Versão do arquivo anexado / Version of attached file:

Versão do Editor / Published Version

Mais informações no site da editora / Further information on publisher's website:

https://link.springer.com/article/10.1007/s13202-019-0727-5

DOI: 10.1007/s13202-019-0727-5

Direitos autorais / Publisher's copyright statement:

ORIGINAL PAPER - PRODUCTION ENGINEERING

A novel localization scheme for scalar uncertainties in ensemble-based data assimilation methods

Ricardo Vasconcellos Soares[1] · Célio Maschio[1] · Denis José Schiozer[1]

Abstract

History matching, also known as data assimilation, is an inverse problem with multiple solutions responsible for generating more reliable models for use in decision-making processes. An iterative ensemble-based method (Ensemble Smoother with Multiple Data Assimilation—ES-MDA) has been used to improve the solution of history-matching processes with a technique called distance-dependent localization. In conjunction, ES-MDA and localization can obtain consistent petrophysical images (permeability and porosity). However, the distance-dependent localization technique is not used to update scalar uncertainties, such as relative permeability; therefore, the variability for these properties is excessively reduced, potentially excluding plausible answers. This work presents three approaches to update scalar parameters while increasing the final variability of these uncertainties to better scan the search space. The three approaches that were developed and compared using a benchmark case are: binary correlation coefficient (BCC), based on correlation calculated by ES-MDA through cross-covariance matrix C_{MD}^f (BCC-C_{MD}); BCC, based on a correlation coefficient between the objective functions and scalar uncertainties (R) (BCC–R); and full correlation coefficient (FCC). We used the work of Soares et al. (J Pet Sci Eng 169:110–125, 2018) as a base case to compare the approaches because although it showed good matches with geologically consistent petrophysical images, it generated an excessive reduction in the scalar parameters. BCC-C_{MD} presented similar results to the base case, excessively reducing the variability of the scalar uncertainties. BCC–R increased the variability in the scalar parameters, especially for BCC with a higher threshold value. Finally, FCC found many more potential answers in the search space without impairing data matches and production forecast quality.

Keywords History matching · ES-MDA · Distance-dependent localization · Non-distance-dependent localization · Correlation-based adaptive localization

List of symbols

α	Inflation factor
λ	Threshold value
ρ	Correlation matrix
σ	Standard deviation
σ^2	Variance
a	Representation of each uncertainty
b	Representation of each observed data point
BCC	Binary correlation coefficient
BHP	Bottomhole pressure
cc	Correlation coefficient (as calculated by Furrer and Bengtsson)
c	Covariance
C_D	Measurement error
C_{DD}^f	Auto-covariance of simulated data
C_{MD}^f	Cross-covariance matrix between uncertainties and simulated data
Const	Constant in NQDS formulation
CPOR	Rock compressibility
d_{obs}	Observed data vector
d_{sim}	Simulated data vector
DWOC	Depth of water–oil contact
EnKF	Ensemble Kalman Filter
ES-MDA	Ensemble smoother with multiple data assimilation
FCC	Full correlation coefficient
h	Euclidean distance
i	Iteration number
j	Model number
K	Kalman gain matrix
k_{rw}	Water relative permeability

✉ Ricardo Vasconcellos Soares
    rsoares@cepetro.unicamp.br

[1] Department of Energy, School of Mechanical Engineering, University of Campinas - Unicamp, Campinas, SP, Brazil

k_{rwiro}	Maximum water relative permeability
k_x	Horizontal permeability in the x-direction
k_y	Horizontal permeability in the y-direction
k_z	Vertical permeability
L	Critical length
m^a	Updated uncertainty vector
m^f	Initial uncertainty vector
MCMC	Markov Chain Monte Carlo
MOD	Vertical permeability multiplier
NDDL	Non-distance-dependent localization
NQDS	Normalized quadratic deviation with sign
Ne	Number of models
Ni	Number of iterations
Nd	Number of observed data
OF	Objective function
pdf	Probability density function
P_{krw}	Water relative permeability exponent
q_l	Liquid rate
q_o	Oil rate
q_w	Produced water rate
q_{wi}	Injected water rate
R	Correlation coefficient
S_{orw}	Residual oil saturation
S_w	Water saturation
S_{wcrit}	Critical water saturation
Tol	Tolerance of observed data
TSV	Truncated singular value

Introduction

History matching is a crucial process during petroleum field management. Historical production and injection data are used to update uncertainties in the reservoir simulation models and generate more consistent models. Reservoir behavior must be predicted with a certain level of confidence as it is used to optimize key factors of oil and gas projects, such as oil production and/or net present value. History matching is a challenging inverse problem, in which the ultimate goal is to determine the uncertain parameters that lead to the known answer (historical production and injection data, for instance).

According to Oliver and Chen (2011), history matching is an ill-posed process, i.e., multiple combinations of the uncertain parameters can match historical data. Therefore, as the exact characteristics of the real reservoir are unknown, a single model is insufficient to represent the reservoir. As several combinations of uncertainties can match historical data, all possible combinations within a defined search space should be considered, taking into account the main geological and operational features.

