DOI: https://doi.org/10.1093/jge/gxad099

Journal of Geophysics and Engineering

Journal of Geophysics and Engineering (2024) **21**, 68–97

https://doi.org/10.1093/jge/gxad099

Advance access publication 30 November 2023

Performance assessment of an iterative ensemble smoother with local analysis to assimilate big 4D seismic datasets applied to a complex pre-salt-like benchmark case

Célio Maschio[1,*], **Gilson Moura Silva Neto**[2], **Alessandra Davolio**[1], **Vinicius de Souza Rios**[1] **and Denis José Schiozer**[1]

[1] Center for Petroleum and Energy Studies (CEPETRO), Universidade Estadual de Campinas (UNICAMP), Campinas, SP, 13.083-896, Brazil

[2] CENPES, Petrobras - Petróleo Brasileiro S.A., Rio de Janeiro, RJ, 21941-915, Brazil

[*]Corresponding author. E-mail: celio@cepetro.unicamp.br

Received: June 6, 2023. **Revised:** October 20, 2023. **Accepted:** November 29, 2023

Abstract

The use of 4D seismic (4DS) (or time-lapse seismic, TLS) in data assimilation (DA) makes the process more complex due to the higher amount of data to be assimilated, requiring more robust methods and better computational resources (processing capacity and memory). The development and application of permanent seismic monitoring technologies have increased in recent years, improving the overall 4D seismic quality in terms of signal resolution and repeatability. However, a massive amount of data is generated from the multiple monitors, making the incorporation of 4DS data in the DA process more complex. Therefore, robust DA methods capable of dealing with huge amount of data effectively and efficiently are essential. This paper aims to assess the performance of an iterative ensemble smoother method, named subspace ensemble randomized maximum likelihood with a local analysis, to assimilate a big dataset. The method was applied in a challenging pre-salt-like benchmark case with eight seismic surveys, one base, and seven monitors. The 4DS data are the impedance ratios (between two consecutive monitors) in 15 seismic horizons, totaling 105 maps to be assimilated. To our best knowledge, this is state of the art in terms of practical applications in DA. It was possible to assimilate all the data simultaneously: the 105 horizons for the 4DS data and the wells' production and pressure data. The DA was successful in terms of results quality and method performance. We also ran a case assimilating only well data for comparison purposes.

Keywords: reservoir simulation; data assimilation; 4D seismic data; big dataset; iterative ensemble smoother

1. Introduction

Brazilian pre-salt is an important oil province, contributing to a significant portion of the country's reserves and daily production (Cruz *et al.* 2016, Abelha and Petersohn 2018, Vasquez *et al.* 2019). Located in Santos and Campos sedimentary basins, the pre-salt reservoirs consist of microbial and coquina rocks at depths that surpass 5000 m (Johann and Monteiro 2016). The task of building geological simulation models for these reservoirs is highly challenging, involving significant technical uncertainties. Some of the main uncertainties related to pre-salt simulation models are reservoir connectivity, facies, and petrophysical properties distributions, response to the enhanced oil recovery strategy, distribution and behavior of faults and fractures, and fluid properties (Moczydlower *et al.* 2012). Furthermore, the projects in this province involve huge investments (Pizarro and Branco 2012). All these characteristics corroborate the importance of mitigating the model uncertainties using all the information available. In this respect, well and time-lapse seismic (TLS) data are complementary sources of information to calibrate the reservoir simulation models. The former provides information abundant in time but scarce in space, especially in offshore projects, which is the case of the pre-salt province. The latter provides information distributed in space, helping to update parameters far from the wells.

Iterative ensemble smoothers are a popular choice for assimilating well and TLS data to improve reservoir models quality (Skjervheim *et al.* 2007, Fahimuddin *et al.* 2010, Emerick and Reynolds 2013a, Emerick 2016). Nevertheless, this type of application is associated with relevant challenges related to pre-salt reservoirs. The reservoirs are big and highly heterogeneous in terms of permo-porous and facies distribution, diagenesis, faults, and fractures (Johann and Monteiro 2016). These characteristics complicate the representation of the reservoir's main features using an ensemble with a limited number of models, each with a restricted number of active cells. Both the ensemble and the model sizes contribute to increase the computational cost of the process. The production involves complex physics, with volatile fluids, often associated with high CO_2 contents, miscible gas injection alternated with water (Moczydlower *et al.* 2012, Johann and Monteiro 2016). Therefore, compositional fluid models are necessary to represent this process, increasing the complexity of both reservoir flow and TLS forward models. The reservoir sizes and the application of seismic monitoring technologies (Johann and Monteiro 2016, Deplante *et al.* 2019) contribute to increase the number of data points, requiring efficient ensemble-based methods to handle big datasets.

Batzle and Wang (1992) proposed classical correlations, which have been the standard method for fluid characterization in quantitative TLS studies (Avseth *et al.* 2005). Nevertheless, the presence of volatile fluids containing significant amounts of CO_2 limits classical correlations' applicability to estimate the speed of sound in the oil phase (Tahani 2012, Altundas *et al.* 2017, Silva Neto *et al.* 2020). There are specific correlations for fluids with such characteristics (Han *et al.* 2012, 2013). However, they require significant extra computations to integrate with compositional reservoir simulators, and one still needs to validate them for conditions that are different from the ranges applied during the experiments.

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

The ensemble randomized maximum likelihood (EnRML) method was primarily proposed as an iterative scheme to improve the performance of the ensemble Kalman filter (Evensen 1994) for highly nonlinear applications (Gu and Oliver 2007). Later, Chen and Oliver (2012, 2013) adapted the method for batch DA (iterative ensemble smoother) instead of the previous sequential approach, thus avoiding the need for time-consuming simulation restarts in reservoir applications. Another difference between filter and smoother is that, in the filter, data are assimilated sequentially in time, being necessary to update both parameters (reservoir properties such as porosity and permeability) and states (dynamic variables such pressure and fluid saturations). In the smoother, all data are assimilated simultaneously, and there is no need to update state variables (Emerick 2016). Recently, Raanes *et al.* (2019) improved the EnRML method conceptually and computationally using the fact that the solution is in the ensemble subspace. Evensen *et al.* (2019) followed their work in the same year and proposed an efficient algorithm to calibrate reservoir models using big datasets, the Subspace EnRML (SEnRML). Silva Neto *et al.* (2021) applied the SEnRML method with local analysis to assimilate TLS data in synthetic applications. They concluded that this method could lead to similar results to the ensemble smoother with multiple data assimilations (Emerick and Reynolds 2013b) with Kalman gain localization. They also reported that the SEnRML with local analysis has the advantage of requiring lower computational costs when there is a big dataset.

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

The SEnRML is a well-established and consolidated method (Evensen *et al.* 2019, Raanes *et al.* 2019). Additionally, the method SEnRML-LA was exhaustively tested in other cases and validated against the method ES-MDA by Silva Neto *et al.* (2021). Therefore, the comparison of the SEnRML-LA with other methods is beyond the scope of this paper.

1.1. *Motivation and objectives*

The main motivation for this work is the lack of practical applications documented in the literature regarding the assimilation of huge 4D seismic dataset in conjunction with well data in complex fields with many seismic surveys. Since the development and application of permanent monitoring technologies have been growing in the last few years, it is essential to seek data assimilation methods capable of dealing with the huge amount of data generated by these technologies. Efficient and effective methods are important to assimilate the data in a suitable time frame to take advantage of the value of these technologies, optimizing the time required to make the decisions in the context of reservoir management.

The main objective of this work is to show a practical application of the method Subspace EnRML (Evensen *et al.* 2019, Raanes *et al.* 2019) with a local analysis scheme (SEnRML-LA), implemented by Silva Neto *et al.* (2021), in a complex pre-salt benchmark case with a huge amount of 4D seismic data. The specific objectives are:

(i) Prove the robustness of the method in assimilating eight seismic surveys simultaneously, together with the well data. The main contribution of the paper is the validation of the method SEnRML-LA in a challenging DA problem typical of Brazilian pre-salt fields, confirming it as an alternative to solve practical problems involving permanent seismic monitoring technologies.

(ii) Compare two DA processes: (a) assimilation of time-lapse seismic and wells data simultaneously (labeled throughout this text as 'TLS-Well') and (b) assimilation of wells data only (labeled as 'Well'). The assimilation of only TLS is not of practical interest and is not tested in this work.

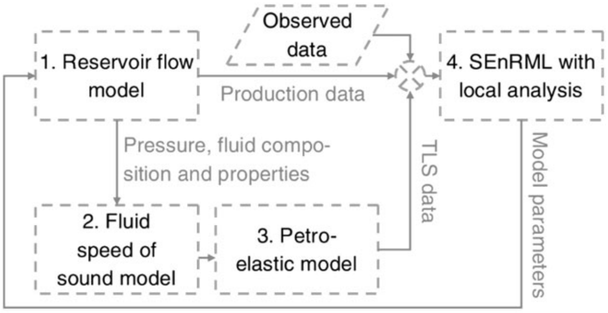

Figure 1. Well and TLS DA workflow.

(iii) Apply the compositional fluid model proposed by Silva Neto *et al.* (2020) in TLS DA in a realistic case.

(iv) Evaluate the TLS benefits in a complex pre-salt-like case.

2. Theoretical background

This work integrates two methods proposed previously to enable the well and seismic DA in a pre-salt-like synthetic application. We apply a compositional fluid model to perform the forward seismic modeling and use an iterative ensemble smoother implementation that is suitable for big datasets and big reservoirs, called SEnRML, with local analysis, to calibrate the parameters. Figure 1 illustrates our workflow, including the reservoir flow model (1), which provides the inputs for the compositional model of the speed of sound in the fluid (2), enabling the PEM computations (3). We ran the reservoir flow models with the compositional reservoir simulator GEM v.2020.1 (CMG 2020). The production and TLS data from the reservoir flow and the PEMs are compared to the observed data. The differences cause model parameter updates through the SEnRML with local analysis (4).

We describe the compositional fluid model and the DA method in the following subsections. Besides the DA workflow, we use a quadratic metric called normalized quadratic deviation with a sign (NQDS) to analyze the well data match. We define this metric in Subsection 2.3 and address the PEM, step (3), in Section 3.4.

2.1. *Compositional fluid model for seismic simulation*

The current work's application considers volatile oil with around 40% CO_2 content. The widely known Batzle and Wang (1992) correlation shows a relatively high deviation in representing the speed of sound in the oil phase with these characteristics (Tahani 2012, Altundas *et al.* 2017). Silva Neto *et al.* (2020) proposed the use of a calibrated cubic EOS to model the speed of sound in the hydrocarbon phases. We apply this model to compute the second step of the workflow depicted in Fig. 1.

Considering the Peng–Robinson EOS (Peng and Robinson 1976, Robinson and Peng 1978) with volume translation (Péneloux *et al.* 1982), the PVT relation is

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

where P is the pressure, T the temperature, $\bar{V}$ the molar volume, R the gas constant, and the parameters a_{PR}, b_{PR}, c_{PR}, and m_{PR} undertake different values for different components. One can calculate them as a function of the acentric factor, critical pressure, P_c, and critical temperature, T_c. Furthermore, it is necessary to apply mixing rules to represent the oil and gas phases as mixtures of components and pseudo-components (Pedersen *et al.* 2015). From the relation defined in Equation (1), one can compute the heat capacity at a constant volume

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

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

$$\frac{\boldsymbol{C_P}}{\boldsymbol{C_V}} = 1 - \frac{\boldsymbol{T}}{\boldsymbol{C_V}} \frac{\left(\left.\frac{\partial \boldsymbol{P}(\boldsymbol{T},\bar{\boldsymbol{V}})}{\partial \boldsymbol{T}}\right|_{\bar{\boldsymbol{V}}}\right)^2}{\left.\frac{\partial \boldsymbol{P}(\boldsymbol{T},\bar{\boldsymbol{V}})}{\partial \bar{\boldsymbol{V}}}\right|_{\boldsymbol{T}}}, \tag{3}$$

in which the pressure partial derivative at a constant molar volume (numerator) and a constant temperature (denominator) are calculated analytically from Equation (1). Finally, the speed of sound in the fluid is

$$\boldsymbol{v_P} = \sqrt{\frac{\boldsymbol{C_P}}{\boldsymbol{C_V}} \times \frac{1}{\boldsymbol{\rho c_T}}}, \tag{4}$$

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

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

2.2. *Subspace ensemble maximum likelihood (SEnRML) with local analysis*

Focusing on the fourth step of the workflow depicted in Fig. 1, the current SEnRML implementation with local analysis follows the revision presented by Raanes *et al.* (2019), the efficient algorithm for big datasets proposed by Evensen *et al.* (2019), and the local analysis scheme of Silva Neto *et al.* (2021). In this section, we present a method summary, highlighting the main features of this algorithm.

The SEnRML method aims at minimizing the objective function

$$\mathcal{J}(\boldsymbol{w_j}) = \frac{1}{2}\boldsymbol{w_j}^T\boldsymbol{w_j} + \frac{1}{2}\left[\boldsymbol{g}(\boldsymbol{x}^a) - \boldsymbol{d_j}\right]^T \boldsymbol{C}_{dd}^{-1}\left[\boldsymbol{g}(\boldsymbol{x}^a) - \boldsymbol{d_j}\right], \tag{5}$$

where $\boldsymbol{g}(\boldsymbol{x}^a)$ is the forward simulation model as a function of the updated parameters, $\boldsymbol{x}^a$, the variable $\boldsymbol{d}_j$ is the perturbed observed data and it follows the distribution $\mathcal{N}(\boldsymbol{d}^{obs}, \boldsymbol{C}_{dd})$, with a covariance matrix of measurement errors $\boldsymbol{C}_{dd}$, and $\boldsymbol{w}_j$ are column vectors that define the changes in the parameters during the calibration for each model. Therefore, the first term on the right of Equation (5) relates to the distance to the prior ensemble, and the second one refers to the data misfit. These two terms form the total cost function of the Bayesian methods (Evensen 2009). The ensemble of updated parameters forms the matrix

$$\boldsymbol{X}^a = \boldsymbol{X}^f + \boldsymbol{AW}, \tag{6}$$

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

$$\boldsymbol{A} = \boldsymbol{X}^f \frac{1}{\sqrt{\boldsymbol{N}-1}}\left(\boldsymbol{I_N} - \frac{1}{\boldsymbol{N}}1_{\boldsymbol{N}}1_{\boldsymbol{N}}^{\boldsymbol{T}}\right) = \boldsymbol{X}^f\Pi_{\boldsymbol{N}}, \tag{7}$$

where N is the ensemble size and the projector Π_N removes the mean and normalizes the matrix by $\sqrt{N-1}$. Note that the change of variables defined in Equation (6) means that the model updates are a linear combination of the prior ensemble anomalies. The algorithm updates the matrix $\boldsymbol{W}$, which defines this linear combination.

One obtains the iterative procedure to update the matrix $\boldsymbol{W}$ and the parameters applying the Gauss–Newton method in the cost function defined by Equation (5). After some manipulations, it is possible to find

$$\boldsymbol{W}^{i+1} = \boldsymbol{W}^i - \gamma\left[\boldsymbol{W}^i - (\boldsymbol{S}^i)^T\left(\boldsymbol{S}^i(\boldsymbol{S}^i)^T + \boldsymbol{C}_{dd}\right)^{-1}\boldsymbol{H}^i\right], \tag{8}$$

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

$$\boldsymbol{S}^i = (\boldsymbol{D}^{sim})^i \Pi_N (\boldsymbol{I}_N + \boldsymbol{W}^i \Pi_N)^{-1}, \tag{9}$$

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

$$\boldsymbol{H}^i = \boldsymbol{S}^i \boldsymbol{W}^i + \boldsymbol{D} - (\boldsymbol{D}^{sim})^i, \tag{10}$$

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

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

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

In the local analysis scheme, we divide the DA problem into independent analyses. In each one of them, we update a predefined subset of the parameters, called local group, using only the part of the dataset that we assume correlated to the group. Each problem follows the same DA procedure, described in Equations (6)–(10).

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

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

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

$$\boldsymbol{\theta} = \frac{1}{\sqrt{\boldsymbol{N}}} \sqrt{2\ln(\boldsymbol{n}_{ac})}, \tag{11}$$

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

$$\boldsymbol{z} = \max\left(1.67 - 0.67\frac{|\boldsymbol{r}|}{\boldsymbol{\theta}}, 0\right), \tag{12}$$

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

2.2.2. Computational efficiency of the SEnRML-LA method. A detailed analysis of the computational requirements with increasing dataset size was performed by Silva Neto *et al.* (2021). The authors made it clear the advantages of the SEnRML-LA method compared to other ensemble-based methods. They carried out several tests in a controlled environment (a dedicated computer) to measure accurately several important indicators such as memory consumption, comparing the SEnRML-LA with the ES-MDA with Kalman gain localization. The authors showed that the ES-MDA memory requirement increased with the square of the dataset size while the growth of the SEnRML-LA memory requirement was linear to the dataset size. This makes it possible to apply the SEnRML-LA to assimilate big datasets, which is the case of the present study.

It is worth nothing that the comparison of memory consumption between ES-MDA and SEnRML-LA may depend on the implementation strategy. In their comparisons, Silva Neto *et al.* (2021) updated the parameters using 5000 rows of the Kalman gain at a time, to reduce the memory requirements at this stage of the analysis scheme. More details about this ES-MDA implementation can be found in Appendix B of Emerick (2016). Other implementations or updated schemes could also be possible.

