, Alexandre A.; REYNOLDS, Albert C. History Matching a Field Case Using the Ensemble Kalman Filter With Covariance Localization. **SPE Reservoir Evaluation & Engineering**, v. 14, n. 04, p. 443–452, 2011. a. DOI: 10.2118/141216-PA.

EMERICK, Alexandre A.; REYNOLDS, Albert C. Ensemble smoother with multiple data assimilation. **Computers & Geosciences**, v. 55, p. 3–15, 2013. DOI: 10.1016/j.cageo.2012.03.011.

EMERICK, Alexandre; REYNOLDS, Albert. Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. **Computational Geosciences**, v. 15, n. 2, p. 251–269, 2011. b. DOI: 10.1007/s10596-010-9198-y.

EVENSEN, Geir. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. **Journal of Geophysical Research**, v. 99, n. C5, p. 10143, 1994. DOI: 10.1029/94JC00572.

EVENSEN, Geir. The Ensemble Kalman Filter: theoretical formulation and practical implementation. **Ocean Dynamics**, v. 53, n. 4, p. 343–367, 2003. DOI: 10.1007/s10236-003-0036-9.

EVENSEN, Geir. Sampling strategies and square root analysis schemes for the EnKF. **Ocean Dynamics**, v. 54, n. 6, p. 539–560, 2004. DOI: 10.1007/s10236-004-0099-2.

EVENSEN, Geir. **Data Assimilation: The Ensemble Kalman Filter**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. DOI: 10.1007/978-3-642-03711-5.

EVENSEN, Geir. Analysis of iterative ensemble smoothers for solving inverse problems. **Computational Geosciences**, v. 22, n. 3, p. 885–908, 2018. DOI: 10.1007/s10596-018-9731-y.

EVENSEN, Geir. Accounting for model errors in iterative ensemble smoothers. **Computational Geosciences**, v. 23, n. 4, p. 761–775, 2019. DOI: 10.1007/s10596-019-9819-z.

EVENSEN, Geir; EIKREM, Kjersti Solberg. Conditioning reservoir models on rate data using ensemble smoothers. **Computational Geosciences**, v. 22, n. 5, p. 1251–1270, 2018. DOI: 10.1007/s10596-018-9750-8.

FENG, Runhai; GRANA, Dario; MUKERJI, Tapan; MOSEGAARD, Klaus. Application of Bayesian Generative Adversarial Networks to Geological Facies Modeling.

Mathematical Geosciences, v. 54, n. 5, p. 831–855, 2022. DOI: 10.1007/s11004-022-09994-w.

FURRER, Reinhard; BENGTSSON, Thomas. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. **Journal of Multivariate Analysis**, v. 98, n. 2, p. 227–255, 2007. DOI: 10.1016/j.jmva.2006.08.003.

GASPARI, Gregory; COHN, Stephen E. Construction of correlation functions in two and three dimensions. **Quarterly Journal of the Royal Meteorological Society**, v. 125, n. 554, p. 723–757, 1999. DOI: 10.1002/qj.49712555417.

GONOG, Liang; ZHOU, Yimin. A Review: Generative Adversarial Networks. 2019. **Anais**. *In*: 2019 14TH IEEE CONFERENCE ON INDUSTRIAL ELECTRONICS AND APPLICATIONS (ICIEA). Xi'an, China: IEEE, 2019. p. 505–510. DOI: 10.1109/ICIEA.2019.8833686.

GOODFELLOW, Ian J.; POUGET-ABADIE, Jean; MIRZA, Mehdi; XU, Bing; WARDE-FARLEY, David; OZAIR, Sherjil; COURVILLE, Aaron; BENGIO, Yoshua. **Generative Adversarial Networks**. arXiv preprint, 2014. DOI: 10.48550/arXiv.1406.2661.

GULRAJANI, Ishaan; AHMED, Faruk; ARJOVSKY, Martin; DUMOULIN, Vincent; COURVILLE, Aaron. **Improved Training of Wasserstein GANs**. arXiv preprint, 2017.

HAMILL, Thomas M.; WHITAKER, Jeffrey S.; SNYDER, Chris. Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. **Monthly Weather Review**, v. 129, n. 11, p. 2776–2790, 2001. DOI: 10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

HAUGEN, Vibeke; NÆVDAL, Geir; NATVIK, Lars-Jøergen; EVENSEN, Geir; BERG, Aina M.; FLORNES, Kristin M. History Matching Using the Ensemble Kalman Filter on a North Sea Field Case. **SPE Journal**, v. 13, n. 04, p. 382–391, 2008. DOI: 10.2118/102430-PA.