Several methods, such as Markov Chain Monte Carlo (MCMC) and gradient-based methods are proven to be capable of matching simulated to observed data. However, these either are computationally demanding or concentrate the response in specific regions (local minimum), failing to find some of the possible models that can represent the reservoir in a proper manner (Yustres et al. 2011; Oliver and Chen 2011).

Ensemble-based methods derived from Ensemble Kalman Filter (EnKF) are a good alternative for history matching. EnKF is a sequential data assimilation algorithm proposed by Evensen (1994) that represents the uncertainties in the models and in the state vector through a finite number of models by a covariance matrix. Furthermore, the ability of storing and propagating this matrix made possible to work with models containing a great amount of data with low computational effort. Usually, EnKF and other ensemble-based methods are used in conjunction with localization techniques. This type of technique was first used in data assimilation problems by Houtekamer and Mitchell (1998), where they utilized distance-dependent localization to constrain the updates of the uncertainties into a defined region based on the observed data. Its use allowed to generate more consistent models, avoiding spurious correlations and underestimation of uncertainties (ensemble collapse), as also mentioned by Aanonsen et al. (2009), Arroyo-Negrete et al. (2008) and Emerick and Reynolds (2011). Note that there are other types of localization techniques, such as non-distance-dependent localization (Furrer and Bengtsson 2007) and streamline-based localization (Arroyo-Negrete et al. 2008).

Alternatively, Emerick and Reynolds (2013) developed the Ensemble Smoother with Multiple Data Assimilation (ES-MDA) which is an iterative ensemble-based method similar to EnKF but it assimilates all historical data in one single update multiple times. Emerick (2016) and Soares et al. (2018) used ES-MDA with distance-dependent localization, and pointed out that the distance-dependent localization technique is used to update only petrophysical uncertainties, which has single values for each grid cell, such as porosity and permeability images. Scalar uncertainties are represented by a single value for the whole model or part of it (rock compressibility, water relative permeability, etc.) and are not updated with the distance-dependent localization technique. Consequently, the variability is strongly reduced, suggesting the potential exclusion of models that can represent the reservoir.

Furthermore, previous knowledge about the ES-MDA methodology demonstrated that the method achieved different values for scalar uncertainties when changing some key parameters, such as inflation factor. Therefore, the method neglected some models with a good representation of the real reservoir. Clearly, the search space of uncertain scalar parameters requires better scanning to improve the representativeness of the reservoir. Consequently, more different scenarios must

be generated to reliably forecast production for future field management decisions.

Objective

The objective of this work is to avoid excessive reduction in variability of the models after history-matching process and the exclusion of models that can represent the reservoir. We achieve this through the development and application of new procedures to update scalar uncertainties using ES-MDA and covariance localization. Thus, the combination of distance-dependent localization to update petrophysical uncertainties and the new procedures to update scalar parameters intends to perform a better scan of the defined search space, avoiding exclusion of potential plausible answers. To test the new procedures, we used the benchmark case, UNISIM-I-H, developed by Avansi and Schiozer (2015), with characteristics and complexity of a real case.

ES-MDA and the localization technique

Emerick and Reynolds (2013) proposed ES-MDA, an iterative method based on the Ensemble Kalman Filter (EnKF). In data assimilation with ensemble-based methods, it is necessary first to run the simulation models, after that, we need to match simulated and historical data and update uncertain parameters. As an iterative method, ES-MDA performs this multiple times, and Emerick and Reynolds proposed the following equation, which is known as the analysis equation:

$$m_j^a = m_j^f + K\left(d_{obs,j} - d_{sim,j}\right) \tag{1}$$

where m^a is updated uncertainty vector; m^f, the initial or previous uncertainty vector; K, the Kalman gain matrix; d_{sim}, the simulated data based on the previous simulation models; d_{obs}, the observed data; and j, the model number. Additionally, the Kalman gain matrix can be defined as:

$$K = C_{MD}^f\left(C_{DD}^f + \alpha_i C_D\right)^{-1} \tag{2}$$

where C_{MD}^f is the cross-covariance matrix between uncertainties of all models and simulated data (d_{sim}), C_{DD}^f is the auto-covariance of d_{sim}, C_D is the measurement error

associated with observed data, α is the inflation factor, and i, the iteration number. α_i represents a weight assigned to each iteration and, C_{MD}^f, C_{DD}^f and C_D are defined in Eqs. 3, 4 and 5, respectively.

$$C_{MD}^f = \frac{1}{Ne - 1} \sum_{j=1}^{Ne} \left(m_j^f - \bar{m}_f\right)\left(d_{sim,j} - \bar{d}_{sim}\right)^T \tag{3}$$

$$C_{DD}^f = \frac{1}{Ne - 1} \sum_{j=1}^{Ne} \left(d_{sim,j} - \bar{d}_{sim}\right)\left(d_{sim,j} - \bar{d}_{sim}\right)^T \tag{4}$$

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

Ne is the number of models per ensemble, Nd the number of data, σ^2 the variance and $\bar{m}^f$ the mean of certain attribute. It is important to point out that the matrix shown in Eq. 5 is diagonal because of the difficulty to measure the correlation between observed data error. More information about the method can be found in Emerick and Reynolds (2013).