2.3. *Normalized quadratic deviation with sign (NQDS)*

To assess the quality of the well data matching on a well-by-well basis, we used the metric NQDS, defined for each data series, j (for example, producer P1 oil rate, producer P2 water rate, and so on) as follows:

$$NQDS_j = \frac{QDS_j}{AQD_j}, \tag{13}$$

where the normalization term AQD stands for acceptable quadratic deviation and is defined according to equation (14):

$$AQD_j = \sum_{i=1}^{Nobs} \left(Tol \times d_i^{obs} + C_p\right)^2, \tag{14}$$

where N_{obs} is the number of observed data, d_i^{obs} is the observed data, and Tol and C_p are tolerances applied to the observed data. The quadratic deviation with sign (QDS) is defined as

$$QDS_j = \frac{LD}{|LD|} \sum_{i=1}^{Nobs} \left(d_i^{sim} - d_i^{obs}\right)^2, \tag{15}$$

where d_i^{sim} is the simulated data and LD (Equation (16)) is the linear deviation used only to define the sign ($LD/|LD|$). The sign is useful to determine the position of the simulated curve in relation to the observed data. In other words, the sign is important to evaluate if an ensemble of simulated curves is biased or not with respect to the history data. If the sign is positive, the simulated curve is above or predominantly above the history, and vice versa:

$$LD = \sum_{i=1}^{Nobs} \left(d_i^{sim} - d_i^{obs}\right). \tag{16}$$

More details about the NQDS can be found in Avansi *et al.* (2016) and Maschio and Schiozer (2016).

2.4. *Global objective function*

To assess the evolution of the SEnRML-LA along the iterations in terms of data mismatch, we computed the cost function for well and 4DS data separately. For well data ($\mathcal{L}_w$), we applied Equation (17), considering a diagonal measurement error matrix:

$$\mathcal{L}_w = \sum_{i=1}^{Nobs} \frac{\left(\boldsymbol{d}_i^{sim} - \boldsymbol{d}_i^{obs}\right)^2}{\boldsymbol{\sigma}_i^2}, \tag{17}$$

where σ_i^2 is the variance of each measured data point, identified by the index i. For 4DS data, we applied the data term of Equation (5) with data error correlations described in the end of Section 3.3.2. It is important to highlight that the method SEnRML (intrinsically) minimizes the whole cost function (following Bayesian formalism). But, in practice, it is useful to analyze (after the DA process) the data mismatch terms separately.

3. Case description and problem setup

3.1. *General case description (UNISIM-IV)*

The case studied in this work is the UNISIM-IV, a reservoir model based on the UNISIM-III benchmark case representing a giant field composed of fractured carbonate karst reservoir from the pre-salt province (Correia *et al.* 2020). The UNISIM-IV is a sector of UNISIM-III and corresponds to one platform's drainage area. The reference model for the UNISIM-IV (UNISIM-IV-R) is represented by a corner-point grid with 170 × 157 × 595 blocks and the dimensions of each block are 50 × 50 × 2 m. The UNISIM-IV-R is a high-resolution geo-cellular model that represents our "real response" with challenges typical of a real field and the total simulation time is ~5 days in a Linux computer with 16 processors.

To create the simulation model, a geological model was firstly created using the same resolution of the

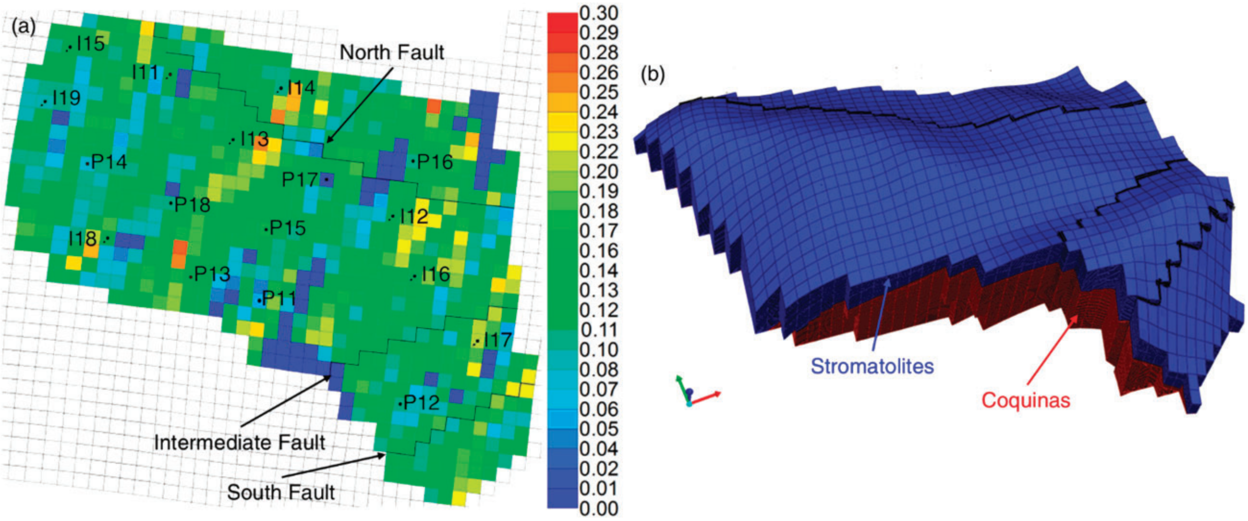

Figure 2. The UNISIM-IV model. The figure depicts an intermediate layer porosity map sample from the prior ensemble (a), including well and fault locations, and the 3D model (b), highlighting the two zones, stromatolites and coquinas. In the well's names, the letter P refers to a producer and I to an injector.

UNISIM-IV-R. This geological model was constrained to the information of well logs. Since it is not feasible to run many simulation models using high-resolution grid, it was necessary to make an upscaling process to a coarser grid. Finally, the simulation model is represented by a corner-point grid with 47 × 39 × 291 blocks. The reservoir simulation model represents partial information from the synthetic truth at the wells' locations. Therefore, the reference model contains unknown characteristics during the simulation studies, mimicking a real field application. The model, depicted in Fig. 2, has two zones. The upper zone consists mainly of stromatolites, while the lower zone corresponds to the coquinas. We present the model geometry and these zones in Fig. 2b. There are three faults in this sector, whose locations are indicated in Fig. 2a. The production strategy consists of 17 vertical wells, eight producers, and nine injectors. The recovery strategy in this benchmark case assumes that the injectors reinject all the produced gas in the reservoir. Furthermore, each injection well operates in WAG cycles of 6 months, except for well I16, which only injects gas. We control the total water injection rate to maintain the average reservoir pressure at a target value of 61 000 kPa (Botechia *et al.* 2021). Table 1 lists the general models' characteristics. For more details regarding the geological models, we refer to Correia *et al.* (2020).

Table 1. General simulation model characteristics.

Characteristic	Value
Horizontal permeability (mD)	0 to 9000 (median ≈ 50)
Vertical permeability (mD)	0 to 900 (median ≈ 1)
Porosity	0 to 0.3 (median ≈ 0.11)
Average depth (m)	≈5500
Initial datum pressure (kPa)	63 000
Average cell size (m)	≈200 × 200 × 5
Number of active cells	77 071
Total number of cells	533 403

Table 2. General fluid properties.

Characteristic	Value
API (°)	27
Initial GOR (m^3/m^3)	415
Oil formation volume factor (m^3/m^3)	2
Initial saturation pressure (kPa)	≈49 000
Initial oil viscosity (cP)	≈0.4
CO_2 in the gas phase (%)	44
CO_2 in reservoir fluid	37%
Temperature at the reservoir (°C)	90

The benchmark reservoir model, which emulates challenges and difficulties typical of complex real fields, includes hydrocarbon fluids whose characteristics reproduce the public report (Petrobras 2015) regarding a pre-salt field. We present the reservoir fluids' main characteristics in Table 2. Among them, it is worth highlighting the high CO_2 content, ~40%. Furthermore, the fluid volatility is associated with the initial gas-oil ratio (GOR) of 415 and the high oil formation volume factor of 2. The oil phase at reservoir conditions presents low viscosity, ~0.4 cP, which is favorable to the recovery process sweep efficiency. As the fluid consists of light oil with a high content of CO_2, we used a compositional fluid model for the reservoir simulations.

All well completions include interval control valves (ICV), allowing the flow control in two zones for the injectors and three intervals for the producers. The injectors' ICV aim at uniformizing the injection of water and gas between

Table 3. Scalar and categorical attributes summary.

Attribute	Type	Quantity	Minimum[a]	Maximum[a]
Fault transmissibility multipliers	Scalar (continuous)	3	−8	0
Well index multipliers	Scalar (continuous)	44	−2	2
Relative permeability tables (stromatolites)	Categorical (7 levels)	1		
Relative permeability tables (coquinas)	Categorical (7 levels)	1		
Equation of state	Categorical (3 levels)	1		

[a] Logarithmic ($\log_{10}$) transformed domain.

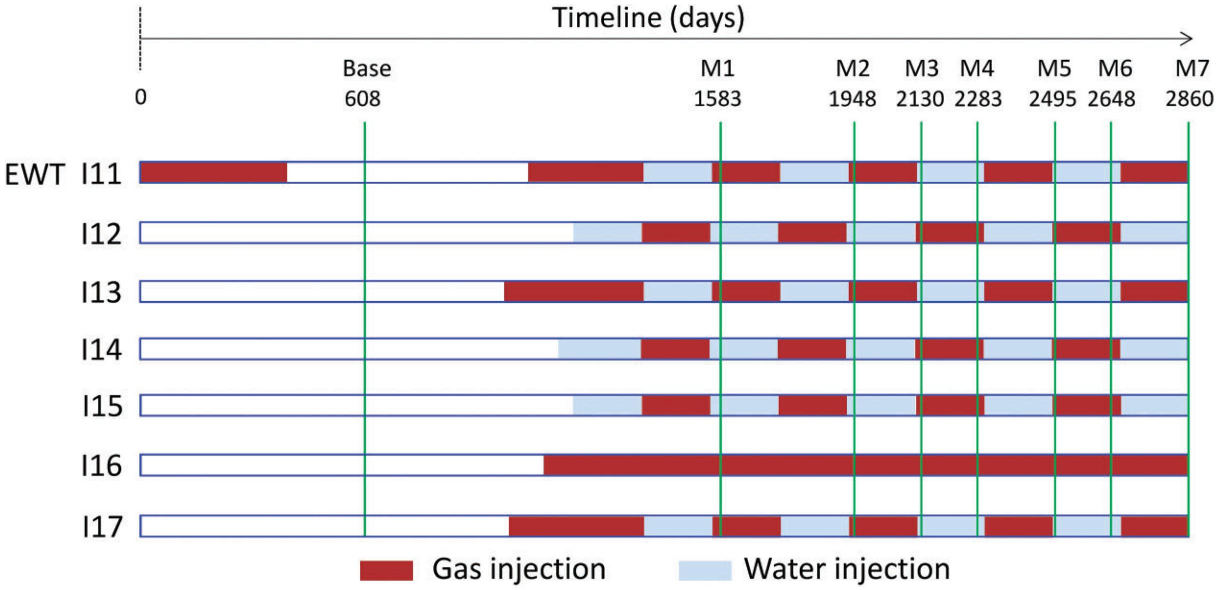

Figure 3. Monitors date according to injection scheme and WAG cycles (M stands for monitor and the extended well test (EWT) is for I11).

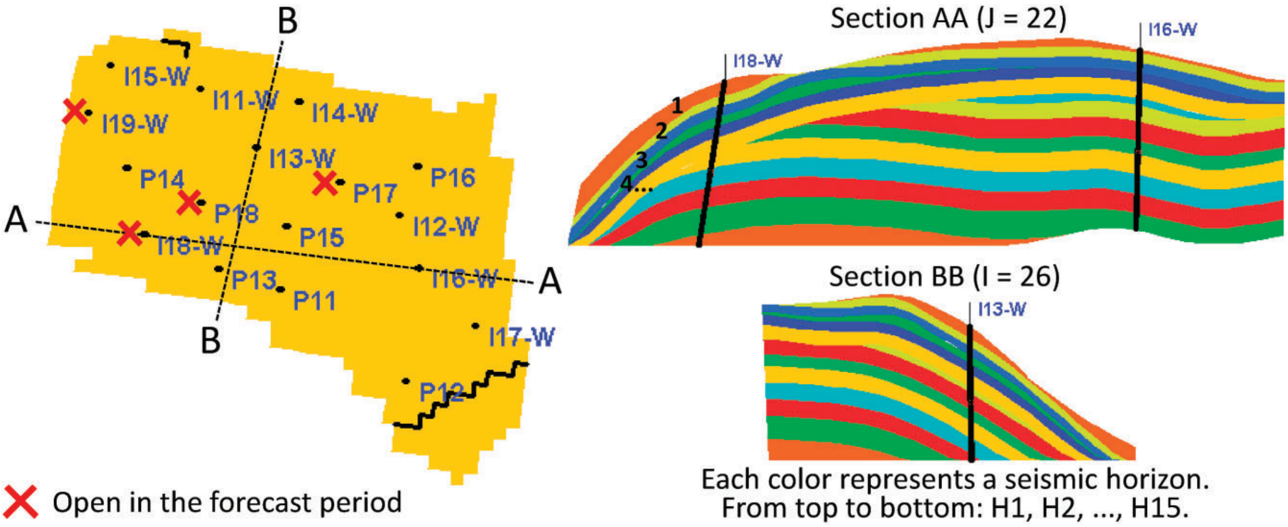

Figure 4. Illustration of the seismic horizons for the UNISIM-IV.

the zones. Therefore, when it detects a predefined level of unbalance between the two intervals, the ICV closes the one that received a larger volume and opens the other. In the producers, the ICV aim at avoiding high GOR production from each interval. If it detects a GOR value above a predefined threshold, it closes the respective production zone. For more detail on the operation and optimization of the ICV for this benchmark, we refer to Botechia *et al.* (2021).

During the history-matching process, each well operates with measured rates as boundary conditions. The total

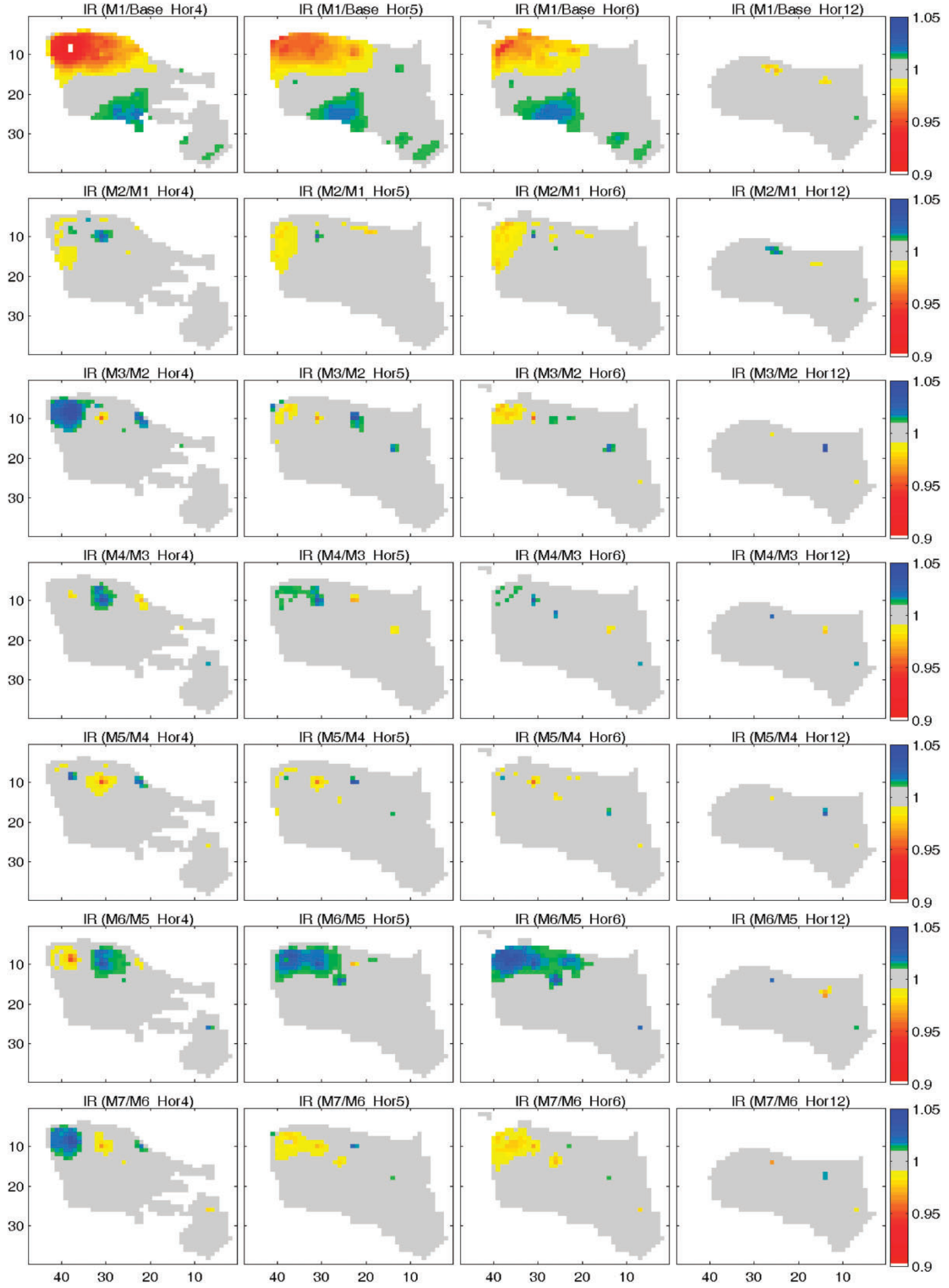

Figure 5. Observed time-lapse maps (acoustic impedance ratios) for the seven pairs of monitors for the horizons 4, 5, 6, and 12.