HEUSEL, Martin; RAMSAUER, Hubert; UNTERTHINER, Thomas; NESSLER, Bernhard; HOCHREITER, Sepp. **GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium**. arXiv preprint, 2018. DOI: 10.48550/arXiv.1706.08500

HONG, Yongjun; HWANG, Uiwon; YOO, Jaeyoon; YOON, Sungroh. How Generative Adversarial Networks and Their Variants Work: An Overview. **ACM Computing Surveys**, v. 52, n. 1, p. 1–43, 2020. DOI: 10.1145/3301282.

HOUTEKAMER, P. L.; MITCHELL, Herschel L. Data Assimilation Using an Ensemble Kalman Filter Technique. **Monthly Weather Review**, v. 126, n. 3, p. 796–811, 1998. DOI: 10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

HOUTEKAMER, P. L.; MITCHELL, Herschel L. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. **Monthly Weather Review**, v. 129, n. 1, p. 123–137, 2001. DOI: 10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

IGLESIAS, Guillermo; TALAVERA, Edgar; DÍAZ-ÁLVAREZ, Alberto. **A survey on GANs for computer vision: Recent research, analysis and taxonomy**. arXiv preprint, 2022. DOI: 10.48550/arXiv.2203.11242

JAITLY, Navdeep; HINTON, Geoffrey E. Vocal Tract Length Perturbation (VTLP) improves speech recognition. **ICML Workshop on Deep Learning for Audio, Speech and Language**, v. 117, p. 21, 2013.

JOHNSON, Justin; ALAHI, Alexandre; FEI-FEI, Li. **Perceptual Losses for Real-Time Style Transfer and Super-Resolution**. arxiv preprint, 2016. DOI: 10.48550/ARXIV.1603.08155.

JUNG, Hyungsik; JO, Honggeun; KIM, Sungil; LEE, Kyungbook; CHOE, Jonggeun. Geological model sampling using PCA-assisted support vector machine for reliable channel reservoir characterization. **Journal of Petroleum Science and Engineering**, v. 167, p. 396–405, 2018. DOI: 10.1016/j.petrol.2018.04.017.

KALMAN, R. E. A New Approach to Linear Filtering and Prediction Problems. **Journal of Basic Engineering**, v. 82, n. 1, p. 35–45, 1960. DOI: 10.1115/1.3662552.

KARRAS, Tero; AITTALA, Miika; HELLSTEN, Janne; LAINE, Samuli; LEHTINEN, Jaakko; AILA, Timo. **Training Generative Adversarial Networks with Limited Data**. arXiv preprint, 2020. DOI: 10.48550/arXiv.2006.06676.

KARRAS, Tero; LAINE, Samuli; AILA, Timo. **A Style-Based Generator Architecture for Generative Adversarial Networks**. arXiv preprint, 2018. DOI: 10.48550/ARXIV.1812.04948.

KARRAS, Tero; LAINE, Samuli; AITTALA, Miika; HELLSTEN, Janne; LEHTINEN, Jaakko; AILA, Timo. **Analyzing and Improving the Image Quality of StyleGAN**. arXiv preprint, 2019. DOI: 10.48550/ARXIV.1912.04958.

KIM, Chanwoo; SHIN, Minkyu; GARG, Abhinav; GOWDA, Dhananjaya. Improved Vocal Tract Length Perturbation for a State-of-the-Art End-to-End Speech Recognition System. 2019. **Anais**. *In*: INTERSPEECH 2019. ISCA, 2019. p. 739–743. DOI: 10.21437/Interspeech.2019-3227.

KIM, Jaejun; PARK, Changhyup; LEE, Kyungbook; AHN, Seongin; JANG, Ilsik. Deep neural network coupled with distance-based model selection for efficient history matching. **Journal of Petroleum Science and Engineering**, v. 185, p. 106658, 2020. a. DOI: 10.1016/j.petrol.2019.106658.

KIM, Sungil; LEE, Kyungbook; LIM, Jungtek; JEONG, Hoonyoung; MIN, Baehyun. Development of ensemble smoother–neural network and its application to history matching of channelized reservoirs. **Journal of Petroleum Science and Engineering**, v. 191, p. 107159, 2020. b. DOI: 10.1016/j.petrol.2020.107159.