To avoid spurious correlations and an ensemble collapse, localization technique is used in the analysis equation (Eq. 1), which introduces a correlation matrix ρ via an element-by-element multiplication, also known as Schur product (∘) (Eq. 6). A common type of localization is the distance dependent. Watanabe and Datta-Gupta (2011) defined it as a function of the distance between the well (observation data point) and any point of the grid that influences the data from this specific well. Therefore, in this type of localization, the regions of influence of all data for each well must be defined, and the updates of petrophysical properties (permeability, porosity) are performed only in these defined regions based on the specified data (oil rate, water rate).

$$m_j^a = m_j^f + \rho \circ K\left(d_{obs,j} - d_{sim,j}\right) \tag{6}$$

There are many alternatives to calculate ρ when using the distance-dependent approach. One much-used formulation is defined by Gaspari and Cohn (1999), where ρ can assume values from 0 to 1 depending on the distance from the well location:

$$\rho(h, L) = \begin{cases} -\frac{1}{4}\left(\frac{h}{L}\right)^5 + \frac{1}{2}\left(\frac{h}{L}\right)^4 + \frac{5}{8}\left(\frac{h}{L}\right)^3 - \frac{5}{3}\left(\frac{h}{L}\right)^2 + 1, & 0 \le h \le L \\ \frac{1}{12}\left(\frac{h}{L}\right)^5 - \frac{1}{2}\left(\frac{h}{L}\right)^4 + \frac{5}{8}\left(\frac{h}{L}\right)^3 + \frac{5}{3}\left(\frac{h}{L}\right)^2 - 5\left(\frac{h}{L}\right) + 4 - \frac{2}{3}\left(\frac{h}{L}\right)^{-1}, & L < h \le 2L \\ 0, & h > 2L \end{cases} \tag{7}$$

here h is the Euclidean distance between any grid cell and well location and L is the critical length, which delineates the region of influence of each type of data from the wells. Thus, grid blocks close to the wells will present a high value of ρ and, consequently, during the assimilation process, will be highly influenced by the data from that well. However, this type of localization is not used to update uncertain scalar parameters as they are used to represent the whole (or part) of the reservoir model.

The ρ matrix comprises the elements ρ_{ab}, where 'a' is the number of rows and 'b,' the number of columns. The rows represent the uncertainties in each grid block and the column represents historical data at specific historical times. Therefore, depending on the location of the grid block in relation to a specific well, ρ_{ab} will be equal to the value calculated by Eq. 7, depending on h and L.

Chen and Oliver (2014) used an iterative ensemble smoother method, similar to ES-MDA, in the Norne field (real case). They highlighted the improvement in the history-matching process over a manual process. Nevertheless, because of the high complexity of the case, they believe that the localization approach they used could be improved.

Morosov and Schiozer (2016) carried out a closed-loop reservoir management and development process using the UNISIM-I-D benchmark case (Avansi and Schiozer 2015) and, during history-matching phases, they used ES-MDA without a localization technique. They reported a strong reduction in the variability of the uncertainties, resulting in the exclusion of several models that could have a good representation of the reservoir. In fact, they observed that simulated responses did not capture the reference model, which represented the real reservoir.

Silva et al. (2017) also presented a closed-loop process for the UNISIM-I-D benchmark with ES-MDA. The main difference from the work of Morosov and Schiozer (2016) was the use of distance-dependent localization. They reported a better response when compared with the reference model. However, localization range was constant for all data and all wells, and it was used to update only petrophysical images.

Soares et al. (2018) proposed a methodology using ES-MDA in history matching with distance-dependent localization based on streamlines. They observed that localization improved the petrophysical images generated after the application of the method. On the other hand, they pointed out that uncertainties such as water relative permeability produced a significant reduction in variability since distance-dependent localization is not used to update scalar uncertainties.

Furrer and Bengtsson (2007) developed a non-distance-dependent localization (NDDL) based on the correlation coefficient (cc) calculated between two variables (a and b), in this case representing uncertainties and observed data, which are used in the definition of ρ, as Eq. 8 shows.

$$\mathrm{cc}_{ab} = \frac{c_{ab}^2}{c_{ab}^2 + (c_{ab}^2 + \sigma_a * \sigma_b)/\mathrm{Ne}} \tag{8}$$

where σ is the standard deviation of a certain attribute; Ne, the number of models; and c, the covariance between two attributes. In this case, c_{ab} corresponds to the values of the elements of $C_{\mathrm{MD}}^{\mathrm{f}}$ in reference to uncertainties and observed data. Furrer and Bengtsson (2007) stated that for small values of cc_{ab}, which mean that when two parameters do not have a strong correlation, ρ_{ab} should be zero. Therefore, it is possible to generate data sparseness, increasing the variability of the final distribution of the uncertainties. To decide when ρ_{ab} should be zero, they specified a threshold value (λ), thus, if the correlation between an uncertainty and a data point is $< \lambda$, ρ_{ab} equals zero and this data point is not used to update this uncertainty. If the correlation is higher than λ, ρ_{ab} equals one and the data is used to update the uncertainty. The covariance between two parameters is calculated as follows.