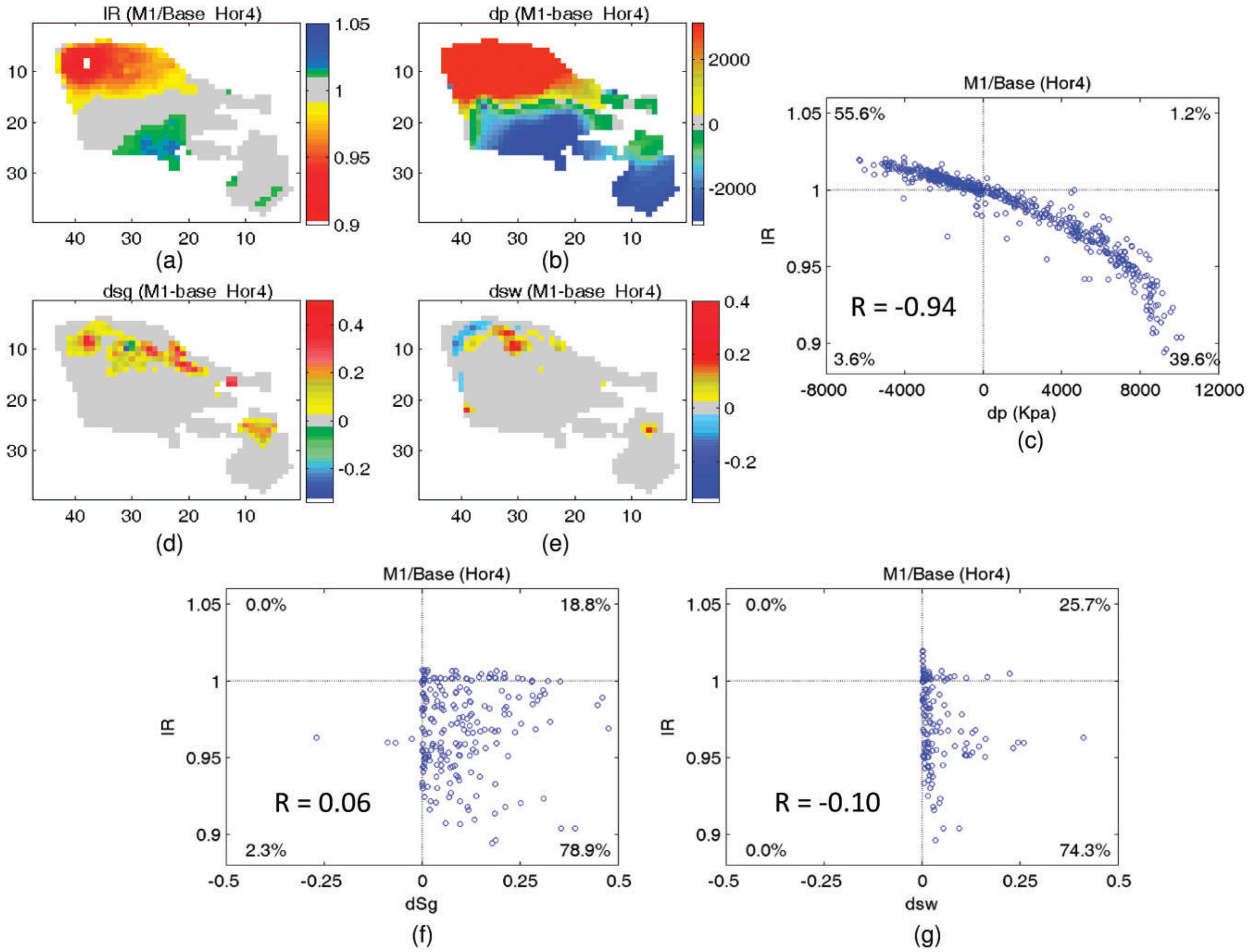

Figure 6. Analysis of the pair of monitors M1/Base for the Horizon 4: (a) map of IR, (b) map of pressure difference (dp), (c) crossplot between dp and IR, (d) and (e) gas and water saturation differences, respectively, (f) crossplot between dSg and IR, and (g) crossplot between dSw and IR.

liquid rate represents this condition for each producer, while the injected water or gas rate plays the same role for each injector. It is worth mentioning that all ICVs must reflect the same states as the actual operation during this period. Failing to report the valve restrictions in the forward simulation will act as a modeling error, which may impair the parameter calibration. Although not considered in this work, mechanical failure may be represented as an uncertain attribute and included in the DA workflow (Evensen 2019).

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

3.2. *Uncertain attributes*

We apply the DA workflow presented in Fig. 1 to calibrate the reservoir simulation models of the UNISIM-IV benchmark case. The simulation models represent partial information of the fine-scale reference model, and they carry geological uncertainties regarding the permeability fields at the three main directions (x, y, and z) and the porosity field. As shown in Correia *et al.* (2020), the porosity is correlated with horizontal permeability. We represent all the four grid-related parameters using an ensemble of 100 geostatistical realizations. The truncated Gaussian simulation is applied for modeling facies, whereas porosity and horizontal permeability were generated using Gaussian simulation. More details about the geological/geostatistical modeling can be found in Correia *et al.* (2020). During the generation of the initial (prior)

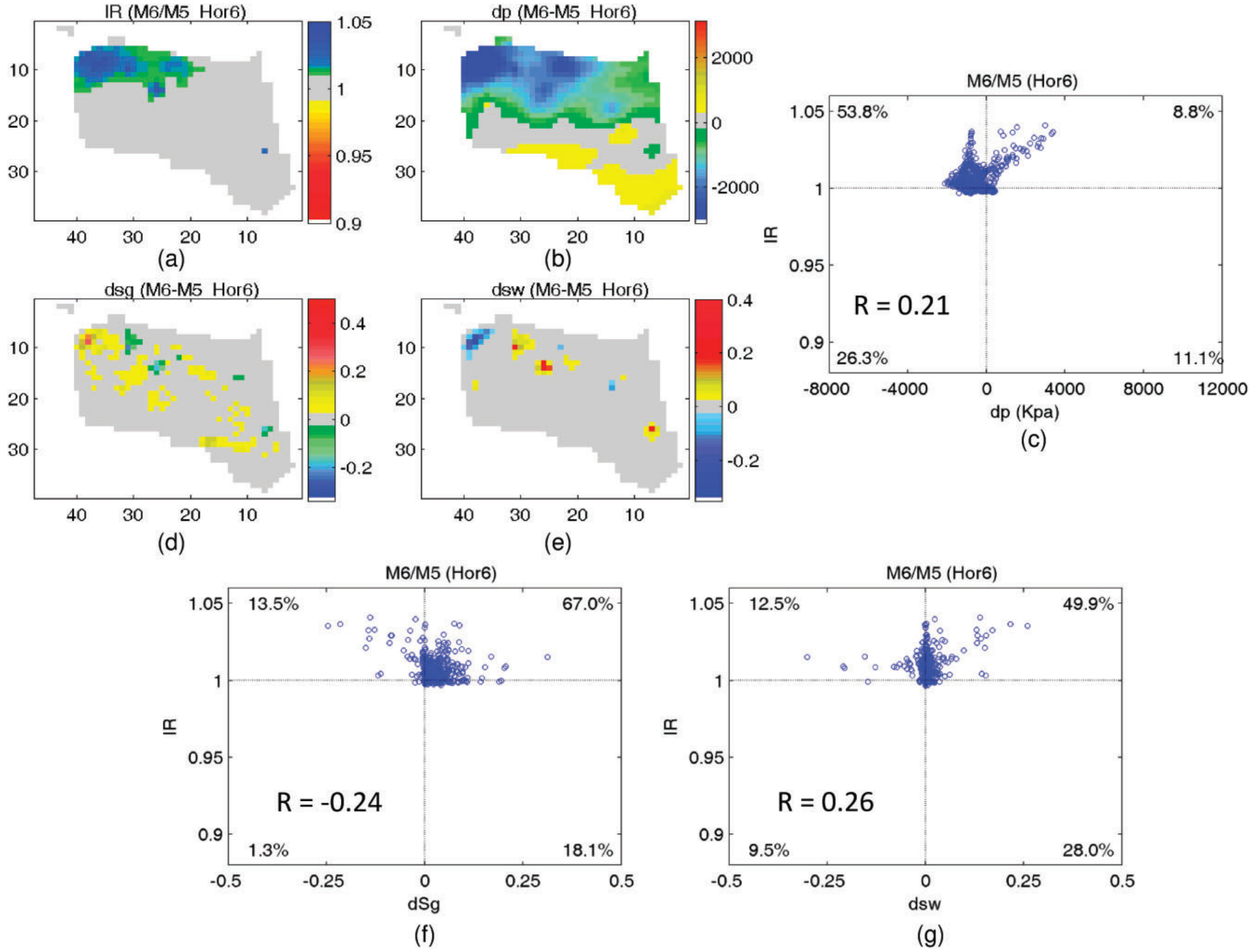

Figure 7. Analysis of the pair of monitors M6/M5 for the Horizon 6: (a) map of IR, (b) map of pressure difference (dp), (c) crossplot between dp and IR, (d) and (e) gas and water saturation differences, respectively, (f) crossplot between dSg and IR, and (g) crossplot between dSw and IR.

geostatistical realizations (in the geostatistical software modeling), the seed number is considered a random variable. However, we did not consider the seed as an update variable in the DA because, during the DA, the petrophysical properties are directly updated by the SEnRML method. This ensemble size is a tradeoff between the representation of the uncertainty statistics and the computational costs to simulate the forward models. Furthermore, we apply a logarithmic transformation to the permeability to compute the DA, the fourth step of Fig. 1. This transformation aims at approximating the problem to linear Gaussian since the permeability commonly has a log-normal distribution.

Besides the grid-related attributes, the benchmark case also considers 50 additional attributes (summarized in Table 3): 47 scalar attributes, $x \in \Re^{1\times 1}$, and three categorical variables. We apply a logarithmic transformation to all scalar parameters during the calibration process. First, there are three transmissibility multipliers related to the three faults presented in Fig. 2a. We represent the faults using a uniform distribution in the transformed domain, from closed to fully opened or equivalently from nearly zero transmissibility to 100% transmissibility. Secondly, each producer has uncertain productivity (well index) at each of the three intervals, represented by a multiplier with a Gaussian distribution, centered at 1 in the transformed domain. The same occurs in each injector but with two intervals. These intervals are related to the ICV operation. The fault transmissibility multiplier has a logarithmic impact in the flow behavior, which is convenient working with a logarithmic distribution. For the well index multiplier, the idea was to allow larger values, but less frequent (less spaced) than smaller ones.

The simulation model also assumes three categorical variables, the relative permeability curves of each zone (Fig. 2b) and the EOS. SEnRML, with local analysis, handles these variables by representing them with auxiliary continuous variables with a prior standard normal distribution. The algorithm selects thresholds to determine the discrete levels based on the prior probability of each category. During the calibration process, the analysis equation updates each model auxiliary variable to determine the calibrated categories.

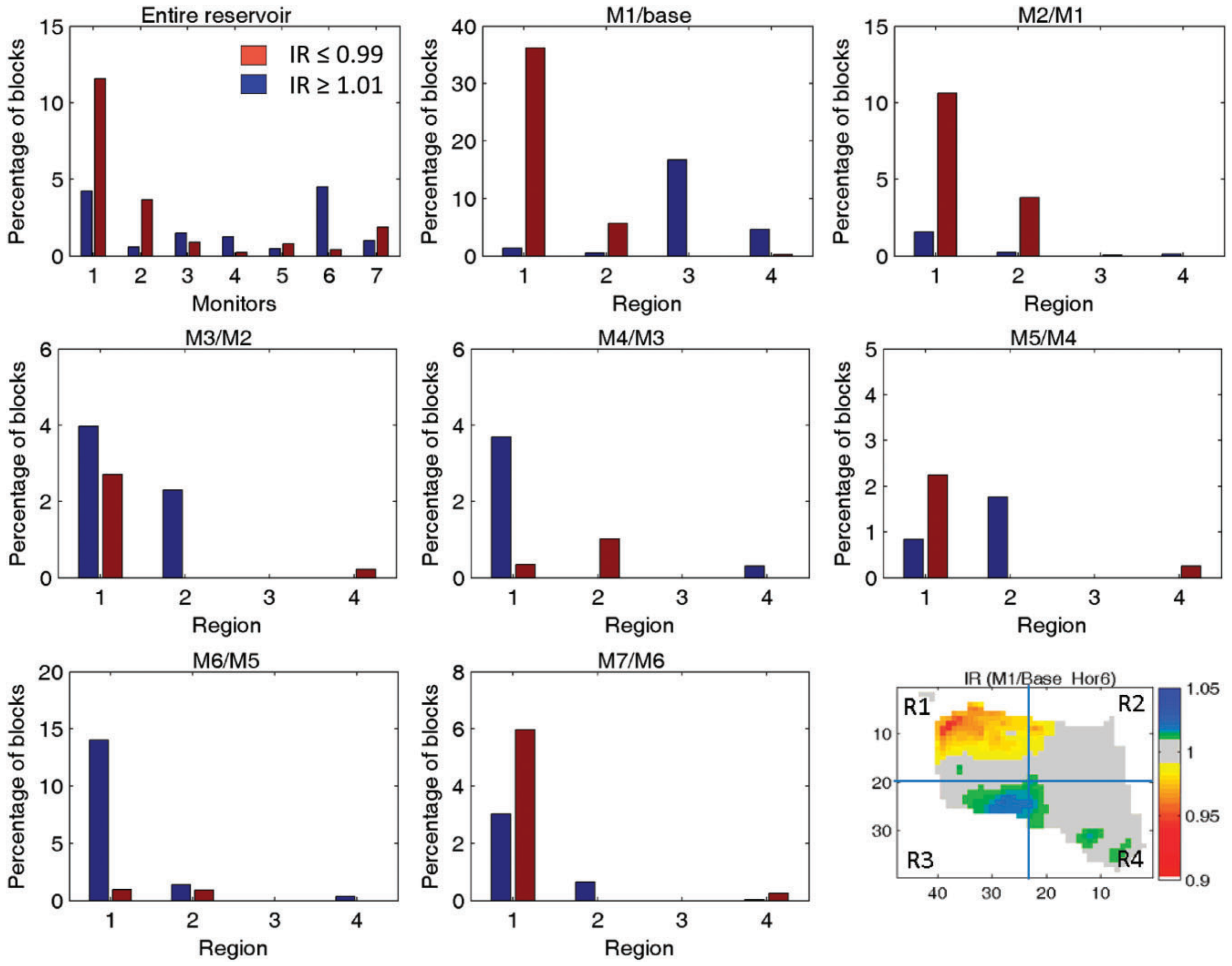

Figure 8. Impedance ratio for the field (seven pairs of monitors) and for the regions (depicted in the bottom-right image) for each pair of monitors.

Note that one should order each variable's categories in a way that generally enables a monotonic tendency between the discrete levels and the impacts on the simulated data. For instance, the relative permeability curves should cause an increase or decrease in water and gas production among the categories. This procedure aims to reduce the nonlinearity of the relation between the parameters and the simulated data.

Finally, the prior ensemble is composed of 100 geostatistical realizations combined with 50 scalar and categorical attributes.

3.3. *Observed data*

The observed data assimilated in this work was generated from the UNISIM-IV-R and is composed of well and 4D seismic data. There are 86 004 data points in total to be assimilated: 69 426 related to TLS data plus 16 578 related to well data. It is worth mentioning that this amount of data is very challenging for most ensemble-based assimilation methods. The description of each kind of data is presented as follows.

3.3.1. *Well data.* The well data considered in this study comprises oil, gas, and water rates, and BHP for each producer. Moreover, each injector provided the BHP and the individual rate, gas, or water, depending on WAG cycles. Each measured data contains uncorrelated noise with a standard deviation of 10% for the rates and ~2% for the pressures. This noise is applied to the reference output to generate more realistic data. For the well data measurement error, we assumed a diagonal covariance matrix (which is a common choice for production data) with a standard deviation equal to 10% of the observed data.

3.3.2. *4D seismic.* The 4D seismic dataset assimilated in this work comprises one baseline acquisition, after 608 days of the beginning of the extended well test and seven monitors in the times shown in Fig. 3, which shows the injection scheme according to the monitors date and WAG

Table 4. SEnRML with local analysis parameters.

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

Table 5. Operational constraints for the production forecast of the simulation models with physical limits for the platform and wells.

Constraint	Value	Applies to
Maximum oil rate (m^3/d)	28 617	Platform
Maximum liquid rate (m^3/d)	28 617	Platform
Maximum water production rate (m^3/d)	23 848	Platform
Maximum water injection rate (m^3/d)	35 771	Platform
Maximum gas production rate (m^3/d)	12 million	Platform
Minimum BHP (kPa)	History period last value	Producer well
Maximum BHP (kPa)	History period last value	Injector well
Maximum liquid rate (m^3/d)	8000	Producer well
Maximum gas injection rate (m^3/d)	4 million	Injector well
Maximum water injection rate (m^3/d)	10 000	Injector well

cycles. It is worth mentioning that the benchmark UNISIM-IV considers permanent seismic monitoring, where most of the monitors were “acquired” every 6 months. We generated the observed data by applying a PEM in the fine-scale reference model results. After that, we extracted maps considering the vertical resolution of seismic data for this case, a permanent monitoring system, and performed scale transference to the simulation scale. At this scale, each seismic map corresponds to five to ten model layers. Each map consists of the acoustic impedance ratio between two consecutive monitors, representing the flow-related variations. There are 15 seismic horizons, as illustrated in Fig. 4, totaling 105 maps to be assimilated, being seven pairs of surveys multiplied by 15 horizons.

Starting with a qualitative analysis of the 4D seismic data, Fig. 5 shows the observed time-lapse maps (acoustic impedance ratios) for the seven pairs of monitors for the horizons 4, 5, 6, and 12. In these maps, “hot” colors (yellow to red) represent impedance reduction (softening effect) and “cold” colors (green to blue) represent impedance increase (hardening effect). To highlight the most important anomalies, impedance variations lower than 1% were set to gray. From the maps in Fig. 5, several important aspects can be highlighted:

(i) The WAG cycles anomalies are clearly identified by changes in the color between successive pair of monitors; see, for example, “M1/Base” to “M3/M2” for the Horizon 4 (column 1) and “M6/M5” to “M7/M6” for the Horizon 6 (column 3). This behavior in impedance ratio variation indicates the injected fluids changing: that is, injected water after a period of injecting gas in the porous media and vice versa. This behavior is predominant in the northwest region that contains most injector locations.