KINGMA, Diederik P.; BA, Jimmy. Adam: **A Method for Stochastic Optimization**. arxiv preprint, 2014. DOI: 10.48550/ARXIV.1412.6980.

LACERDA, Johann M.; EMERICK, Alexandre A.; PIRES, Adolfo P. Methods to mitigate loss of variance due to sampling errors in ensemble data assimilation with non-local model parameters. **Journal of Petroleum Science and Engineering**, v. 172, p. 690–706, 2019. DOI: 10.1016/j.petrol.2018.08.056.

LACERDA, Johann M.; EMERICK, Alexandre A.; PIRES, Adolfo P. Using a machine learning proxy for localization in ensemble data assimilation. **Computational Geosciences**, v. 25, n. 3, p. 931–944, 2021. DOI: 10.1007/s10596-020-10031-0.

LALOY, Eric; HÉRAULT, Romain; LEE, John; JACQUES, Diederik; LINDE, Niklas. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. **Advances in Water Resources**, v. 110, p. 387–405, 2017. DOI: 10.1016/j.advwatres.2017.09.029.

LE, Duc H.; EMERICK, Alexandre A.; REYNOLDS, Albert C. An Adaptive Ensemble Smoother With Multiple Data Assimilation for Assisted History Matching. **SPE Journal**, v. 21, n. 06, p. 2195–2207, 2016. DOI: 10.2118/173214-PA.

LI, Liangping; STETLER, Larry; CAO, Zhendan; DAVIS, Arden. An iterative normal-score ensemble smoother for dealing with non-Gaussianity in data assimilation. **Journal of Hydrology**, v. 567, p. 759–766, 2018. DOI: 10.1016/j.jhydrol.2018.01.038.

LIPTÁK, Béla G. (ORG.). **Instrument engineers’ handbook**. 4th ed ed. Boca Raton, FL: CRC Press, 2003.

LIPTON, Zachary C.; TRIPATHI, Subarna. **Precise Recovery of Latent Vectors from Generative Adversarial Networks**. arxiv preprint, 2017. DOI: 10.48550/ARXIV.1702.04782.

LIU, Ning; OLIVER, Dean S. Ensemble Kalman filter for automatic history matching of geologic facies. **Journal of Petroleum Science and Engineering**, v. 47, n. 3–4, p. 147–161, 2005. DOI: 10.1016/j.petrol.2005.03.006.

LORENTZEN, Rolf J.; BHAKTA, Tuhin; GRANA, Dario; LUO, Xiaodong; VALESTRAND, Randi; NÆVDAL, Geir. Simultaneous assimilation of production and seismic data: application to the Norne field. **Computational Geosciences**, v. 24, n. 2, p. 907–920, 2020. DOI: 10.1007/s10596-019-09900-0.

LORENTZEN, Rolf J.; FLORNES, Kristin M.; NÆVDAL, Geir. History Matching Channelized Reservoirs Using the Ensemble Kalman Filter. **SPE Journal**, v. 17, n. 01, p. 137–151, 2012. DOI: 10.2118/143188-PA.

LUCIC, Mario; KURACH, Karol; MICHALSKI, Marcin; GELLY, Sylvain; BOUSQUET, Olivier. **Are GANs Created Equal? A Large-Scale Study**. arXiv preprint, 2018. DOI: 10.48550/arXiv.1711.10337

LUO, Xiaodong. Novel iterative ensemble smoothers derived from a class of generalized cost functions. **Computational Geosciences**, v. 25, n. 3, p. 1159–1189, 2021. DOI: 10.1007/s10596-021-10046-1.

LUO, Xiaodong; BHAKTA, Tuhin. Automatic and adaptive localization for ensemble-based history matching. **Journal of Petroleum Science and Engineering**, v. 184, p. 106559, 2020. DOI: 10.1016/j.petrol.2019.106559.

LUO, Xiaodong; BHAKTA, Tuhin; NÆVDAL, Geir. Correlation-Based Adaptive Localization With Applications to Ensemble-Based 4D-Seismic History Matching. **SPE Journal**, v. 23, n. 02, p. 396–427, 2018. DOI: 10.2118/185936-PA.