$$c_{ab} = \frac{1}{\mathrm{Ne} - 1} \sum_{j=1}^{Ne} (a_j - \bar{a})(b_j - \bar{b}) \tag{9}$$

Additionally, Luo et al. (2016) highlighted that some data, especially seismic, may not be associated with specific physical locations (such as scalar uncertainties). Therefore, the use of distance-dependent localization becomes an issue. Because of this, they proposed a correlation-based adaptive localization to be used with seismic data in a history-matching process, where they considered a threshold value to update the parameters, following the same logic of Furrer and Bengtsson (2007). Finally, they noted that the proposed methodology should also work for production data and values ranging from 0 to 1 could be used for ρ, as in the distance-dependent localization.

Lacerda et al. (2019) compared several methods (NDDL—Furrer and Bengtsson (2007), correlation-based—Luo et al. (2016), among others) with the same objective, and because most of these methods need some internal parameters that can have a large effect on the final response, their use might require the calibration of these parameters.

Methodology

In this work, we propose and compare three different approaches to increase the variability of scalar uncertainties without impairing data match, model consistency or production forecast. Figure 1 shows the different approaches used.
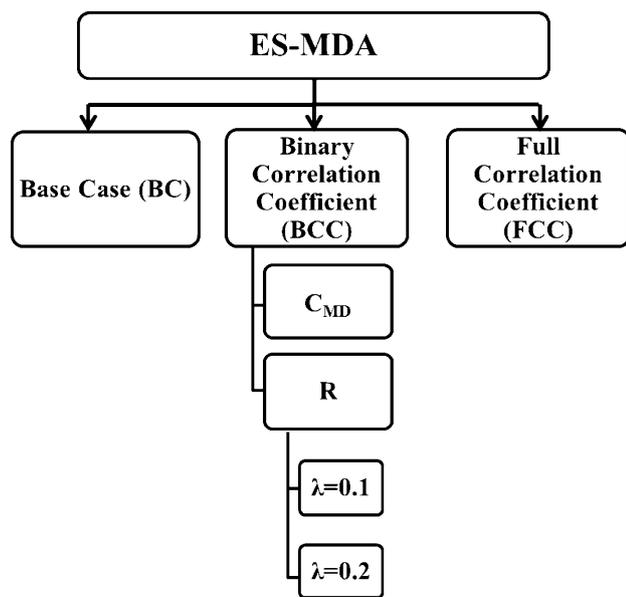
Fig. 1 Different approaches using ES-MDA and localization

Table 1 Base case (BC) characteristics

Parameter	Value
Base case	
Number of iterations (Ni)	6
Number of models (Ne)	500
C_D	10%
α_i	6 (constant)
Truncated singular value (tsv) in matrix inversion	0.99

Each approach is explained below.

Base case

The Base Case (BC) was developed by Soares et al. (2018), where ES-MDA and distance-dependent localization based on streamlines were used to update only petrophysical uncertainties. Note that during the simulation of the historical period, they specified the liquid rate for producer wells and water rate for injector wells as boundary condition, and during the assimilation process (Eq. 1), they used oil rate (q_o), water rate (q_w), liquid rate (q_l) and bottom-hole pressure (BHP) of producer wells and water rate (q_{wi}) and BHP for injector wells to update the uncertainties. All data were used to update scalar uncertainties.

Table 1 shows the characteristics of the application of ES-MDA.

Binary correlation coefficient (BCC)

The characteristics of the base case were the same for this and the approach FCC. The only difference concerns the updates of the scalar parameters. To update these uncertainties, we analyzed the correlation between historical data and uncertain scalar parameters through correlation coefficients. We then selected the data that presented the highest correlation and used only those to update each scalar uncertainty. As the name suggests (binary correlation coefficient), a threshold value of the correlation coefficient must be defined to determine whether data (oil rate, water rate, etc.) is used to update a particular scalar parameter. Therefore, to use this approach, ρ_{ab} (used with the localization technique) is set as 1, if the data are used to update a scalar uncertainty; and as 0, if not.

Note that localization can be defined as the procedure used to limit the influence of historical data to a specific region (Aanonsen et al., 2009). Thus, despite using the same technique (using ρ) to update scalar and petrophysical uncertainties, defining ρ_{ab} for uncertain petrophysical and scalar parameters represents different procedures. While specific regions can be assigned to petrophysical parameters, the same is not true for scalar uncertainties, which are usually related to the whole reservoir model. For that reason, we preferred not to use non-distance-dependent localization or adaptive localization nomenclature, as some authors did (Furrer and Bengtsson, 2007; Luo et al. 2016).

Finally, the difference between C_{MD} and R relies on the method used to calculate the correlation coefficient between historical data and scalar uncertainties. The explanation of each procedure follows.