(ii) In the northwest region of the M1/Base, Horizons 4, 5, and 6, there is a huge softening-anomaly, which is a combination of pore-pressure with localized increased gas saturation. In this region, the acoustic impedance variation (reduction) reaches 10%. This happens because of a strong pressure increase caused by injection and a volcanic rock that is acting as a pressure barrier (Davolio *et al.* 2021).

(iii) In the center-south region of M1/Base, Horizons 4, 5, and 6, there is a huge hardening anomaly where the impedance increase reaches ∼5%. As can be seen from Fig. 2a, this region contains only producer wells. Generally, hardening-anomalies are typically found in vicinity of water injector wells. However, this is not the case here. To better understand these anomalies, we analyzed the correlation between the pressure differential and the acoustic impedance ratio. Figure 6a and b show the maps of IR and dp (pressure difference between M1 and Base), respectively, for M1/Base Hor4. Note that the signature of the IR anomalies is very similar to the pressure anomalies. It is noteworthy that this analysis is possible because we have the dp map from the reference model (ground truth). To complement the

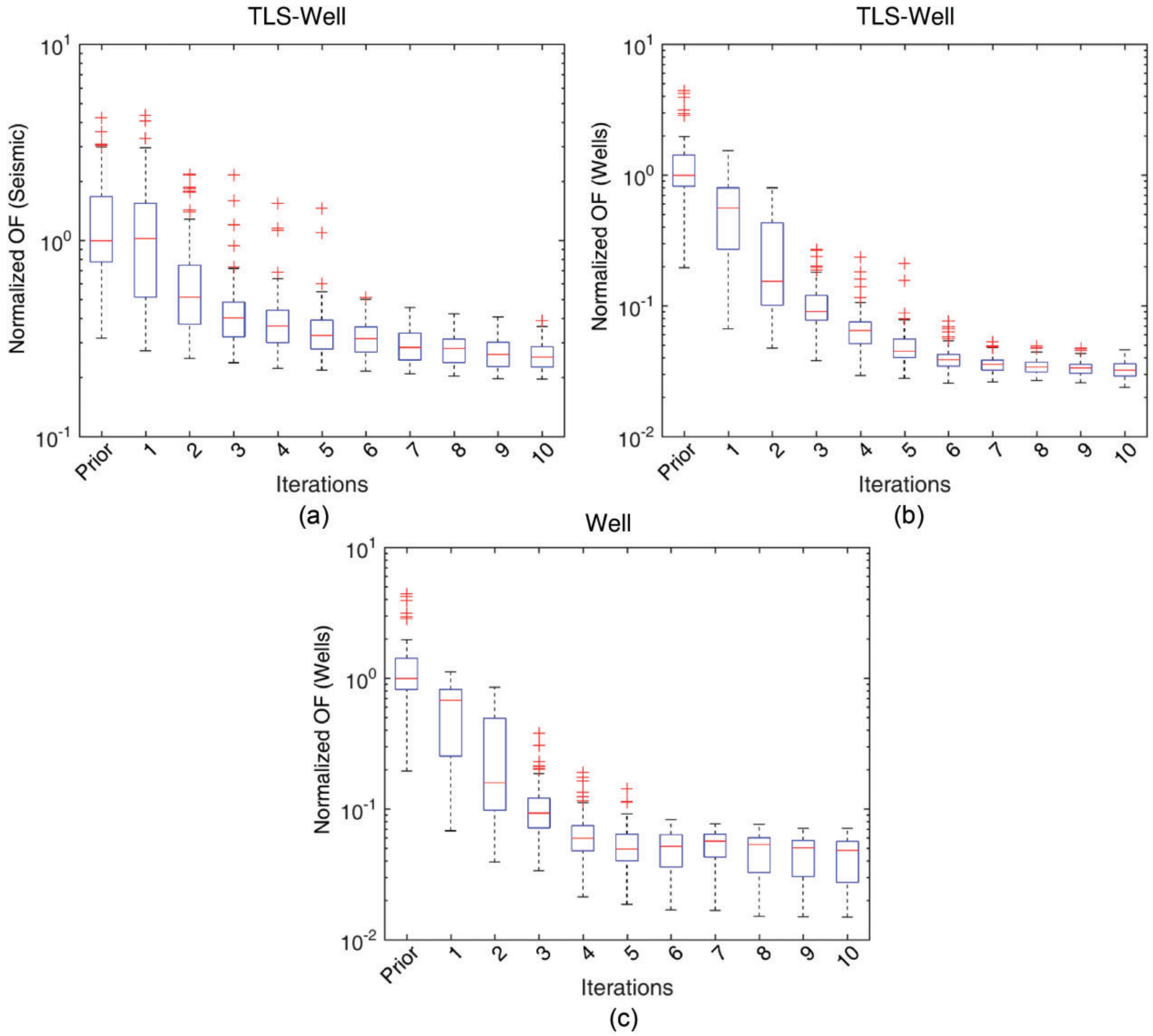

Figure 9. Evolution of the objective function along the iterations for the seismic (a) and wells (b) data for the case TLS-Well and wells data (c) for the Well case.

analysis, Fig. 6c depicts a strong correlation between dp and IR showing that the hardening anomaly in the center-south region is caused by a pressure drop (depletion).

(iv) Still, considering M1/Base, the effect of injected fluid saturations in the acoustic impedance ratio is also analyzed in Fig. 6d–g. Figure 6 parts d and e show, respectively, the gas and water saturation differences between M1 and Base (Horizon 4), and Fig. 6f and g show the corresponding cross-plots. We can see that there is a poor correlation between the fluid saturations (gas and water) and the impedance ratio. The concentration of points in the bottom-right quadrant in the plots in Fig. 6f and g is due to the predominance of the pressure effect.

(v) Figure 7 shows the same analysis (presented in the previous paragraphs) for M6/M5 (Horizon 6). Based on this figure, we can see that for later monitors, there is a balance between the influence of pressure and fluid situations in the impedance variation. The correlation between dp and IR is not too strong as in M1/Base.

Additional analysis of the seismic data along the monitors is presented in Fig. 8. The histograms exhibit the percentage of blocks whose impedance ratio is ≤0.99 (red) and ≥1.01 (blue), representing impedance variation (decrease or increase) ≥1%. The IR for the entire reservoir is shown in Fig. 8 (top-left plot) where it can be noted an alternation between softening and hardening effects along the monitors. In Monitors 1, 2, 5, and 7 there is predominance of softening effects and in others, the hardening effects predominate. To go into more detail on the IR variation, we divided the reservoir in four regular regions, as shown in the bottom-right image of Fig. 8, and computed the percentage of blocks with

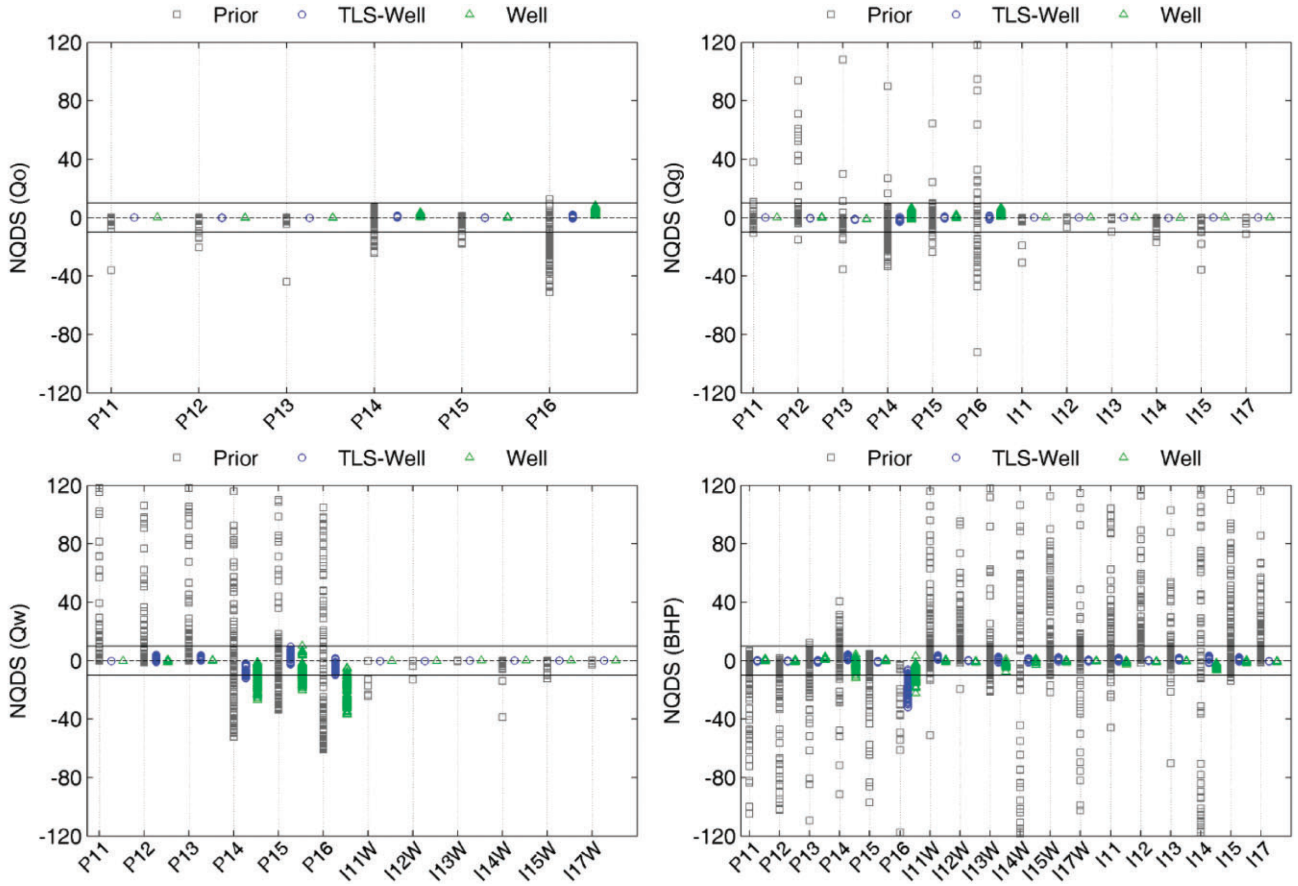

Figure 10. NQDS for oil rate (Q_o) for the producer wells, water (Q_w) and gas (Q_g) rates and BHP for producers and injectors comparing the prior ensemble (gray) and the posterior ensemble for the cases TLS-Well (blue) and Well (green). The closer the NQDS is to the range centered at the zero line, the better the results are.

significant hardening and softening effects, similar to what we did for the entire reservoir, however, in this case for each region and pair of monitors.

Similar to the entire reservoir, we can see (Fig. 8) that there is an alternation in softening and hardening effects along the monitors (due to WAG cycles). The alternation of positive and negative anomalies indicates that the fluid substitution plays a key role in the observed data behavior of the monitor pairs. It can also be noted that most changes occur predominantly in Region 1 (northwest part of the reservoir), where there are more injectors and the volcanic rock previously mentioned. Another point is that the IR variation tends to reduce along the monitors. This suggests that physical effects with opposite influences, hardening and softening, may attenuate the time-lapse signal.

We highlight that the 4D seismic data used here correspond to a quasi-ideal data, since they were generated from the application of the PEM to the reference model (without running the whole seismic forward modeling). Therefore, these data do not suffer from common problems of real data such as low vertical resolution and data repeatability issues. We named them "quasi-ideal" data because of the information loss caused by scale transference to the simulation scale. As the focus of the work is to evaluate the potential of SEnRML-LA to assimilate big data, using this 4D seismic data is a fair practice.

In the DA process, we perturbed the seismic data using Gaussian noise with standard deviation (σ_{4D}) equal to 0.2% (IR = 0.002) and 600 m exponential correlation length. We estimated the correlation length based on synthetic seismic modeling that considered the noise characteristics expected in permanent reservoir monitoring surveys. Furthermore, the standard deviation allowed balanced cost-function contributions between seismic and production data in the prior ensemble. To define the value of σ_{4D}, we simulated the prior ensemble models and computed the production data cost function for all models. Then, we computed the 4DS cost function for all models for several values of σ_{4D} comparing the cost-function distributions in a boxplot. The value of σ_{4D} that resulted in a 4DS cost-function distribution closer to the production data cost function was adopted. This strategy for the selection of the diagonal elements of the data error matrix tries to reduce the effect of the larger number of data points from the seismic acquisitions when compared to the

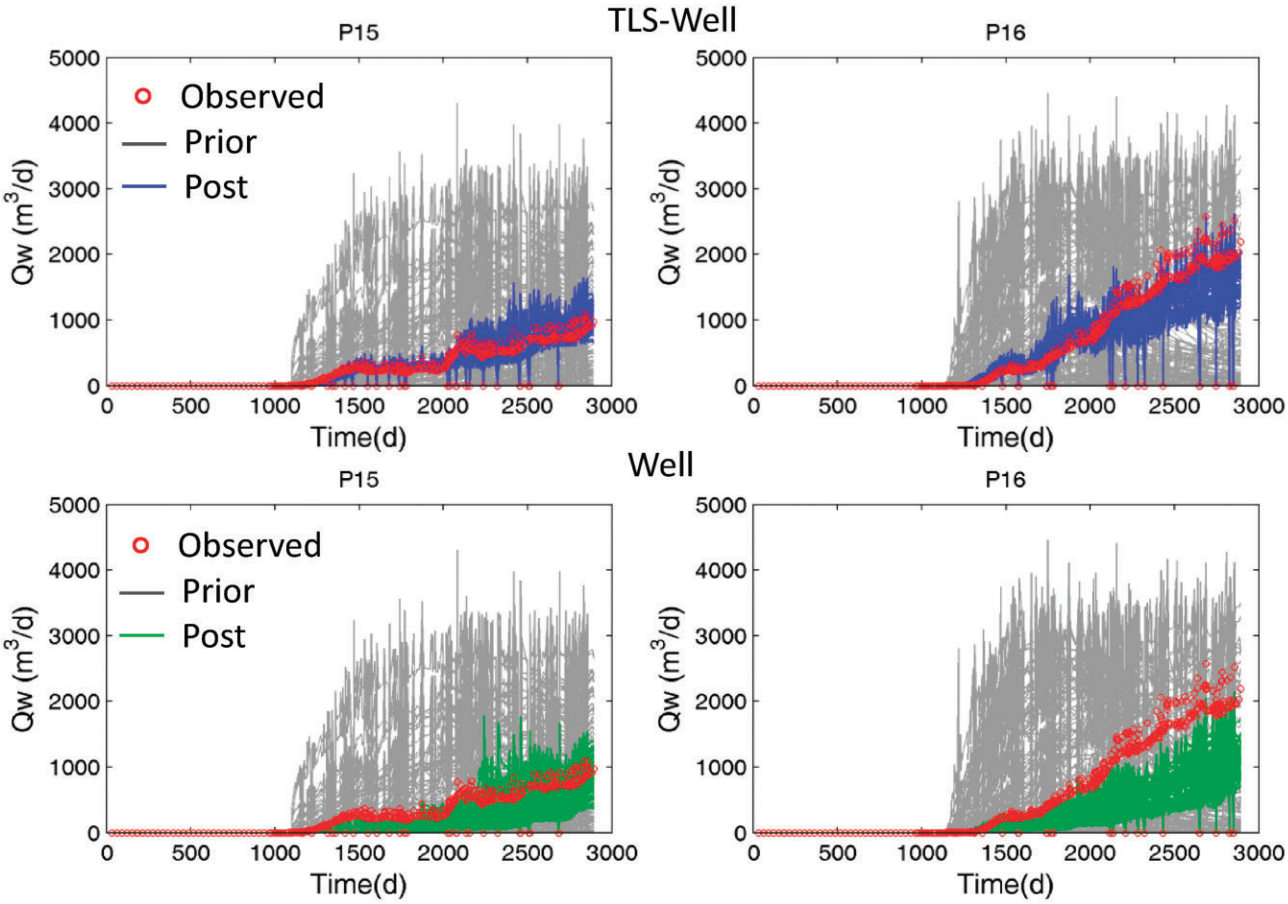

Figure 11. Water rate curves for P15 and P16 comparing the prior ensemble (gray) and the posterior ensemble for the cases TLS-Well (blue) and Well (green).

well data. Furthermore, it tends to reduce the risk of excessive uncertainty reduction with an associated risk of biasing the solutions.

3.4. *Petroelastic model*

The PEM applied in this work, the third step of Fig. 1, considers a mixture of three minerals in the rock, calcite, dolomite, and quartz, with fractions of 85, 11, and 4%, respectively. The dry-rock properties as a function of effective pressure were represented by logarithmic functions fitted using laboratory data regarding a pre-salt field, as detailed in Costa *et al.* (2016) and Silva *et al.* (2020). Declining exponential functions matching data from a pre-salt reservoir represented the dry-rock moduli as a porosity function were used. We considered different pressure and porosity laws for each reservoir zone, stromatolites, and coquinas. Vasquez *et al.* (2019) and Silva *et al.* (2020) reported that the Gassmann equation (Gassmann 1951; see Mavko *et al.* 2009) provides a reasonable representation of the fluid substitution in reservoirs similar to ours. Therefore, we apply the Gassmann equation to the current PEM. The Batzle and Wang (1992) equations represent the water phase bulk modulus, while the fluid model (Silva Neto *et al.* 2020) described in Section 2.1 enables computing the gas and oil bulk moduli and densities. It is noteworthy that, in general, there are uncertainties related to the petroelastic modeling. However, dealing with PEM's uncertainty during the DA process is an open research subject and it is not treated in this paper.

3.5. *Configuration of the SEnRML method with local analysis*

The SEnRML method with the local analysis scheme has some user-defined parameters. Table 4 shows the list of parameters defined for the application in UNISIM-IV. The well data localization lengths related to the drainage area and each well's influence region were estimated from streamlines (Emerick and Reynolds 2011, Soares *et al.* 2018). Ten iterations with 100 models per iteration were carried out for each case (TLS-Well and Well). Silva Neto *et al.* (2021) showed that 10 iterations of the method are enough to reach convergence.