LUO, Xiaodong; LORENTZEN, Rolf J.; VALESTRAND, Randi; EVENSEN, Geir. Correlation-Based Adaptive Localization for Ensemble-Based History Matching: Applied To the Norne Field Case Study. **SPE Reservoir Evaluation & Engineering**, v. 22, n. 03, p. 1084–1109, 2019. DOI: 10.2118/191305-PA.

LUO, Xiaodong; STORDAL, Andreas S.; LORENTZEN, Rolf J.; NÆVDAL, Geir. Iterative Ensemble Smoother as an Approximate Solution to a Regularized Minimum-Average-Cost Problem: Theory and Applications. **SPE Journal**, v. 20, n. 05, p. 962–982, 2015. DOI: 10.2118/176023-PA.

MA, Xiang; BI, Linfeng. A robust adaptive iterative ensemble smoother scheme for practical history matching applications. **Computational Geosciences**, v. 23, n. 3, p. 415–442, 2019. DOI: 10.1007/s10596-018-9786-9.

MASCHIO, Célio; AVANSI, Guilherme; SCHIOZER, Denis; SANTOS, Alberto. Study Case for History Matching and Uncertainties Reduction based on UNISIM-I Field. 2015. Available in: https://www.unisim.cepetro.unicamp.br/benchmarks/br/unisim-i/unisim-i-h.

MASCHIO, Célio; SANTOS, Alberto; CORREIA, Manuel. Case Study for Field History Matching and Uncertainties Reduction based on UNISIM-II. 2018. Available in: https://www.unisim.cepetro.unicamp.br/benchmarks/br/unisim-ii/unisim-ii-h.

MASCHIO, Célio; SCHIOZER, Denis José. A new parameterization method for data assimilation and uncertainty assessment for complex carbonate reservoir models based on cumulative distribution function. **Journal of Petroleum Science and Engineering**, v. 183, p. 106400, 2019. DOI: 10.1016/j.petrol.2019.106400.

MAUCEC, Marko; DE MATOS RAVANELLI, Fabio M.; LYNGRA, Stig; ZHANG, Shu J.; ALRAMADHAN, Aymen A.; ABDELHAMID, Osama A.; AL-GARNI, Saad A. Ensemble-Based Assisted History Matching With Rigorous Uncertainty Quantification Applied to a Naturally Fractured Carbonate Reservoir. 2016. **Anais**. *In*: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION. Dubai, UAE: SPE, 2016. p. D011S017R007. DOI: 10.2118/181325-MS.

MCELHOE, Bruce A. An Assessment of the Navigation and Course Corrections for a Manned Flyby of Mars or Venus. **IEEE Transactions on Aerospace and Electronic Systems**, v. AES-2, n. 4, p. 613–623, 1966. DOI: 10.1109/TAES.1966.4501892.

MESCHEDER, Lars; GEIGER, Andreas; NOWOZIN, Sebastian. **Which Training Methods for GANs do actually Converge?** arXiv preprint, 2018. DOI: 10.48550/arXiv.1801.04406.

MEYER, Franz O.; PRICE, Rex C.; AL-RAIMI, Saleh M. Stratigraphic and Petrophysical Characteristics of Cored Arab-D Super-k Intervals, Hawiyah Area, Ghawar Field, Saudi Arabia. **GeoArabia**, v. 5, n. 3, p. 355–384, 2000. DOI: 10.2113/geoarabia0503355.

MORENO, David L.; AANONSEN, Sigurd I. Continuous Facies Updating Using the Ensemble Kalman Filter and the Level Set Method. **Mathematical Geosciences**, v. 43, n. 8, p. 951–970, 2011. DOI: 10.1007/s11004-011-9347-4.

MOSSER, Lukas; DUBRULE, Olivier; BLUNT, Martin J. **Stochastic seismic waveform inversion using generative adversarial networks as a geological prior**. arXiv preprint, 2018. DOI: 10.48550/arXiv.1806.03720.

MOSSER, Lukas; DUBRULE, Olivier; BLUNT, Martin J. **DeepFlow: History Matching in the Space of Deep Generative Models**. arXiv preprint, 2019. DOI: 10.48550/arXiv.1905.05749.

MOTAMED, Saman; ROGALLA, Patrik; KHALVATI, Farzad. Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. **Informatics in Medicine Unlocked**, v. 27, p. 100779, 2021. DOI: 10.1016/j.imu.2021.100779.