Cross-covariance calculated by ES-MDA (C_{MD})

In this approach, we used a procedure similar to NDDL developed by Furrer and Bengtsson (2007). Thus, we used the correlation coefficient (cc_{ab}) calculated based on the C_{MD}^f to define which data were used to update each scalar uncertainty. This definition was based on a threshold value equal to 0.05, as suggested by Furrer and Bengtsson (2007). In other words, for correlation coefficient (cc_{ab}) values between observed data and scalar uncertainties less than 0.05, $\rho_{ab}=0$, and for cc_{ab} values higher than 0.05, $\rho_{ab}=1$. Petrophysical uncertainties were updated using the methodology based on Soares et al. (2018), as mentioned in the previous section.

Correlation coefficient based on objective functions and scalar uncertainties (R)

As mentioned previously, the novelty of this approach is the method used to compute the correlation coefficient between

scalar uncertainties and historical data. The correlation coefficient (R) we use is based on the work of Maschio and Schiozer (2016), where they calculated the correlation between uncertain parameters and the Objective Function (OF).

$$R_{ab} = \frac{c_{ab}}{\sqrt{\sigma_a * \sigma_b}} \tag{10}$$

where, in this approach, a is the uncertainties and b is the OF, representing simulated data.

The OF used was the Normalized Quadratic Deviation with Sign (NQDS). Basically, NQDS can be understood as the quadratic deviation with sign between d_{sim} and d_{obs} of all historical data of a variable (q_o, for instance) for a well divided by a determined tolerance plus a constant for each type of data. Thus, NQDS works as a multiobjective function and demands the analysis of the definition of variables. For this approach, we set q_o, q_w, q_{wi} and BHP as the OFs. Mathematically, NQDS is represented as follow.

$$\text{NQDS} = \frac{\sum_{i=1}^{Nd}(d_{sim,i} - d_{obs,i})}{\left|\sum_{i=1}^{Nd}(d_{sim,i} - d_{obs,i})\right|} * \frac{\sum_{i=1}^{Nd}(d_{sim,i} - d_{obs,i})^2}{\sum_{i=1}^{Nd}(\text{Tol} * d_{obs,i} + \text{Const})^2} \tag{11}$$

here Nd is the total number of observed data, Tol is the tolerance as a percentage and Const is the constant to avoid division by zero (for details, see Maschio and Schiozer 2016).

In addition, a threshold value was assigned (λ) to define what data were used to update the uncertain scalar parameter. As a basic guideline, Maschio and Schiozer (2016) suggested that values of λ that indicate a moderate correlation can be adequate (between 0.1 and 0.4). For our case, we found that values higher than 0.2 were too extreme and we tested λ as 0.1 and 0.2. Therefore, $\rho_{ab} = 0$ for R_{ab} values lower than the threshold value assigned (λ), and $\rho_{ab} = 1$ for R_{ab} values higher than λ. For $\lambda = 0.2$, fewer data points are used to update the uncertainties and, consequently, we expect higher variability. Note that this procedure was used only for scalar parameters.

Full correlation coefficient (FCC)

FCC uses ρ_{ab} values equal to the absolute value of R_{ab}, which ranges from 0 to 1. Thus, in this case, uncertain scalar parameters that are more influenced by certain data, represented by the OFs, will have a higher value of R_{ab} and, consequently, of ρ_{ab}. For example, if water relative permeability (k_{rw}) has a larger influence on q_w than the BHP from Well 1, then, $\rho_{qw\,krw}$ will be higher than $\rho_{BHP\,krw}$. Therefore, if a given uncertain attribute has more influence on certain data, this data will have a higher impact when updating this specific attribute.

In summary, this approach assigns a weight factor to better identify the influence of specific observed data to

update scalar uncertainties, avoiding excessive reduction. The combination of using distance-dependent localization and correlation coefficient to update petrophysical and scalar uncertainties, respectively, is a novel methodology aiming to generate models capable of representing the reservoir without excluding some potential plausible answers.

Comparison between approaches

To analyze and compare each approach, we used the methodology proposed by Soares et al. (2018), which follows these steps:

A. Data match analysis

Data match analysis is performed through the multiobjective function, NQDS. We analyze the variables q_o, q_w, q_{wi} and BHP.

B. Uncertainty reduction

To assess uncertainty reduction, the variability of the final ensemble and convergence of the responses must be verified, especially for the scalar uncertainties. To measure this variability, we check the range of the posterior distribution.

C. Filter

The filter process is responsible for selecting acceptable models according to data match analysis and a threshold value of NQDS. Note that for a model to be considered acceptable, all analyzed variables must be within the defined threshold value.

D. Production forecast

Finally, with the filtered models from the previous step (C), production is forecast, assessing whether the oil and water cumulative production encompasses the reference model.