3.6. *Production forecast*

The objective of the DA is not only matching the observed data. As part of the solution of the whole problem, it is

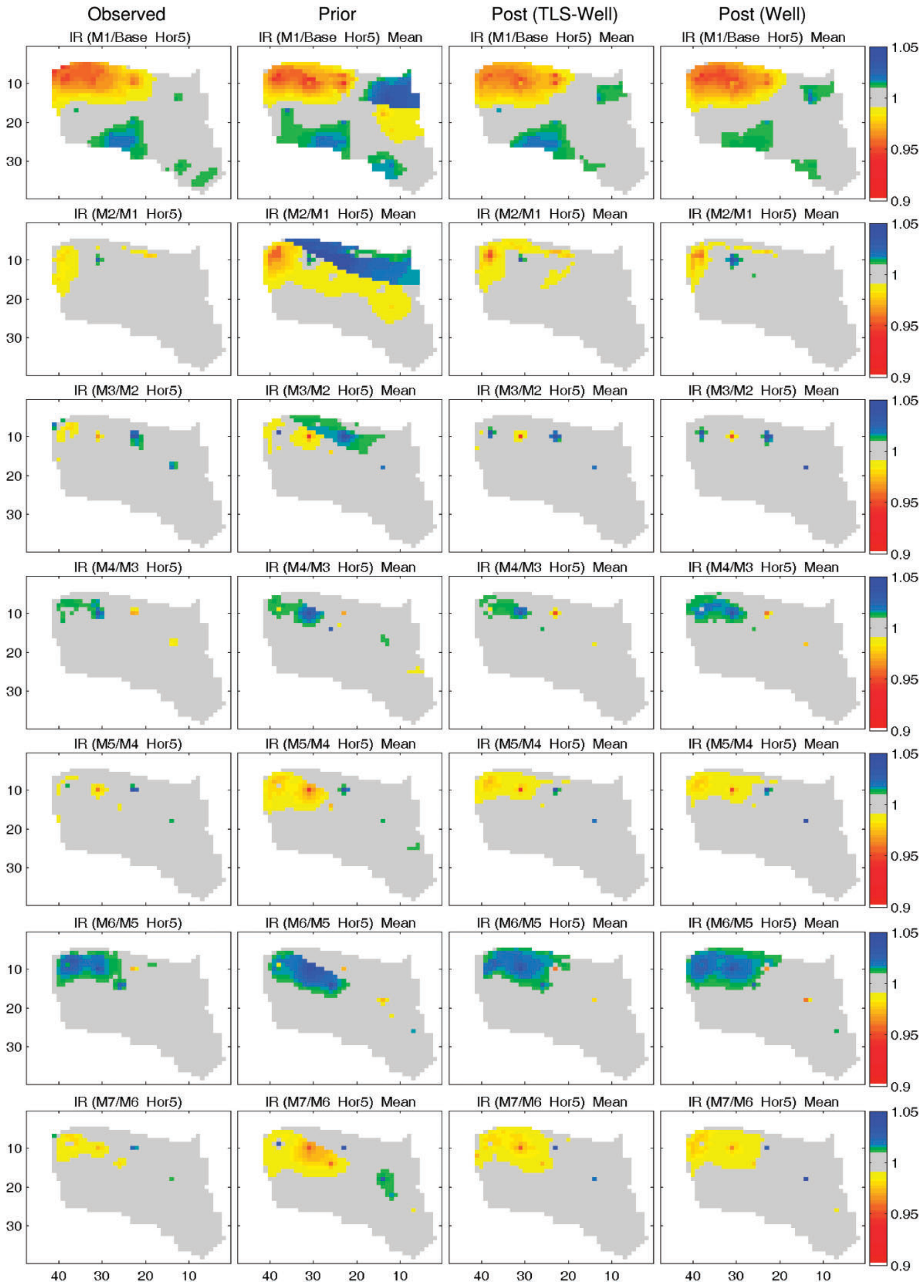

Figure 12. Time-lapse maps (acoustic impedance ratios) for the seven pairs of monitors for the Horizon 5 comparing: observed (first column), prior (second column) and posterior for cases TLS-Well and Well (third and fourth column, respectively).

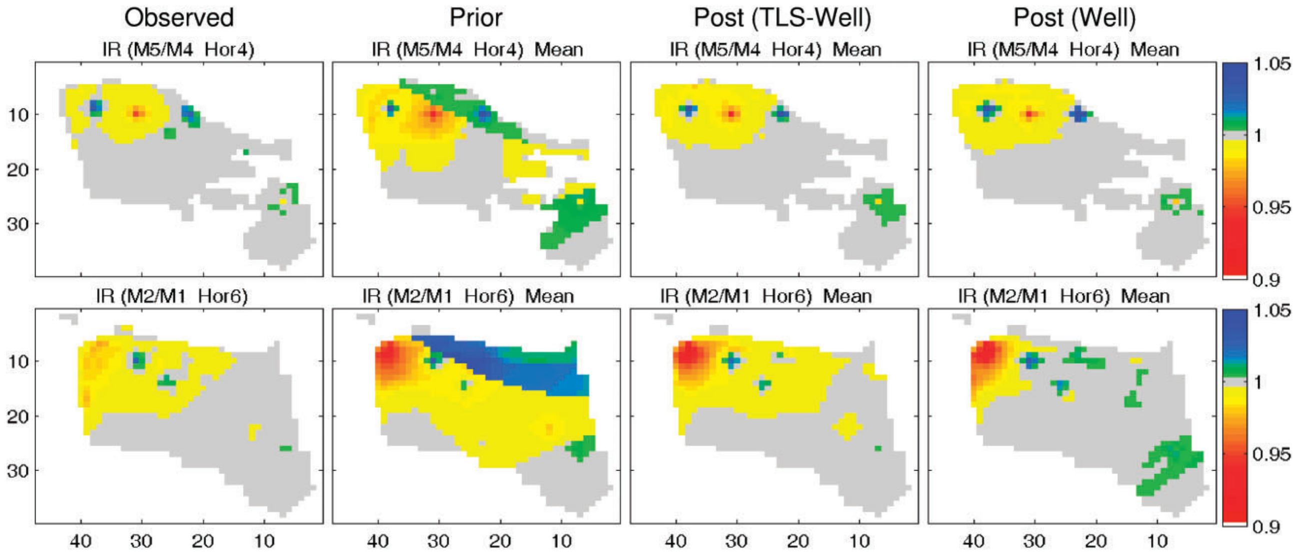

Figure 13. WAG anomalies resulting from injected water after a period of injecting gas in the porous media (bluish changes surrounded by a hot area in the northwest part of the reservoir).

essential to assess the forecast capability of the matched models. In this work, we perform, after the DA, a human-intervention free forecast, whose objective is to avoid the effects of reservoir management rules, for instance GOR and water-cut limits, and assess the models regarding the reservoir properties quality. The motivation to do so is that there are cases where a minor difference in a model can lead to a substantial effect due to differences in reservoir management operations, for instance, ICV changes and/or well shut-in due to GOR or water-cut limits. For example, two models may be very similar (in terms of reservoir properties) and due to a minimal difference between the simulated and allowable GOR, certain wells can close in one model but not in another, hindering the analysis. Therefore, for human-intervention free purposes, we ran the reference case and the simulation models without reservoir management rules. It is worth mentioning that for production strategy optimization problems, it is necessary to consider reservoir management rules. Table 5 shows the basic operational constraints for the production forecast of the simulation models. Note that we considered some physical limits for the platform and wells.

4. Results and discussions

This section is organized in five subsections: first, the evolution of the objective function (Subsection 4.1) is presented to show the global consistency of the method. In the following, the well (Subsection 4.2) and seismic (Subsection 4.3) data match are presented and discussed. The next subsection treats the analysis of the attributes' variability (Subsection 4.4) and, finally, the production forecast assessment is presented in Subsection 4.5.

4.1. *Evolution of the objective function*

Figure 9a and b show the evolution of the global objective function (see Section 2.4) along the iterations for the seismic (Fig. 9a) and wells (Fig. 9b) for the case TLS-Well. For each iteration (and the prior) a box plot generated using the 100 models is depicted. The OF is normalized for the prior median value so that the median shown in the box for the prior is equal to 1. Some points deserve consideration: (i) first, it is possible to note that there is a good convergence of the method, we can see a gradual and consistent reduction of the OF, showing consistency of the DA process; and (ii) there is a reduction of 75% in the seismic OF and a reduction of 97% in the wells OF, both related to the prior median value, which means an expressive OF reduction for both seismic and wells data.

Figure 9c shows the wells OF evolution for the Well case (DA using only well data). Although there are some minor differences, the overall shape of the OF evolution is similar to the case TLS-well. There is a reduction of 95% in the wells OF compared to 97% for the case TLS-Well. It is worth mentioning that the inclusion of TLS in the DA was beneficial to the process since it improved not only the seismic data matching but also well data matching. Rosa *et al.* (2023) also concluded that using 4DS data allowed generating better-calibrated models than assimilating production data only.

4.2. *Well data match*

The well-by-well data match, in terms of NQDS, is shown in Fig. 10 for oil rate (Q_o) for the producer wells, water (Q_w) and gas (Q_g) rates and BHP for producers and injectors. To compute the NQDS, we considered the tolerance as 10%

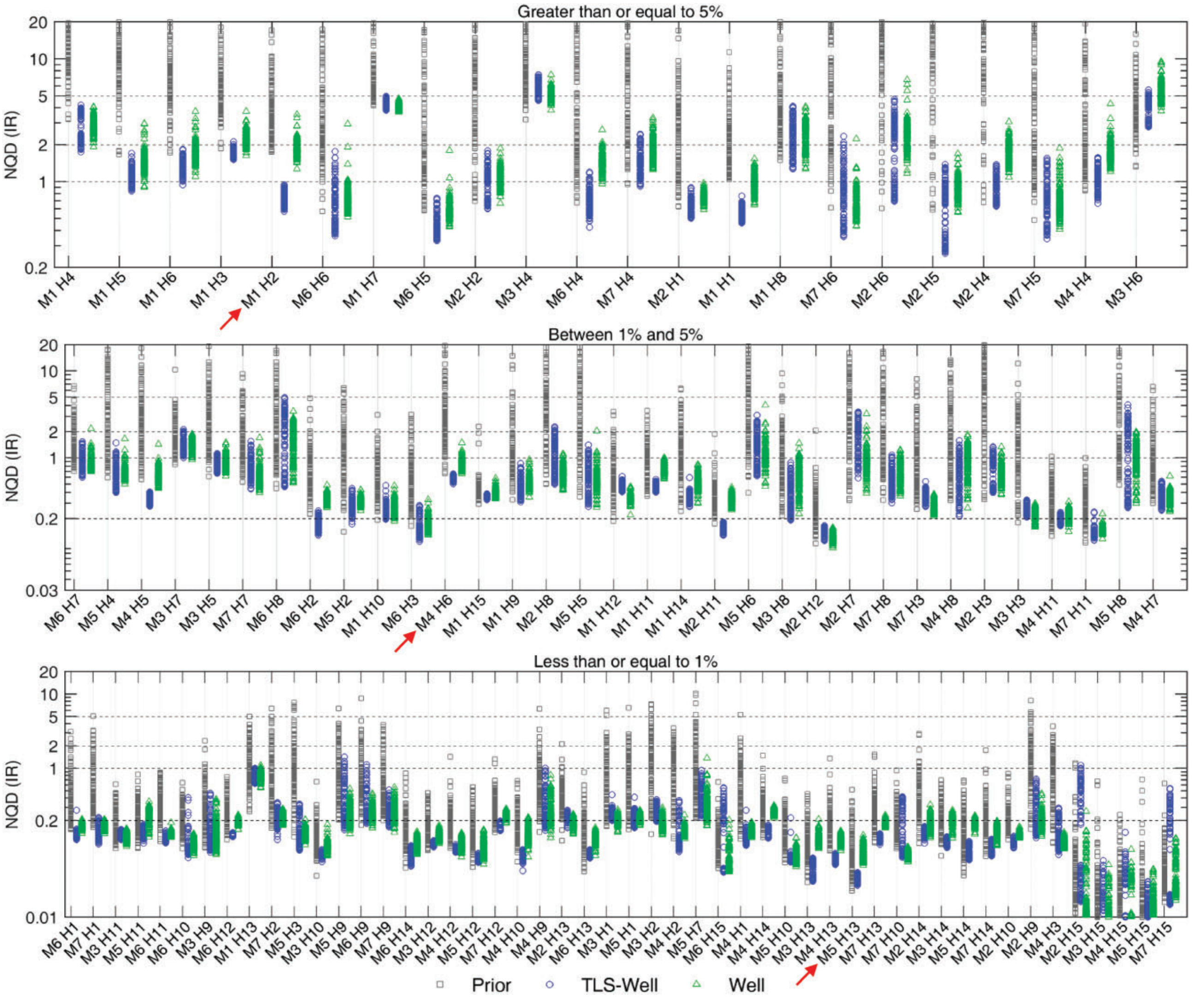

Figure 14. NQD for the 105 acoustic impedance ratios maps classified according to the percentage of blocks (three groups) with impedance variation greater than 1% (the red arrows indicate the monitors analyzed in Fig. 15). The closer the NQD is to the zero line, the better the results are.

$(Tol = 0.1)$ of the rate measurements, the constant (C_p) for the produced water rate is equal to 40 m^3/d, and the BHP tolerance is 3% $(Tol = 0.03)$.

There is a significant reduction in the NQDS distribution for the posterior ensemble for both cases (Well and TLS-Well) compared with the prior ensemble, which agrees with the analysis of the well OF shown in Fig. 9b and c. However, especially for the water rate, the results from TLS-Well are better for three out of the six producer wells (P14, P15, and P16). The NQDS plots show that the oil and gas rate were properly matched for both cases. Analyzing the BHP behavior from the prior ensemble, one can note that there is a significant pressure deviation with an underestimation tendency in the producers, while the injectors exhibit an opposite tendency.

The results in Fig. 10 for wells P14, P15, and P16 indicate that TLS data may improve well data match. At the first glance, this result seems unusual, as a new data source increases the complexity of the DA problem. However, well and TLS data are partially redundant. This redundancy act as an increased data weight and tends to shift the DA result toward a better data match, when searching the optimal solution of the summation of the two cost-function terms in Equation (5).

Figure 11 shows the water rate curves for P15 and P16 comparing the prior ensemble (gray) and the posterior ensemble for the cases TLS-Well (blue) and Well (green). The set of observed data points above the main curve of observed data is due to the well stops defined in the reference case. This happens due to the group control: when a well stop, the simulator increases the rate of other wells following apportionment rules.

Analyzing the posterior simulation curves, one can see that there is a trend of underestimation of water production

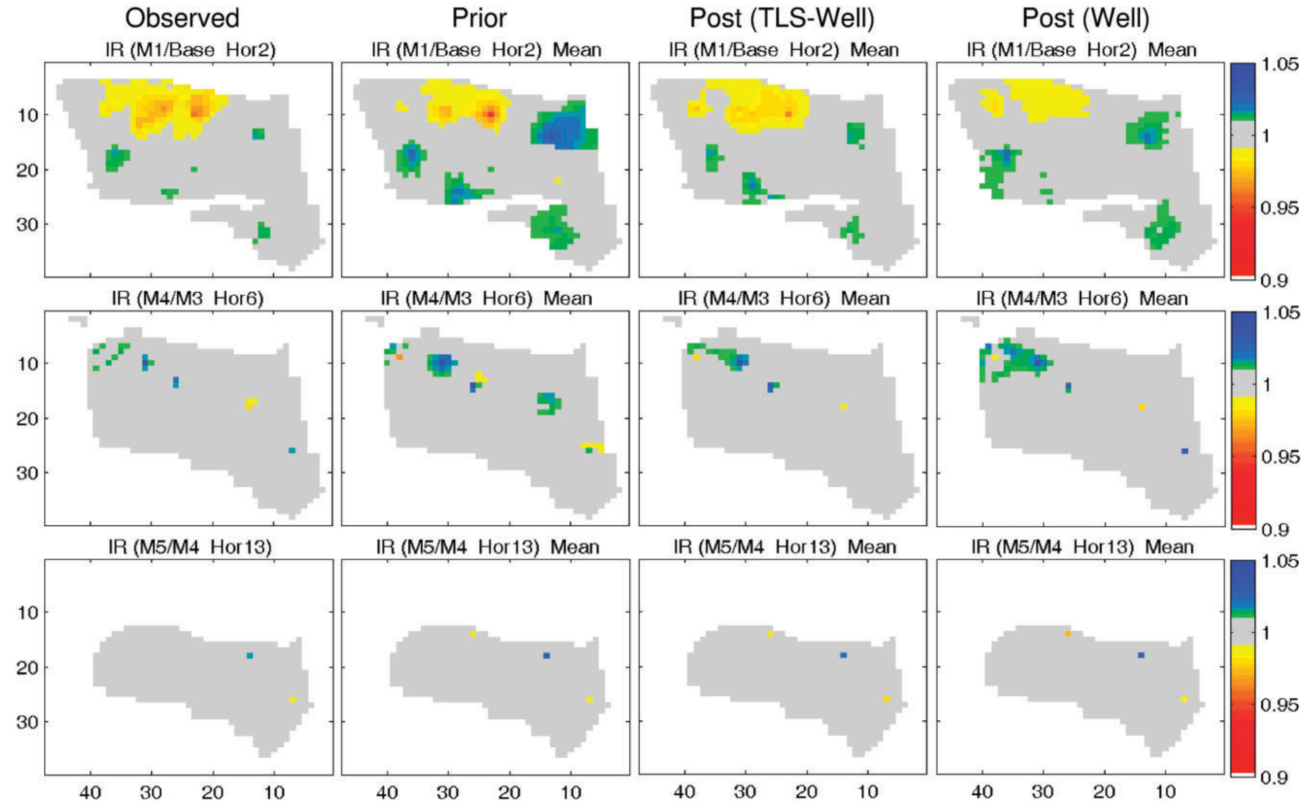

Figure 15. IR maps for each pair of monitors indicated with a red arrow in Fig. 14 representative of each group.

for the Well case, as also depicted in the NQDS plot (Fig. 10). On the other hand, the inclusion of the seismic data improved the well data match. The seismic data provides spatially rich information related to the reservoir fluid distribution and this allows the assimilation method to better detect the influence between a given attribute (relative permeability, for example) and the fluid movement, and as a result improving the well data match. It is noteworthy that the seismic dataset used in this work is not contaminated by issues related to real seismic acquisitions and processing, such as noise and repeatability problems, for example. However, even being a synthetic dataset, it is not perfect because there are scale issues, which cause information loss. In real cases, the more advanced the acquisition and processing technologies are, the higher the seismic quality is; in other words, the higher the capability of representing the true reservoir fluid distribution.