NÆVDAL, Geir; MANNSETH, Trond; VEFRING, Erlend H. Near-Well Reservoir Monitoring Through Ensemble Kalman Filter. 2002. **Anais**. *In*: SPE/DOE IMPROVED OIL RECOVERY SYMPOSIUM. Tulsa, Oklahoma: SPE, 2002. p. SPE-75235-MS. DOI: 10.2118/75235-MS.

ODENA, Augustus; DUMOULIN, Vincent; OLAH, Chris. Deconvolution and Checkerboard Artifacts. **Distill**, v. 1, n. 10, p. e3, 2016. DOI: 10.23915/distill.00003.

OLIVER, Dean S.; CHEN, Yan. Recent progress on reservoir history matching: a review. **Computational Geosciences**, v. 15, n. 1, p. 185–221, 2011. DOI: 10.1007/s10596-010-9194-2.

OLIVER, Dean S.; FOSSUM, Kristian; BHAKTA, Tuhin; SANDØ, Ivar; NÆVDAL, Geir; LORENTZEN, Rolf Johan. 4D seismic history matching. **Journal of Petroleum Science and Engineering**, v. 207, p. 109119, 2021. DOI: 10.1016/j.petrol.2021.109119.

OLIVER, Dean S.; REYNOLDS, Albert C.; LIU, Ning. **Inverse Theory for Petroleum Reservoir Characterization and History Matching**. Cambridge University Press, 2008. DOI: 10.1017/CBO9780511535642.

OTT, Edward; HUNT, Brian R.; SZUNYOGH, Istvan; ZIMIN, Aleksey V.; KOSTELICH, Eric J.; CORAZZA, Matteo; KALNAY, Eugenia; PATIL, D. J.; YORKE, James A. A local ensemble Kalman filter for atmospheric data assimilation. **Tellus A: Dynamic Meteorology and Oceanography**, v. 56, n. 5, p. 415, 2004. DOI: 10.3402/tellusa.v56i5.14462.

Petrobras. 2022. Available em: https://petrobras.com.br/en/our-activities/performance-areas/oil-and-gas-exploration-and-production/pre-salt/.
Acessed on 26 Jan. 2023.

RAFIEE, Javad; REYNOLDS, Albert C. Theoretical and efficient practical procedures for the generation of inflation factors for ES-MDA. **Inverse Problems**, v. 33, n. 11, p. 115003, 2017. DOI: 10.1088/1361-6420/aa8cb2.

RANAZZI, Paulo Henrique; SAMPAIO, Marcio Augusto. Influence of the Kalman gain localization in adaptive ensemble smoother history matching. **Journal of Petroleum Science and Engineering**, v. 179, p. 244–256, 2019. a. DOI: 10.1016/j.petrol.2019.04.079.

RANAZZI, Paulo Henrique; SAMPAIO, Marcio Augusto. 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. 413, 2019. b. DOI: 10.1007/s40430-019-1935-0.

RAZAK, Syamil Mohd; JAFARPOUR, Behnam. Conditioning generative adversarial networks on nonlinear data for subsurface flow model calibration and uncertainty quantification. **Computational Geosciences**, v. 26, n. 1, p. 29–52, 2022. DOI: 10.1007/s10596-021-10112-8.

REMY, Nicolas; BOUCHER, Alexandre; WU, Jianbing. **Applied Geostatistics with SGeMS: A User’s Guide**. Cambridge University Press, 2009. DOI: 10.1017/CBO9781139150019.

ROSS, Andrew Slavin; DOSHI-VELEZ, Finale. **Improving the Adversarial Robustness and Interpretability of Deep Neural Networks by Regularizing their Input Gradients**. arxiv preprint, 2017. DOI: 10.48550/ARXIV.1711.09404.

RWECHUNGURA, R.; DADASHPOUR, M.; KLEPPE, J. Advanced History Matching Techniques Reviewed. 2011. **Anais**. *In*: SPE MIDDLE EAST OIL AND GAS SHOW AND CONFERENCE. Manama, Bahrain: SPE, 2011. p. SPE-142497-MS. DOI: 10.2118/142497-MS.

SACHER, William; BARTELLO, Peter. Sampling Errors in Ensemble Kalman Filtering. Part I: Theory. **Monthly Weather Review**, v. 136, n. 8, p. 3035–3049, 2008. DOI: 10.1175/2007MWR2323.1.