Application: UNISIM-I-H

We applied the methodology to a benchmark case with a known response, UNISIM-I-H. Avansi and Schiozer (2015) developed a reference model (UNISIM-I-R) based on data from the Namorado field, an offshore field located in Campos Basin (Brazil). UNISIM-I-R is a very fine grid model containing $326 \times 234 \times 157$ cells (3,408,633 active cells) with a resolution of $25 \times 25 \times 1$ m. Therefore, UNISIM-I-R
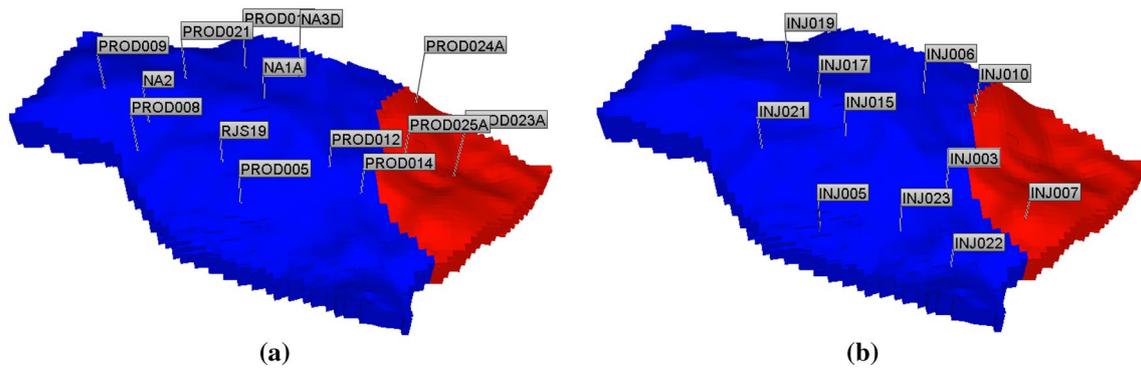
Fig. 2 Main block of UNISIM-I-H in blue and east block in red. Producer wells (**a**) and injector wells (**b**)

is assumed to be the reference case, which corresponds to the true model of the benchmark.

In order to test and validate new methodologies, Avansi and Schiozer developed the UNISIM-I-H, which is a coarser model derived from the reference model (UNISIM-I-R), and despite being a synthetic case, it has the characteristics and complexity of a real field.

UNISIM-I-H represents a reservoir composed by consolidated sandstones with a sealing fault, which divides the field into two production areas: main block and East block, as shown in Fig. 2. Avansi and Schiozer (2015) discretized the model in $81 \times 58 \times 20$ cells and each cell measures, on average, $100 \times 100 \times 8$ m. The model has 14 producers and 11 injectors (only water injection), and a black-oil simulator is used. Among the 14 producers, 4 are verticals and 10 are horizontal located at the top of the reservoir. Injectors are located at the basis of the field. The historical period is 11 years and forecast time is from 11 to 30 years. Figure 2 also shows the distribution of the wells.

As mentioned previously, uncertainties were divided into two groups, scalar and petrophysical. Avansi and Schiozer (2015) and Avansi et al. (2016) performed a careful modeling of the benchmark. Table 2 shows the uncertain attributes

considered and their initial parameterization. Regarding the petrophysical uncertainties, for instance, prior images come from geostatistical realizations generated by a commercial tool. In addition, they used data from well logging to check quality of the images. Regarding the scalar uncertainties, the authors carried out a sensitivity analysis to select the most influential uncertain parameters to be used during history matching. In addition, to represent water relative permeability curve, we used Corey (1954) formulation:

$$k_{rw} = k_{rwiro} * \left(\frac{S_w - S_{wcrit}}{1 - S_{wcrit} - S_{orw}} \right)^{P_{krw}} \tag{12}$$

where S_w is the water saturation, S_{wcrit} is the critical water saturation and S_{orw} is the residual oil saturation.

Avansi and Schiozer, (2015) selected the range of the uncertain parameters based on the knowledge of the Namorado field while avoiding values that are not compatible with the case. Finally, because UNISIM-I-H is a benchmark case and different authors have been using it (Morosov and Schiozer 2016; Silva et al. 2017; Soares et al. 2018), it is important to keep some basic features of the case in order to validate and compare different methodologies.

Table 2 Uncertainties defined for the UNISIM-I-H case

Type	Uncertainty	Minimum	Maximum	Mean	Distribution
Petrophysical	Porosity	0.00	0.31	–	–
	Permeability I (k_x) (mD)	1	5000	–	–
	Permeability J (k_y) (mD)	1	5000	–	–
	Permeability Z (k_z) (mD)	1	500	–	–
	NTG ratio	0	1	–	–
Scalar	Water–oil contact from East block (DWOC) (m)	3169	3179	3174	Triangular
	Rock compressibility (CPOR) (kgf/cm^2)$^{-1}$	1.0×10^{-5}	9.6×10^{-5}	5.3×10^{-5}	Triangular
	k_z multiplier (MOD)	0.01	3.00	1.50	Triangular
	Maximum water relative permeability (k_{rwiro})	0.15	0.52	0.33	Triangular
	Water relative permeability exponent (P_{krw})	1.50	3.30	2.30	Triangular

Table 3 Parameters used in NQDS calculation

Data	Tolerance (%)	Constant
q_w	10	20 m³/day
q_o	10	0 m³/day
q_{wi}	5	0 m³/day
BHP	10	0 kgf/cm²

Results and discussion

A. Data match

To analyze the data match, we computed NQDS using the tolerance and constant values shown in Table 3.