4.3. *Seismic data match*

This section shows in detail the seismic DA by assessing the match quality qualitatively and quantitatively. Starting with qualitative analysis, Fig. 12 shows the time-lapse maps (acoustic impedance ratios) for the seven pairs of monitors for the Horizon 5. The color legend follows the same scheme as Fig. 5. The first column is the observed data and the second is the prior ensemble mean. The third and fourth columns are the posterior (last iteration) ensemble mean for the cases TLS-Well and Well, respectively. For the Well case, we run the reservoir simulations again for the models of the last iteration to generate the data necessary to compute the impedance ratios, since during the DA for this case only well data were generated. From the maps in Fig. 12, some important aspects can be highlighted:

(i) The first three pairs of monitors show the compartmentalization of the prior models, characterized by the transition of colors from one side of the main fault to another, more pronounced in M2/M1.
(ii) In general, the high deviation in the prior models, especially in the three first pairs of monitors, was corrected after the DA for both cases, TLS-Well and Well. The post models follow the same alternation trend in the impedance ratio variation along the monitors verified in the observed seismic data.
(iii) For the set of maps shown in Fig. 12, the match quality of the cases TLS-Well and Well is qualitatively similar comparing with the observed maps, although there are some differences. For example, in the pairs M1/Base and M4/M3, the shape of some anomalies for the case TLS-Well is closer to the observed maps compared to the Well case.

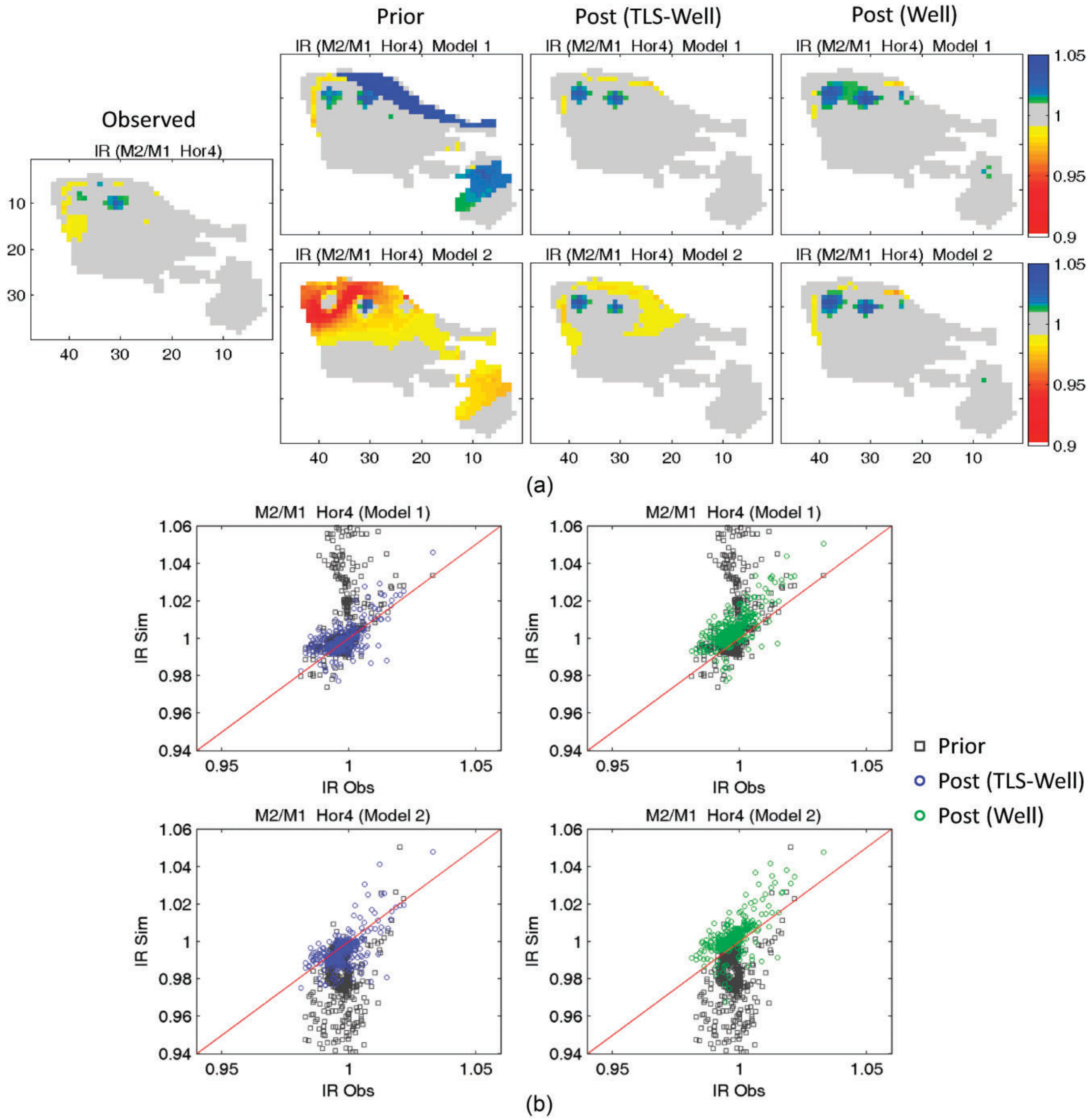

Figure 16. IR maps for the pair of monitors M2/M1 (Horizon 4) for Models 1 and 2 (a) and the corresponding cross-plots (b).

The 4D seismic data used in this work is capable of detecting WAG anomalies, resulting from injected water after a period of injecting gas in the porous media. This kind of anomaly is highlighted in Fig. 13 (M5/M4 Hor4 and M2/M1 Hor6) where it is possible to observe bluish changes surrounded by a hot area in the northwest part of the reservoir. In M5/M4 Hor4, the observed anomalies are correctly represented by the post-ensemble mean for both cases (TLS-Well and Well), whereas for 'M2/M1 Hor6' the anomalies are better represented in the case TLS-Well.

The assessment of the seismic data match is complemented by a quantitative analysis. Figure 14 shows the NQD (|NQDS|) for the 105 pairs of monitors. The NQD (absolute value) was chosen in this case to permit logarithmic scale plots, more convenient than linear scale in this analysis. To compute the NQD, an additive tolerance [constant C_p in Equation (14)] of 0.005 (which represent an impedance variation of 0.5%) was used. In the NQD plots, M1 represents the pair M1/Base, M2 the pair M2/M1 and so on, and H (1 to 15) represents the 15 horizons. To facilitate the analysis, we separated the maps in three groups according to the

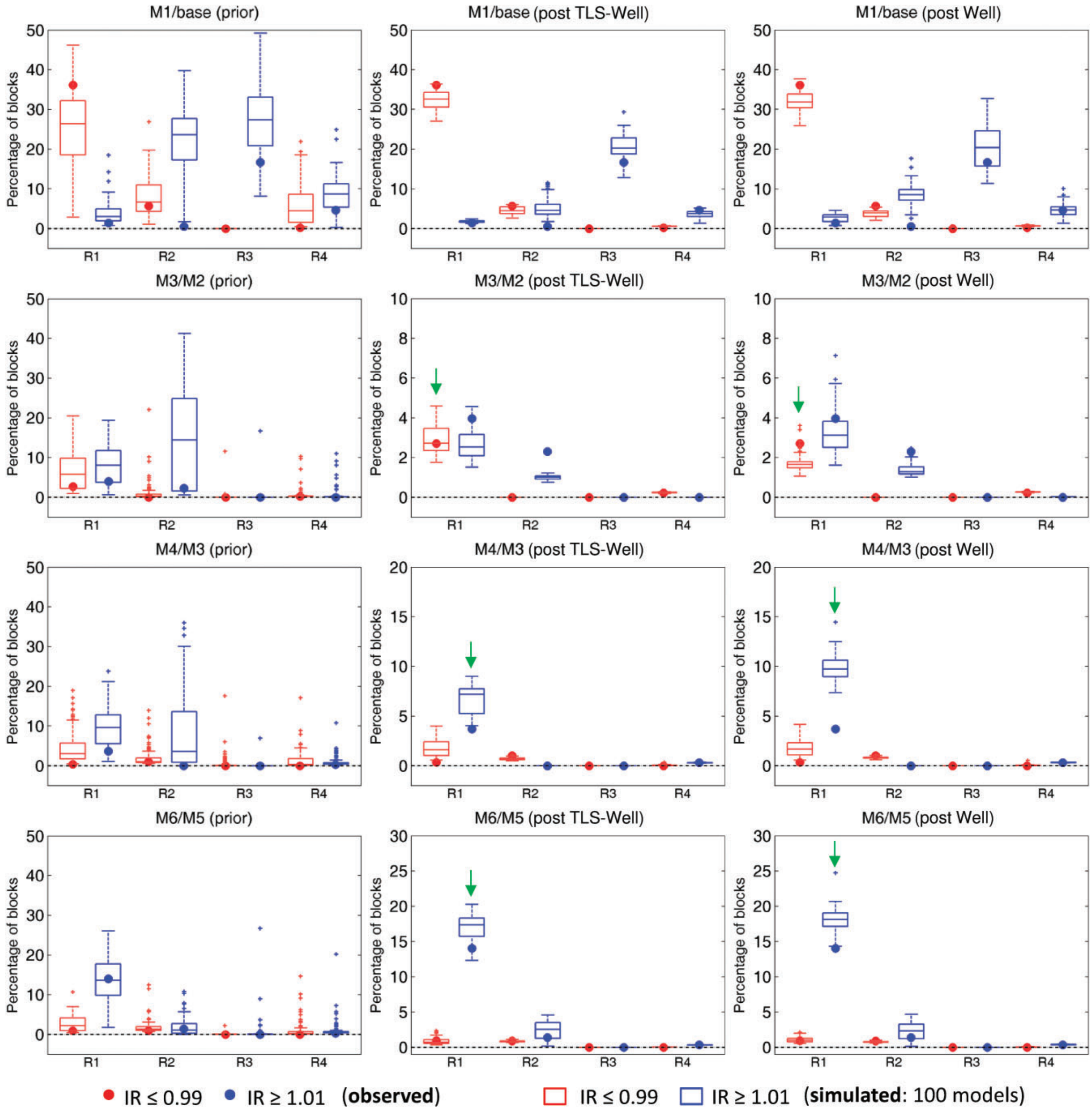

Figure 17. Boxplots of the percentage of blocks with impedance variation (decrease or increase) greater than or equal to 1% for four pairs of monitors for the four regions shown in Fig. 8. The green arrows indicate where the results from post TLS-Well were better than the post Well.

percentage of blocks with impedance variation greater than 1%. In the first group, composed of 22 maps, this percentage is ≥5%. In the second, composed of 33 maps, the percentage is between 1 and 5% and in the third, composed of the remaining 50 maps, the percentage is ≤1%. Note that the order of magnitude of the NQD values decreases from the first to the third group. These plots allow us to assess and compare quantitatively all the 105 pairs of monitors, comparing the pre-and the post-ensemble for both assimilation cases. Clearly, we can see that the case assimilating TLS and wells together provides better results comparing to the case assimilating only well data. For most pairs of monitors, the impedance ratio deviation is lower in the TLS-Well.

To complement the analysis, Fig. 15 shows one pair of monitors representative of each group shown in Fig. 14. These monitors are indicated with a red arrow in Fig. 14. Note that in M1/Base (Hor2) there is a big hardening anomaly in the northeast region of the reservoir in the prior ensemble. In the post-ensemble case TLS-Well this anomaly is much smaller than the prior and very similar to the observed map. In the Well case, although the size of the anomaly was reduced in the matched models, it is still bigger than the

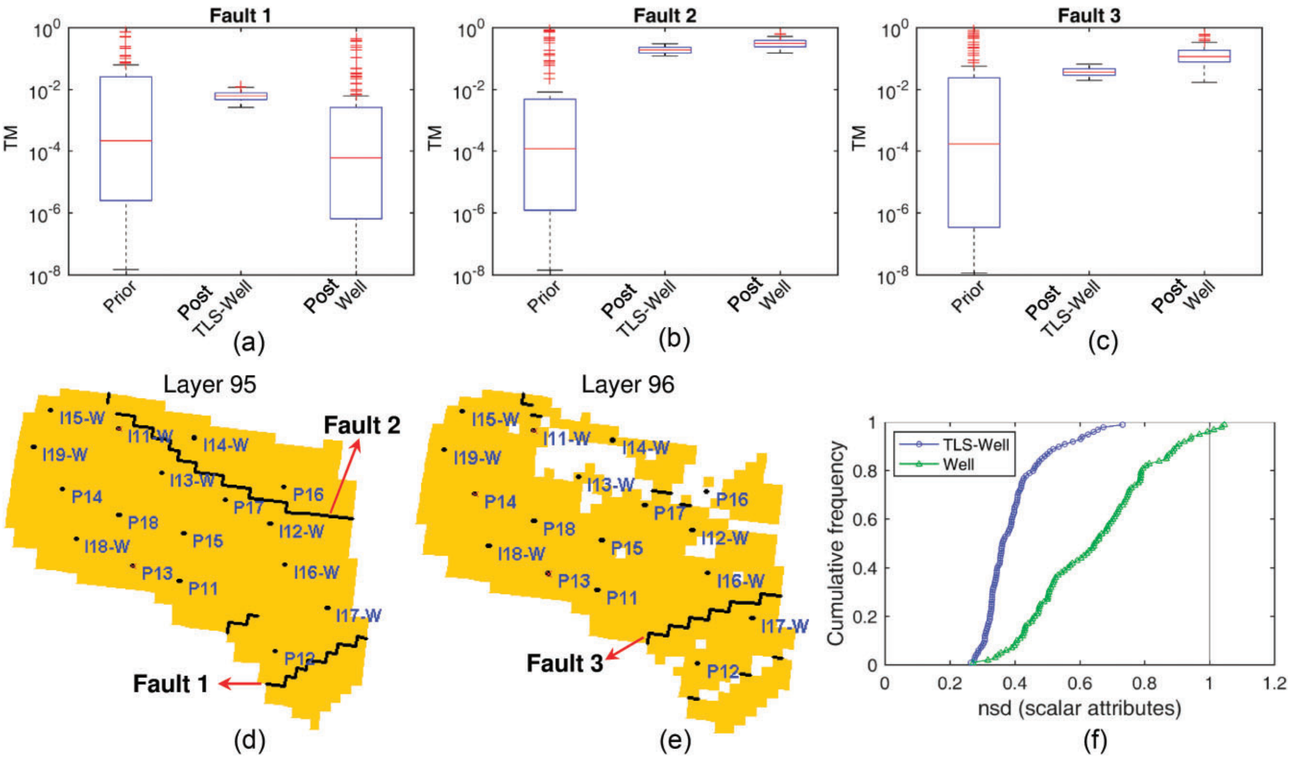

Figure 18. (a–c) The variability of the transmissibility multiplier (TM) of the faults comparing prior, and posterior ensembles for the cases TLS-Well and Well; (d and e) position of the faults; and (f) NSD (post divided by the prior) of all scalar attributes.

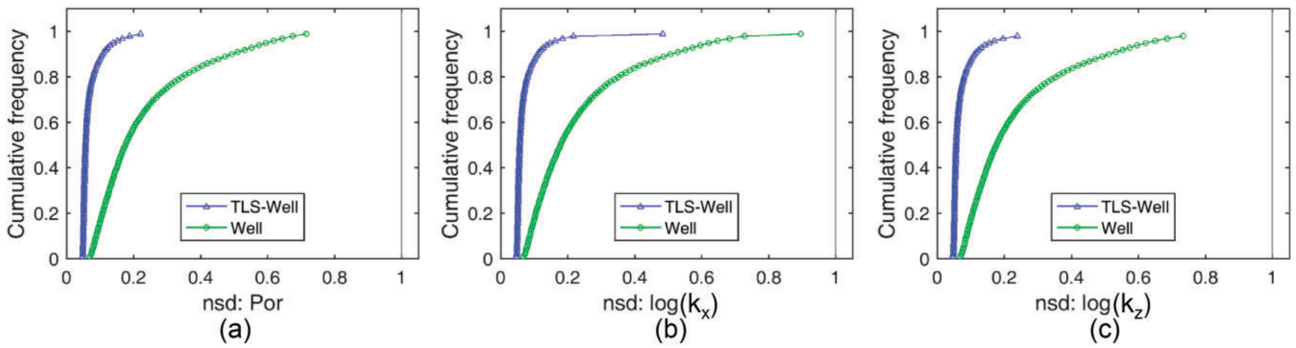

Figure 19. Normalized posterior standard deviation of grid properties with respect to the prior distribution (represented by the vertical black line, NSD = 1): (a) porosity, (b) horizontal permeability in the *x*-direction (k_x) and (c) vertical permeability (k_z); comparison between TLS-Well and Well cases.

observed map. The same can be said for another hardening anomaly in the south region of the reservoir. In the TLS-Well case, this anomaly is also more similar to the observed map compared to the Well case.