SALIMANS, Tim; GOODFELLOW, Ian; ZAREMBA, Wojciech; CHEUNG, Vicki; RADFORD, Alec; CHEN, Xi. **Improved Techniques for Training GANs**. arXiv preprint, 2016. DOI: 10.48550/arXiv.1606.03498.

SANDFORT, Veit; YAN, Ke; PICKHARDT, Perry J.; SUMMERS, Ronald M. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. **Scientific Reports**, v. 9, n. 1, p. 16884, 2019. DOI: 10.1038/s41598-019-52737-x.

SARMA, P.; CHEN, W. H. Generalization of the Ensemble Kalman Filter Using Kernels for Non-gaussian Random Fields. 2009. **Anais**. *In*: SPE RESERVOIR SIMULATION SYMPOSIUM. The Woodlands, Texas: SPE, 2009. p. SPE-119177-MS. DOI: 10.2118/119177-MS.

SARMA, Pallav; DURLOFSKY, Louis J.; AZIZ, Khalid. Kernel Principal Component Analysis for Efficient, Differentiable Parameterization of Multipoint Geostatistics.

Mathematical Geosciences, v. 40, n. 1, p. 3–32, 2008. DOI: 10.1007/s11004-007-9131-7.

SEBACHER, Bogdan; HANEA, Remus. Channelized reservoir estimation using a low-dimensional parameterization based on high-order singular value decomposition. **Computational Geosciences**, v. 24, n. 2, p. 509–531, 2020. DOI: 10.1007/s10596-019-09856-1.

SEBACHER, Bogdan; HANEA, Remus; HEEMINK, Arnold. A probabilistic parametrization for geological uncertainty estimation using the ensemble Kalman filter (EnKF). **Computational Geosciences**, v. 17, n. 5, p. 813–832, 2013. DOI: 10.1007/s10596-013-9357-z.

SEBACHER, Bogdan; HANEA, Remus; STORDAL, Andreas S. An adaptive pluri-Gaussian simulation model for geological uncertainty quantification. **Journal of Petroleum Science and Engineering**, v. 158, p. 494–508, 2017. DOI: 10.1016/j.petrol.2017.08.038.

SEBACHER, Bogdan; STORDAL, Andreas S.; HANEA, Remus. Bridging multipoint statistics and truncated Gaussian fields for improved estimation of channelized reservoirs with ensemble methods. **Computational Geosciences**, v. 19, n. 2, p. 341–369, 2015. DOI: 10.1007/s10596-014-9466-3.

SEILER, A.; EVENSEN, G.; SKJERVHEIM, J. A.; HOVE, J.; VABØ, J. G. Advanced Reservoir Management Workflow Using an EnKF Based Assisted History Matching Method. 2009. **Anais***. In*: SPE RESERVOIR SIMULATION SYMPOSIUM. The Woodlands, Texas: SPE, 2009. p. SPE-118906-MS. DOI: 10.2118/118906-MS.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on Image Data Augmentation for Deep Learning. **Journal of Big Data**, v. 6, n. 1, p. 60, 2019. DOI: 10.1186/s40537-019-0197-0.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M.; FURHT, Borko. Text Data Augmentation for Deep Learning. **Journal of Big Data**, v. 8, n. 1, p. 101, 2021. DOI: 10.1186/s40537-021-00492-0.

SILVA, Vinícius Luiz Santos; EMERICK, Alexandre Anozé; COUTO, Paulo; ALVES, José Luis Drummond. History matching and production optimization under uncertainties – Application of closed-loop reservoir management. **Journal of Petroleum Science and Engineering**, v. 157, p. 860–874, 2017. DOI: 10.1016/j.petrol.2017.07.037.

SIMONYAN, Karen; ZISSERMAN, Andrew. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. arxiv preprint, 2014. DOI: 10.48550/ARXIV.1409.1556.

SKJERVHEIM, J. A. A.; EVENSEN, G.; HOVE, J.; VABØ, J. G. An Ensemble Smoother for assisted History Matching. 2011. **Anais***. In*: SPE RESERVOIR SIMULATION SYMPOSIUM. The Woodlands, Texas, USA: SPE, 2011. p. SPE-141929-MS. DOI: 10.2118/141929-MS.