NQDS values were similar for all variables. Figure 3 shows NQDS for q_w, which presented a symmetrical response for most wells, with values close to zero. The exception, PROD023A, tended to overestimate water

production for all approaches. Figure 4 illustrates water production for wells PROD012A (Fig. 4a) and PROD023A (Fig. 4b) under the FCC approach. PROD012A presented water rate curves close to the historical points, as NQDS values suggested, while PROD023A tended to overestimate values, especially at the beginning of water production.

B. Data match

In Fig. 5, it is possible to see the distribution of the scalar uncertainties, where the y-axis represents their values. As suggested by Soares et al. (2018) and as we can see in Fig. 5, both BC and BCC-C_{MD} caused a large reduction in the variability in scalar uncertainties. The only exception was DWOC because water–oil contact is uncertain only in the East block, thus, the only data used for updates were from wells located in that part of the field (PROD023A, PROD024A, PROD025A, INJ007 and INJ010). Furthermore, the final convergence for different responses, as
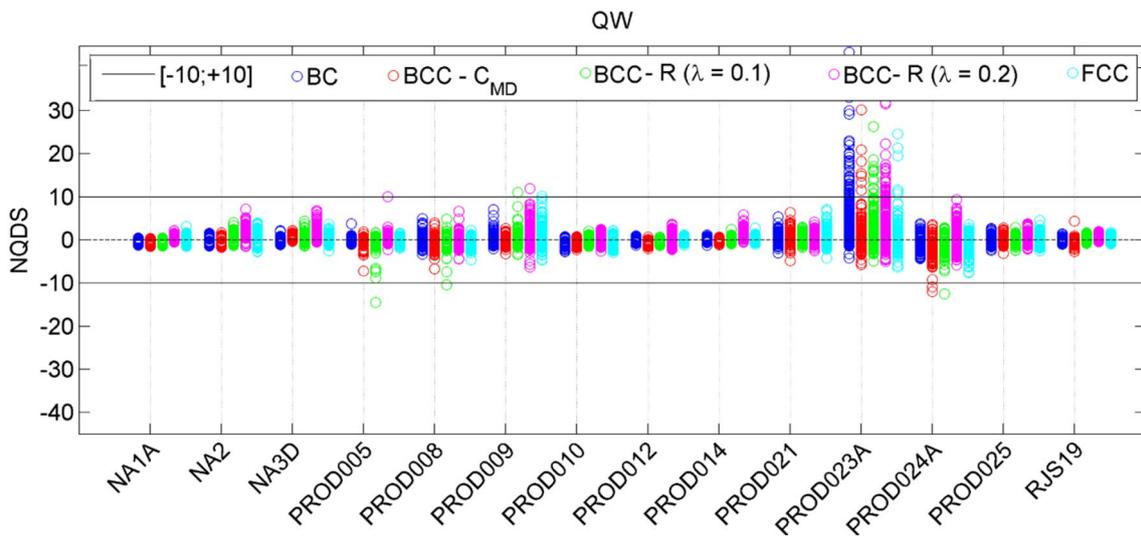
Fig. 3 NQDS for q_w