Figure 16a shows the IR maps for the pair of monitors M2/M1 (Horizon 4) for Models 1 and 2 from the prior ensemble and from the posterior ensembles for the cases TLS-Well and Well comparing with the observed map. Note that the prior anomalies in Model 1 are totally different compared with Model 2. In Model 1, there is a hardening anomaly in the region above the fault. In this model, the fault transmissibility multiplier is practically zero, characterizing a sealing fault, which causes a hydraulic isolation area in the region above the fault. The limit of the anomaly follows exactly the fault position.

In the post model (after the DA), this problem is corrected and the anomalies are closer to the observed map. The behavior of Model 2 (prior) is completely different because the fault transmissibility multiplier characterizes a fully opened fault. There is a predominance of softening effect with a smaller hardening anomaly in the center. As in Model 1, in the post model the anomalies are closer to the observed map. Figure 16b depicts cross-plots for both models comparing the prior and post for the two assimilation cases,

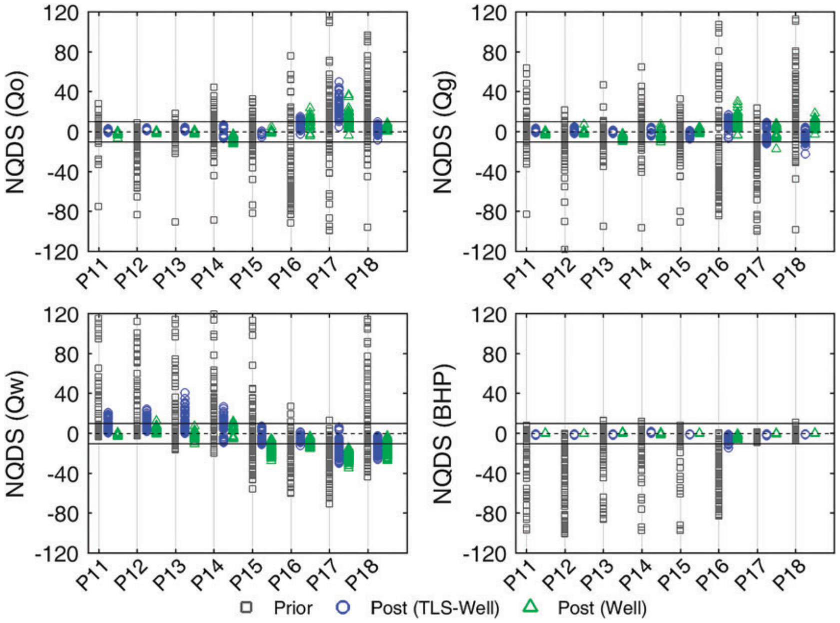

Figure 20. NQDS (computed in the forecast period) for all producers. The closer the NQDS is to the range centered at the zero line, the better the results are.

showing that for the case TLS-Well the posterior points (blue) is more symmetrically distributed around the 45° line when compared to the Well case (green points), indicating that the TLS-Well case results are in better agreement with the observed data.

Figure 17 shows boxplots of the percentage of blocks with impedance variation (decrease or increase) ≥1% for four pairs of monitors for the four regions shown in Fig. 8 (bottom-right image). The first column is the prior, and the second and third columns correspond to the cases TLS-Well and Well, respectively. Red boxes represent softening effects and blue boxes represent hardening effects. The circles represent the percentage corresponding to the observed seismic. It is possible to observe, especially in Region R1, the alternation of impedance variation. For example, in M1/Base (still considering R1) there is a predominance of softening effects, and this behavior gradually changes until the predominance of hardening is reached in M6/M5. We see that the prior ensemble exhibits a high variability. For instance, the percentage of blocks for softening in Region 1 varies from 3 to 46%. On the other hand, the posterior ensembles for both cases significantly reduce the variability, encompassing, in general, the observed value. The green arrows indicate where the results from TLS-Well were better than Well, in the sense that the TLS-Well results tend to better encompass the observed data in the posterior distribution.

4.4. Analysis of the attributes' variability

In Fig. 18a–c, we analyze the three faults variability before (prior) and after the DA (post) comparing the cases TLS-Well and Well. It is interesting to observe that the Faults 2 and 3 converged to similar ranges for both assimilation cases. Another observation is that both cases reduced the variability of these faults in a similar way. We can see from Fig. 18d and e that the Faults 2 and 3 are located in regions of the reservoir close to the wells' locations. This explains why both cases reduced the uncertainty in the transmissibility multipliers of these two faults similarly: in the Well case, there was enough well data to constrain the faults transmissibility, in other words, the fault effects are captured by the wells. For the TLS-Well case this result was expected because, besides the well data, there was the spatial information from the TLS. For the Fault 1, we can see that the results are different from Faults 2 and 3. For the TLS-Well case, the uncertainty reduction was similar to Faults 2 and 3, however, the Well case practically maintained the prior distribution. This can be explained by the fact that the Fault 1 is located in a region more distant from the wells. Therefore, the wells' data were not enough to constrain this fault transmissibility and the method (SEnRML-LA) correctly maintained the uncertainty for this attribute. On the other hand, the spatial information from the TLS enabled the uncertainty reduction of this fault (TLS-Well case).

Figure 18f shows the cumulative frequency curves for the normalized standard deviation (NSD), of the posterior ensembles for all scalar attributes. The NSD is the posterior standard deviation normalized with respect to the prior standard deviation. These curves indicate that the DA process using the TLS data have a more significant uncertainty reduction. For example, for the Well case, the ratio between the post and the prior (that measures the uncertainty

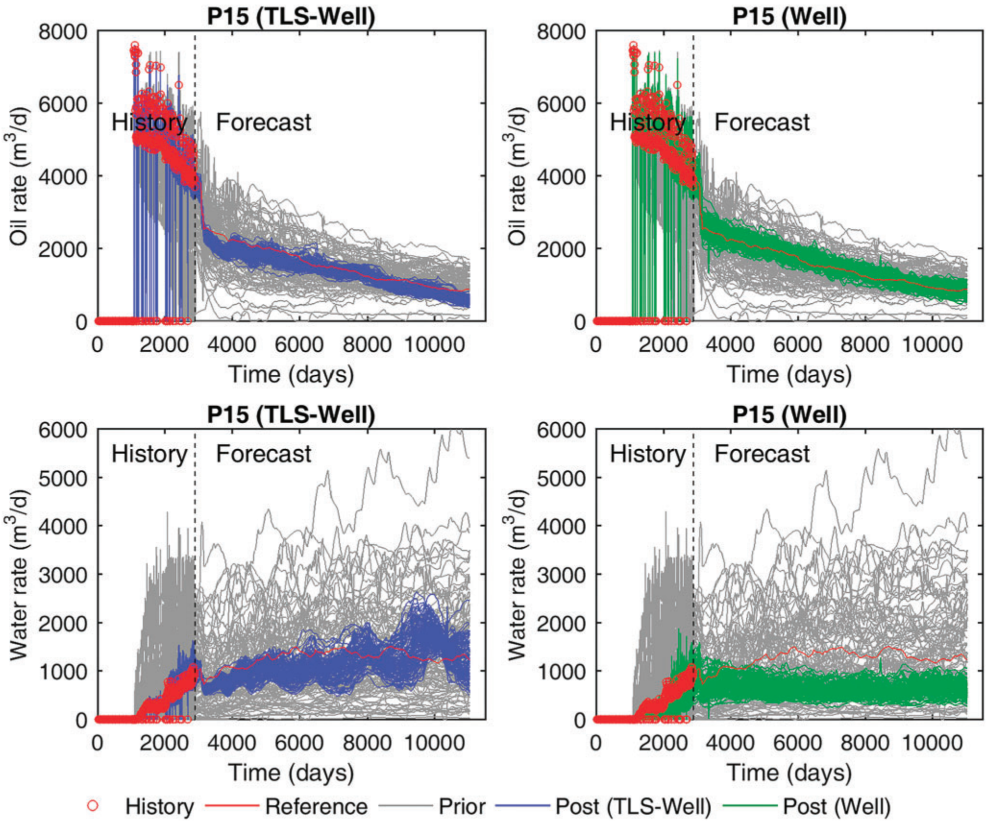

Figure 21. Oil and water production forecast for the Well P15 comparing the prior and posterior ensembles for cases TLS-Well and Well.

reduction) for 50% of the attributes (P50) is lower than or equal to 0.64; while for the TLS-Well case, the P50 is 0.36, indicating a higher uncertainty reduction.

In Fig. 19, we analyze the posterior variability of the grid properties (porosity and permeability in *x* and z directions) normalized with respect to the prior distribution (represented by the vertical black line, NSD = 1). The *y*-permeability is not shown because its behavior is very similar to the *x*-permeability. First, the TLS data allows a more effective uncertainty reduction compared to the case assimilating only well data. Second, comparing with the scalar attributes, the assimilation process using the TLS leads to a more significant uncertainty reduction, which is normally expected because both the grid properties and the TLS data are spatially distributed or, in other words, the grid properties tended to be more strongly conditioned by the TLS than the scalar attributes. However, it is worth mentioning that this explanation is valid for the case studied and we could not necessarily generalize for all cases. Another aspect is that, in general, the higher the amount of data to be assimilated, the higher the uncertainty reduction.

4.5. *Production forecast*

To assess the forecast quality of the models (see Section 3.6) in a well-by-well basis, we first computed the NQDS in the forecast period (using the same tolerances applied to the history period). Note that this is possible because we know the reference solution. In real cases, this may be accomplished by dividing part of the history data to assimilate and part to compare with the models in the extrapolated (forecast) period (an example of this kind of analysis can be seen in Maschio *et al.* 2022). The NQDS for the producer wells (forecast period) is shown in Fig. 20. Overall, these plots show significant uncertainty reduction in the forecast period for both cases, especially for oil and gas rates and BHP, compared with the prior ensemble. Nevertheless, incorporating TLS data does not seem to improve the water rate forecast for some wells, especially wells P11, P12, and P13. This can be explained by the fact that these wells do not produce water during the history period, not contributing, as a consequence, to water-related anomalies in the TLS maps. On the other hand, for the wells P15 and P16 (which produce more water during the history) the water rate forecast is better, as shown in the NQDS distribution for these wells. Another complimentary explanation (related to the previous one) is the fact that these wells (specially P11 and P13) are located close to the hardening-pressure anomaly in the center-south region (M1/Base, Fig. 5), in other words, the absence of water anomaly in the vicinity of this region made it difficult to forecast water production in these wells.

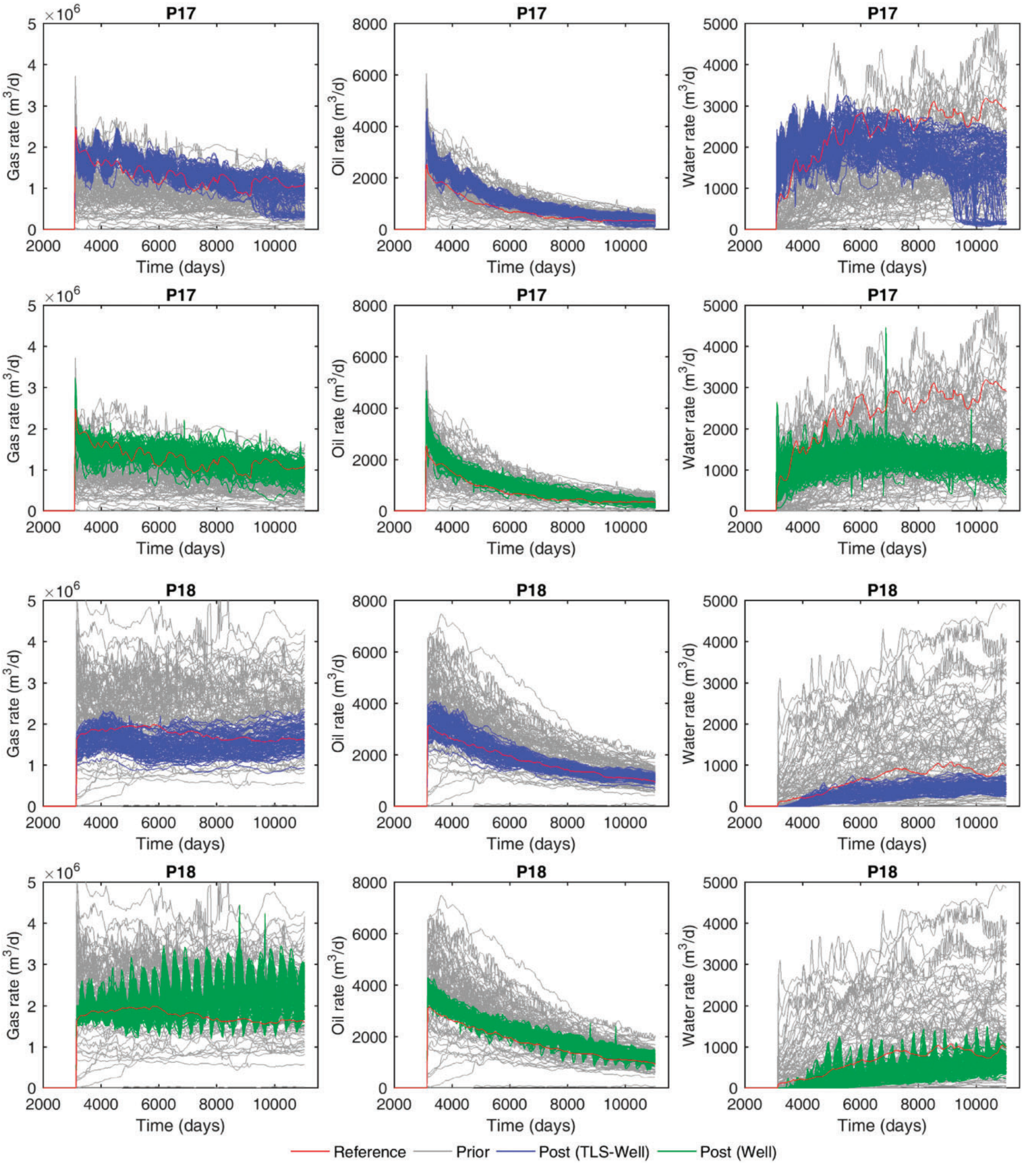

Figure 22. Gas, oil, and water rates for the wells P17 and P18 (opened after the end of the history period) comparing the prior and the posterior ensembles for cases TLS-Well and Well with the reference.

In Fig. 21 we present the oil and water rate for the Well P15. We can see that the better water rate matching in the case TLS-Well (as shown in Figs 10 and 11) reflects in a better water rate forecast when the seismic data is assimilated. Note (from Fig. 11) that P15 water rate is significant. Regarding the oil production, the sudden oil rate drop in the beginning of the forecast period is due to the opening of the producers P17 and P18. To honor the group production, the rate of some wells is reduced automatically by the simulator. But it is important to observe that the models follow adequately the reference behavior in both cases (TLS-Well and Well) in the forecast period.

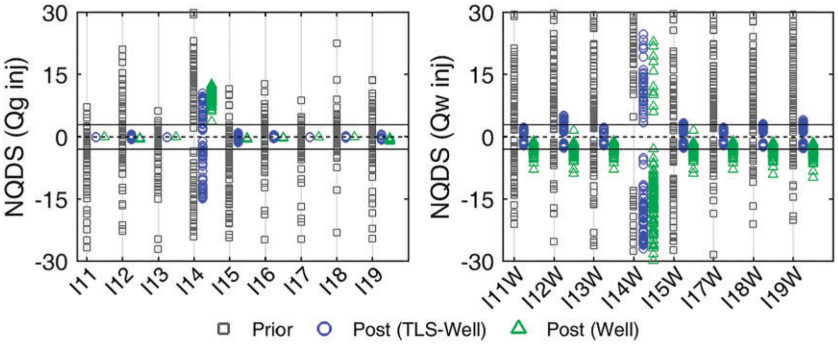

Figure 23. NQDS (computed in the forecast period) for all injectors. The closer the NQDS is to the range centered at the zero line, the better the results are.

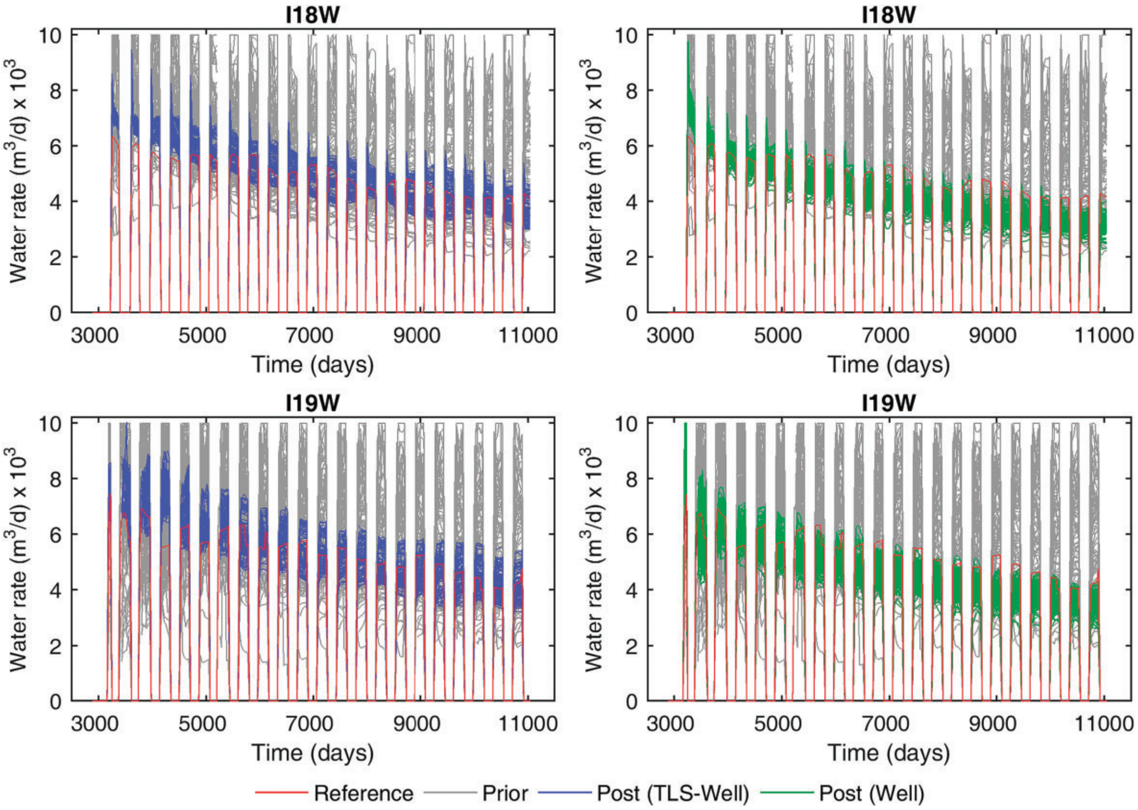

Figure 24. Water rate for the injectors I18W and I19W (opened after the end of the history period) comparing the case TLS-Well and Well.