SKJERVHEIM, J. A.; EVENSEN, G.; AANONSEN, S. I.; RUUD, B. O.; JOHANSEN, T. A. Incorporating 4D Seismic Data in Reservoir Simulation Models Using Ensemble Kalman Filter. **SPE Journal**, v. 12, n. 03, p. 282–292, 2007. DOI: 10.2118/95789-PA.

SOARES, Ricardo Vasconcellos; LUO, Xiaodong; EVENSEN, Geir; BHAKTA, Tuhin. Handling Big Models and Big Data Sets in History-Matching Problems through an Adaptive Local Analysis Scheme. **SPE Journal**, v. 26, n. 02, p. 973–992, 2021. DOI: 10.2118/204221-PA.

SOARES, Ricardo Vasconcellos; MASCHIO, Célio; SCHIOZER, Denis José. 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, 2018. DOI: 10.1016/j.petrol.2018.05.059.

SØNDERBY, Casper Kaae; CABALLERO, Jose; THEIS, Lucas; SHI, Wenzhe; HUSZÁR, Ferenc. **Amortised MAP Inference for Image Super-resolution**. arXiv preprint, 2017. DOI: 10.48550/arXiv.1610.04490.

SRIVASTAVA, Nitish; HINTON, Geoffrey; KRIZHEVSKY, Alex; SUTSKEVER, Ilya; SALAKHUTDINOV, Ruslan. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **Journal of Machine Learning Research**, v. 15, n. 1, 2014.

STRÉBELLE, Sébastien. **Sequential simulation drawing structures from training images**. 2000. Stanford University, Stanford, CA, USA, 2000. Available in: https://searchworks.stanford.edu/view/4701249.

STRÉBELLE, Sébastien. Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics. **Mathematical Geology**, v. 34, n. 1, p. 1–21, 2002. DOI: 10.1023/A:1014009426274.

SZEGEDY, Christian; VANHOUCKE, Vincent; IOFFE, Sergey; SHLENS, Jonathon; WOJNA, Zbigniew. **Rethinking the Inception Architecture for Computer Vision**. arXiv preprint, 2015. DOI: 10.48550/ARXIV.1512.00567.

THANH-TUNG, Hoang; TRAN, Truyen. **On Catastrophic Forgetting and Mode Collapse in Generative Adversarial Networks**. arxiv preprint, 2018. DOI: 10.48550/ARXIV.1807.04015.

TRAN, Ngoc-Trung; TRAN, Viet-Hung; NGUYEN, Ngoc-Bao; NGUYEN, Trung-Kien; CHEUNG, Ngai-Man. On Data Augmentation for GAN Training. **IEEE Transactions on Image Processing**, v. 30, p. 1882–1897, 2021. DOI: 10.1109/TIP.2021.3049346.

VAN LEEUWEN, Peter. A consistent interpretation of the stochastic version of the Ensemble Kalman Filter. **Quarterly Journal of the Royal Meteorological Society**, v. 146, n. 731, p. 2815–2825, 2020. DOI: 10.1002/qj.3819.

VAN LEEUWEN, Peter Jan. Comment on “Data Assimilation Using an Ensemble Kalman Filter Technique”. **Monthly Weather Review**, v. 127, n. 6, p. 1374–1377, 1999. DOI: 10.1175/1520-0493(1999)127<1374:CODAUA>2.0.CO;2.

VAN LEEUWEN, Peter Jan; EVENSEN, Geir. Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. **Monthly Weather Review**, v. 124, n. 12, p. 2898–2913, 1996. DOI: 10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.

VO, Hai X.; DURLOFSKY, Louis J. A New Differentiable Parameterization Based on Principal Component Analysis for the Low-Dimensional Representation of Complex Geological Models. **Mathematical Geosciences**, v. 46, n. 7, p. 775–813, 2014. DOI: 10.1007/s11004-014-9541-2.

VO, Hai X.; DURLOFSKY, Louis J. Data assimilation and uncertainty assessment for complex geological models using a new PCA-based parameterization. **Computational Geosciences**, v. 19, n. 4, p. 747–767, 2015. DOI: 10.1007/s10596-015-9483-x.

WANG, Xuguang; BISHOP, Craig H. A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes. **Journal of the Atmospheric Sciences**, v. 60, n. 9, p. 1140–1158, 2003. DOI: 10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.