To complement the analyses of the producer wells, Fig. 22 shows gas, oil, and water rates for the wells P17 and P18 comparing the prior and the posterior ensembles for cases TLS-Well and Well with the reference. These wells open after the end of the history period. Therefore, they are examples of how the models can predict a new well performance (without history). For P17, the distribution of gas and oil rate curves (posterior ensemble) is similar for TLS-Well and Well cases; however, the water rate for the TLS-Well is closer to the reference when compared to the Well case. P18 water rate is a little bit underestimated for both cases, but this well does not produce much water. P18 gas and oil production confirm the TLS benefits for this well since the reference solution is better represented in the TLS-Well posterior curves compared to the Well case. Note that the TLS-Well posterior curves are more symmetrically distributed around the reference. Another observation is that the oscillations in the P18 posterior curves for the case Well do not appear in the TLS-Well case, in which the behavior of the models is very similar to the reference solution.

The wells P17 and P18 are in a central portion of the reservoir and the Well case results for P17 and P18 reflect the lack of spatial information to condition the reservoir property perturbation far from the wells that operate during the history. In this sense, the TLS data provides complementary information to the well data, evidencing the benefit of the seismic data.

Besides the production wells, it is also important to analyze the injector wells. Figure 23 shows the NQDS (computed in the forecast period) for all injectors. Note that the injected water forecast for all injectors is better for the TLS-Well case, since the NQDS distribution is closer to the range around zero (and more symmetrical) compared to the Well case.

These results indicate that the WAG injection forecasted by the models conditioned to the well and seismic data is more balanced than the WAG injection forecasted by the models conditioned only to the well data, evidencing the benefits of the TLS data. Figure 24 shows the water rate for the injectors I18W and I19W comparing the cases TLS-Well and Well. As the producer P17 and P18, the injectors I18 and I19 also open after the end of the history period. The explanation for these results is similar to the producers, that is, the TLS data provides additional (spatial) information for a more appropriate reservoir properties updating far from the wells operating during history period.

The case TLS-Well took 135 hours (elapsed-time) to run and Well took 116 hours. The case TLS-Well took more time than Well because it is necessary to extract much more information from the reservoir output to compute the impedance ratio. For both cases, the reservoir simulations were distributed (and parallelized) in a cluster of Linux computers. Each simulation used eight processors (in parallel) in a computational environment where 200 processors are available for each user. It is important to note that this time depends on several factors such as the characteristics of the reservoir model, computational resources available and number of licenses for the reservoir simulator. The SEnRML-LA was run in a shared Windows machine Intel(R) Xeon(TM) CPU @ 2.20 GHz and 64 GB memory available.

5. Conclusions

In this work, we performed well and TLS DA in a realistic case that represents challenges similar to a Brazilian pre-salt reservoir. The case studied corresponds to a carbonate reservoir with a light oil and high CO_2 content. All produced gas is reinjected back to reservoir in a WAG strategy for enhanced oil recovery (and CO_2 discard). A compositional fluid model was considered for the reservoir simulations and a recent proposed approach (Silva Neto *et al.* 2020) was applied to estimate oil and gas properties in the petro-elastic model. The case also considers ICV in the wells, and permanent seismic monitoring. We employed the SEnRML method with local analysis to assimilate seven pairs of seismic monitors together with the well data. The specific conclusions are:

(i) The method SEnRML with local analysis (SEnRML-LA), proposed by Silva Neto *et al.* (2021), can handle big datasets originated from multiple monitor acquisitions, being an alternative to solve practical problems involving permanent seismic monitoring technologies.

(ii) The problem solved in this work with the SEnRML-LA is the state of the art in DA process. It was possible to assimilate all the data simultaneously, including the 105 horizons for the TLS and the wells' production and pressure data.

(iii) The DA was successful in terms of the quality of the results and method performance. Good convergence was verified for both objective functions (seismic and wells) for the TLS-Well case. For the Well case, good convergence was also verified for the well objective functions.

(iv) The assimilation of TLS data together with well data provided better well matching compared with the case assimilating only well data. The TLS data provide information that is coherent and partially redundant to the well data. Therefore, it allows a better well data match with Bayesian methods by shifting the optimal solutions toward a smaller data cost function.

(v) The case assimilating only well data allowed a reasonable seismic data match, indicating the well data reveal part of the information associated with the TLS data. However, a detailed analysis showed that TLS-Well case provided better seismic data match.

(vi) Our data indicates that it is possible to detect WAG effects through TLS data in conditions close to ours, although the combination of hardening and softening effects may attenuate the anomalies in late monitor pairs.

(vii) Seismic simulation using a compositional fluid model provided useful information for reservoir parameters calibration in a miscible gas injection alternating water synthetic case.

(viii) The assimilation of TLS data jointly with well data improved the reservoir forecast compared to the prior ensemble and the one calibrated only with production data. This result corroborates the importance of TLS data in a pre-salt reservoir application. Moreover, it demonstrates that integrating the compositional fluid model and SEnRML with local analysis is a viable solution to take advantage of the TLS information.

(ix) The assimilation of TLS data improved the reservoir characterization in regions far from the wells opened during the history period, improving the TLS data match and the production/injection forecast of new wells. It was shown that the two new producers and the two new injectors (opened after the history period) provided better forecast results for the TLS-Well case.

Finally, this work showed the benefit the TLS data gathered from multiple monitors using permanent monitoring technologies and proved the robustness of the SEnRML-LA method in assimilating all seismic monitors simultaneously.

Acknowledgments

We conducted this work with the support of Libra Consortium (Petrobras, Shell Brasil, TotalEnergies, CNOOC, CNPC, and PPSA) within the ANP R&D levy as 'commitment to research and development investments' (grant no. 5850.0105307.17.9) and Energi Simulation. The authors are grateful for the support of the Center of Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil), and the Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil). Also, a special thanks to CMG and Schlumberger for software licenses.

Conflict of interest. None declared.

Data availability

The data will be available on request.

References

Abelha M, Petersohn E. The state of the art of the Brazilian pre-salt exploration. in *AAPG 2018 Annual Convention & Exhibition*. Salt Lake City, Utah, US: American Association of Petroleum Geologists (AAPG), 2018.

Altundas B, Chugunov N, Ramakrishnan TS *et al.* Quantifying the effect of CO_2 dissolution on seismic monitoring of CO_2 in CO_2-EOR. *Soc Explor Geophys* 2017; pp. 3771–5. https://doi.org/10.1190/segam2017-17792996.1

Avansi GD, Maschio C, Schiozer DJ. Simultaneous history-matching approach by use of reservoir-characterization and reservoir-simulation studies (SPE-179740-PA). *SPE Reservoir Eval Eng* 2016;**19**:694–712. https://doi.org/10.2118/179740-PA

Avseth P, Mukerji T, Mavko G. *Quantitative Seismic Interpretation: Applying Rock Physics Tools to Reduce Interpretation Risk*. Cambridge, UK; New York: Cambridge University Press, 2005.

Batzle M, Wang Z. Seismic properties of pore fluids. *Geophysics* 1992;**57**:1396–408. https://doi.org/10.1190/1.1443207

Botechia VE, Araújo de Lemos R, Hohendorff Filho JC *et al.* Well and ICV management in a carbonate reservoir with high gas content. *J Pet Sci Eng* 2021;**200**:108345. https://doi.org/10.1016/j.petrol.2021.108345

Chen Y, Oliver DS. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math Geosci* 2012;**44**:1–26. https://doi.org/10.1007/s11004-011-9376-z

Chen Y, Oliver DS. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput Geosci* 2013;**17**:689–703. https://doi.org/10.1007/s10596-013-9351-5

Chen Y, Oliver DS. Localization and regularization for iterative ensemble smoothers. *Comput Geosci* 2017;**21**:13–30. https://doi.org/10.1007/s10596-016-9599-7

CMG, *WinProp Fluid Property Characterization Tool, Version 2015.10*. Calgary, Alberta, Canada: Computer Modelling Group Ltd, 2015. https://www.cmgl.ca/winprop.

CMG, *GEM Compositional & Unconventional Simulator, Version 2020.10* Computer Modelling Group Ltd, 2020. https://www.cmgl.ca/gem.

Correia M, Botechia V, Pires L *et al.* UNISIM-III: benchmark case proposal based on a fractured Karst reservoir. in *ECMOR XVII –17th European Conference on the Mathematics of Oil Recovery*, 14-17 September, Online Event. 2020.

Costa MM, Silva EP, Santos MS *et al.* Quantification of 4D seismic response in pre-salt carbonates via time shift analysis. in *3rd EAGE/SBGf Workshop 2016: Quantitative Seismic Interpretation of Lacustrine Carbonates* European Association of Geoscientists and Engineers, EAGE, 2016; pp. 85–9.

Cruz ROM, Rosa MB, Branco CCM *et al.* Lula NE pilot project - an ultra-deep success in the Brazilian pre-salt. in *Offshore Technology Conference*, OTC, Houston, TX, USA, 2016; p. D011S007R001. https://doi.org/10.4043/27297-MS

Davolio A, Santos JMC, Rosa DR *et al.* 4D seismic detectability on a pre-salt like reservoir. in *Second EAGE Conference on Pre-Salt Reservoir: from Exploration to Production* 8–10 September 2021, Online. 2021.

Deplante C, Costa M, Santos MS *et al.* Using full wave seismic modeling to test 4D repeatability for Libra pre-salt field. in *16th International Congress of the Brazilian Geophysical Society*, Sociedade Brasileira de Geofísica, Rio de Janeiro. 2019.

Emerick A, Reynolds A. Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Comput Geosci* 2011;**15**:251–69. https://doi.org/10.1007/s10596-010-9198-y

Emerick AA. Analysis of the performance of ensemble-based assimilation of production and seismic data. *J Pet Sci Eng* 2016;**139**:219–39. https://doi.org/10.1016/j.petrol.2016.01.029.

Emerick AA, Reynolds AC. History-matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation. in *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2013a. https://doi.org/10.2118/163675-MS

Emerick AA, Reynolds AC. Ensemble smoother with multiple data assimilation. *Comput Geosci* 2013b;**55**:3–15. https://doi.org/10.1016/j.cageo.2012.03.011

Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res* 1994;**99**:10143. https://doi.org/10.1029/94JC00572

Evensen G. *Data Assimilation*. Berlin, Heidelberg: Springer, 2009. https://doi.org/10.1007/978-3-642-03711-5

Evensen G. Accounting for model errors in iterative ensemble smoothers. *Comput Geosci* 2019;**23**:761–75. https://doi.org/10.1007/s10596-019-9819-z

Evensen G, Raanes PN, Stordal AS *et al.* Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. *Front Appl Math Stats* 2019;**5**:47. https://doi.org/10.3389/fams.2019.00047

Fahimuddin A, Aanonsen SI, Skjervheim J-A. 4D Seismic history matching of a real field case with EnKF: use of local analysis for model updating. in *SPE Annual Technical Conference and Exhibition*. Florence, Italy: Society of Petroleum Engineers, 2010. https://doi.org/10.2118/134894-MS

Gaspari G, Cohn SE. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteoro. Soc.* 1999;**125**:723–57. https://doi.org/10.1002/qj.49712555417

Gassmann F. Über die elastizität poröser medien: vierteljahrss-chrift der Naturforschenden Gesellschaft in Zurich. 1951;**96**:1–23. The English translation of this paper is available at http://sepwww.stanford.edu/sep/berryman/PS/gassmann.pdf

Gu Y, Oliver DS. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal* 2007;**12**:438–46. https://doi.org/10.2118/108438-PA

Han D, Sun M, Liu J. Velocity and density of CO$_2$-oil miscible mixtures. *Soc Explor Geophys* 2012; pp. 1–5. https://doi.org/10.1190/segam2012-1242.1

Han D, Sun M, Liu J. Velocity and density of oil-HC-CO$_2$ miscible mixtures. *Soc Explor Geophys* 2013; pp. 2831–5. https://doi.org/10.1190/segam2013-1104.1

Johann PRS, Monteiro RC. Geophysical reservoir characterization and monitoring at Brazilian pre-salt oil fields. in *Offshore Technology Conference, OTC*, Houston, TX, USA, 2016; p. D021S020R003. https://doi.org/10.4043/27246-MS

Luo X, Bhakta T. Automatic and adaptive localization for ensemble-based history matching. *J Pet Sci Eng* 2020;**184**:106559. https://doi.org/10.1016/j.petrol.2019.106559

Maschio C, Avansi GD, Silva FBM *et al.* Data assimilation for uncertainty reduction using different fidelity numerical models. *J Pet Sci Eng* 2022;**209**:109851. https://doi.org/10.1016/j.petrol.2021.109851

Maschio C, Schiozer DJ. Probabilistic history matching using discrete Latin Hypercube sampling and nonparametric density estimation. *J Pet Sci Eng* 2016;**147**:98–115. https://doi.org/10.1016/j.petrol.2016.05.011

Mavko G, Mukerji T, Dvorkin J. *The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media*, 2nd edn. Cambridge, UK; New York: Cambridge University Press, 2009.

Moczydlower B, Salomão MC, Branco CC *et al.* Development of the Brazilian pre-salt fields – when to pay for information and when to pay for flexibility. in *SPE Latin America and Caribbean Petroleum Engineering Conference*, SPE, Mexico City, Mexico, 2012; p. SPE-152860-MS. https://doi.org/10.2118/152860-MS

Pedersen KS, Christensen PL, Shaikh JA. *Phase Behavior of Petroleum Reservoir Fluids*, 2nd edn. Boca Raton: CRC Press, Taylor & Francis Group, 2015.

Péneloux A, Rauzy E, Fréze R. A consistent correction for Redlich-Kwong-Soave volumes. *Fluid Phase Equilib* 1982;**8**:7–23. https://doi.org/10.1016/0378-3812(82)80002-2

Peng D-Y, Robinson DB. A new two-constant equation of state. *Ind Eng Chem* 1976;**15**:59–64. https://doi.org/10.1021/i160057a011

Petrobras, Teste de Longa Duração e Sistemas de Produção Antecipada de Libra – Bacia de Santos – II. Caracterização da Atividade. In *Portuguese* Available in: http://licenciamento.ibama.gov.br/Petroleo/Producao/Producao%20-%20Bacia%20de%20Santos%20-%20TLD%20e%20SPAs%20de%20Libra%20-%20Petrobras/EIA/II_2_CaracAtividade/II_2-CaracAtividade.pdf (last accessed on September 8, 2022). 2015.

Pizarro JODS, Branco CCM. Challenges in implementing an EOR project in the pre-salt province in deep offshore Brasil. in *SPE EOR Conference at Oil and Gas West Asia*, Muscat, Oman: Society of Petroleum Engineers. 2012; https://doi.org/10.2118/155665-MS

Raanes PN, Stordal AS, Evensen G. Revising the stochastic iterative ensemble smoother. *Nonlinear Processes Geophys* 2019;**26**:325–38. https://doi.org/10.5194/npg-26-325-2019

Robinson DB, Peng D-Y. The characterization of the heptanes and heavier fractions for the GPA Peng-Robinson programs (Research Report RR-28). Edmonton, Alberta: Gas Processors Association. 1978.

Rosa DR, Schiozer DJ, Davolio A. Data assimilation of production and multiple 4D seismic acquisitions in a deepwater field using ensemble smoother with multiple data assimilation. *SPE Reservoir Eval Eng* 2023;**26**:1528–40 (SPE-215812-PA).

Silva EPA, Davólio A, Santos MS *et al.* 4D petro-elastic modeling based on a pre-salt well. *Interpretation* 2020;**8**:639–49. https://doi.org/10.1190/INT-2019-0099.1

Silva Neto GM, Rios VdS, Davolio A *et al.* Improving fluid modeling representation for seismic data assimilation in compositional reservoir simulation. *J Pet Sci Eng* 2020;**194**:107446. https://doi.org/10.1016/j.petrol.2020.107446

Silva Neto GM, Soares RV, Evensen G *et al.* Subspace ensemble randomized maximum likelihood with local analysis for time-lapse-seismic-data assimilation. *SPE J* 2021;**26**:1011–31. https://doi.org/10.2118/205029-PA

Skjervheim J-A, Evensen G, Aanonsen SI *et al.* Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter. *SPE J* 2007;**12**:282–92. https://doi.org/10.2118/95789-PA

Soares RV, Maschio C, Schiozer DJ. Applying a localization technique to Kalman Gain and assessing the influence on the variability of models in history matching. *J Pet Sci Eng* 2018;**169**:110–25. https://doi.org/10.1016/j.petrol.2018.05.059

Tahani H. *Determination of the velocity of sound in reservoir fluids using an equation of state. Doctoral Thesis*, Heriot-Watt University, Edinburgh, United Kingdom, 2012.

Vasquez GF, Morschbacher MJ, Justen JCR. Experimental efforts to access 4D feasibility and interpretation issues of Brazilian presalt carbonate reservoirs. *Interpretation* 2019;**7**:SH1–SH18. https://doi.org/10.1190/INT-2018-0203.1