WATANABE, Shingo; DATTA-GUPTA, Akhil. Use of Phase Streamlines for Covariance Localization in Ensemble Kalman Filter for Three-Phase History Matching. **SPE Reservoir Evaluation & Engineering**, v. 15, n. 03, p. 273–289, 2012. DOI: 10.2118/144579-PA.

WHITE, Tom. **Sampling Generative Networks**. arXiv preprint, 2016. DOI: 10.48550/ARXIV.1609.04468.

WIATRAK, Maciej; ALBRECHT, Stefano V.; NYSTROM, Andrew. **Stabilizing Generative Adversarial Networks: A Survey**. arXiv preprint, 2019. DOI: 10.48550/ARXIV.1910.00927.

WONG, Sebastien C.; GATT, Adam; STAMATESCU, Victor; MCDONNELL, Mark D. Understanding Data Augmentation for Classification: When to Warp?, 2016. **Anais.** *In*: 2016 INTERNATIONAL CONFERENCE ON DIGITAL IMAGE COMPUTING: TECHNIQUES AND APPLICATIONS (DICTA). Gold Coast, Australia: IEEE, 2016. p. 1–6. DOI: 10.1109/DICTA.2016.7797091.

YANG, Yulin; JIN, Rize; XU, Caie. On the Effects of Skip Connections in Deep Generative Adversarial Models, 2020. **Anais.** *In*: CSAI 2020: 2020 4TH INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE. Zhuhai China: ACM, 2020. p. 57–61. DOI: 10.1145/3445815.3445825.

YEO, M. J.; JUNG, S. P.; CHOE, J. Covariance Matrix Localization Using Drainage Area in an Ensemble Kalman Filter. **Energy Sources, Part A: Recovery, Utilization, and Environmental Effects**, v. 36, n. 19, p. 2154–2165, 2014. DOI: 10.1080/15567036.2010.511437.

ZHANG, Chengkai; SONG, Xianzhi; AZEVEDO, Leonardo. U-net generative adversarial network for subsurface facies modeling. **Computational Geosciences**, v. 25, n. 1, p. 553–573, 2021. DOI: 10.1007/s10596-020-10027-w.

ZHANG, Kai et al. Multi-source information fused generative adversarial network model and data assimilation based history matching for reservoir with complex geologies. **Petroleum Science**, v. 19, n. 2, p. 707–719, 2022. DOI: 10.1016/j.petsci.2021.10.007.

ZHANG, Xinru et al. CarveMix: A simple data augmentation method for brain lesion segmentation. **NeuroImage**, v. 271, p. 120041, 2023. DOI: 10.1016/j.neuroimage.2023.120041.

ZHANG, Yanfen; OLIVER, Dean S. Improving the Ensemble Estimate of the Kalman Gain by Bootstrap Sampling. **Mathematical Geosciences**, v. 42, n. 3, p. 327–345, 2010. DOI: 10.1007/s11004-010-9267-8.

ZHANG, Yanfen; OLIVER, Dean S. History Matching Using the Ensemble Kalman Filter With Multiscale Parameterization: A Field Case Study. **SPE Journal**, v. 16, n. 02, p. 307–317, 2011. DOI: 10.2118/118879-PA.

ZHAO, Shengyu; LIU, Zhijian; LIN, Ji; ZHU, Jun-Yan; HAN, Song. **Differentiable Augmentation for Data-Efficient GAN Training**. arXiv preprint, 2020a. DOI: 10.48550/ARXIV.2006.10738.

ZHAO, Yu; FOROUZANFAR, Fahim; REYNOLDS, Albert C. History matching of multi-facies channelized reservoirs using ES-MDA with common basis DCT. **Computational Geosciences**, v. 21, n. 5–6, p. 1343–1364, 2017. DOI: 10.1007/s10596-016-9604-1.

ZHAO, Zhengli; ZHANG, Zizhao; CHEN, Ting; SINGH, Sameer; ZHANG, Han. **Image Augmentations for GAN Training**. arXiv preprint, 2020b. DOI: 10.48550/ARXIV.2006.02595.

ZHOU, Haiyan; GÓMEZ-HERNÁNDEZ, J. Jaime; HENDRICKS FRANSSEN, Harrie-Jan; LI, Liangping. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. **Advances in Water Resources**, v. 34, n. 7, p. 844–864, 2011. DOI: 10.1016/j.advwatres.2011.04.014.