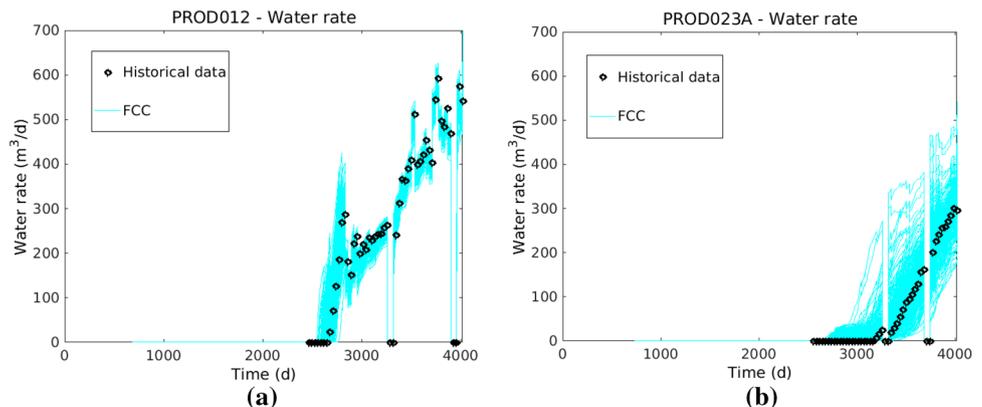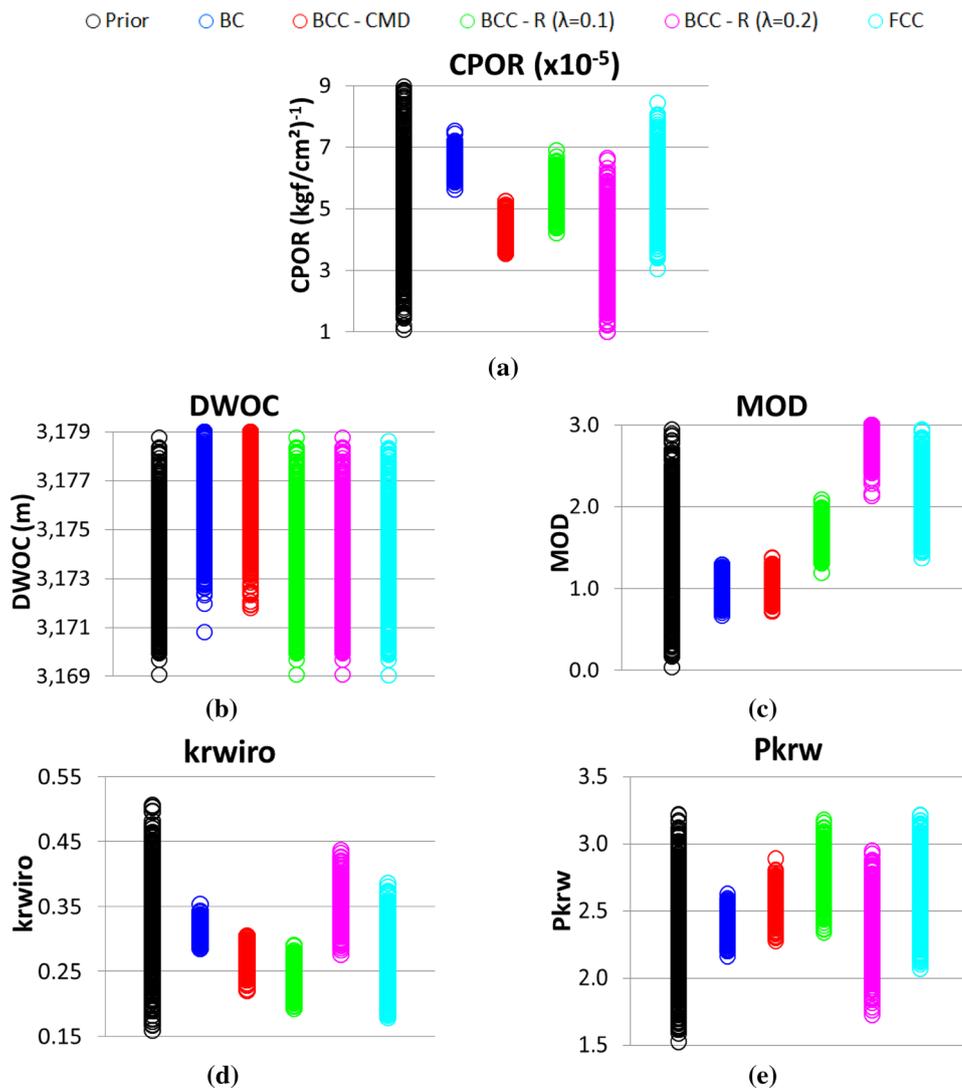**Fig. 4** Water rate: PROD012-FCC (**a**); and PROD023A-FCC (**b**)

Fig. 5 Scalar uncertainties: CPOR (**a**), DWOC (**b**), MOD (**c**), k_{rwiro} (**d**) and P_{krw} (**e**)

CPOR and k_{rwiro} values demonstrated, indicated the presence of other possible responses for this specific problem.

BCC–R ($\lambda = 0.1$) showed that there are other possible responses when analyzing only BC and BCC-C_{MD}, as the convergence of MOD (Fig. 5c) illustrates. Moreover, because fewer data points were used to update the uncertainties, BCC–R ($\lambda = 0.1$) presented higher variability in the final answer. This higher variability, in conjunction with lower data mismatch, is a key point for the proposed methodology, as we aim to represent all possible models that might have a good representation of the real reservoir and, therefore, generate a more reliable forecast. However, the final response for MOD (Fig. 5c), for example, depicted that BCC–R ($\lambda = 0.1$) also found different answers to previous applications.

For BCC–R ($\lambda = 0.2$), the most influential data were used to update scalar uncertainties, generating higher variability

for this approach compared with BCC–R ($\lambda = 0.1$), as CPOR (Fig. 5a) and P_{krw} (Fig. 5e) values show. Consequently, BCC–R ($\lambda = 0.2$) generated even more variability while respecting historical data. Nevertheless, as observed for BCC–R ($\lambda = 0.1$), when using $\lambda = 0.2$, MOD values concentrated in different values, demonstrating that the methodology can be further improved.

Note that for DWOC, BCC–R ($\lambda = 0.1$) and BCC–R ($\lambda = 0.2$) presented the same distribution as the initial (prior) ensemble of models. This was because the correlation between the objective function and this variable was weak, i.e., R_{ab} values were < 0.1 and 0.2, respectively. Consequently, $\rho_{ab} = 0$, resulted in $m^a = m^f$, as shown in Eq. 6.

Finally, FCC generated more likely levels of influence for each data point when updating scalar uncertainties and, because of this, FCC achieved better variability for almost all attributes. With the exception of MOD, FCC encompassed
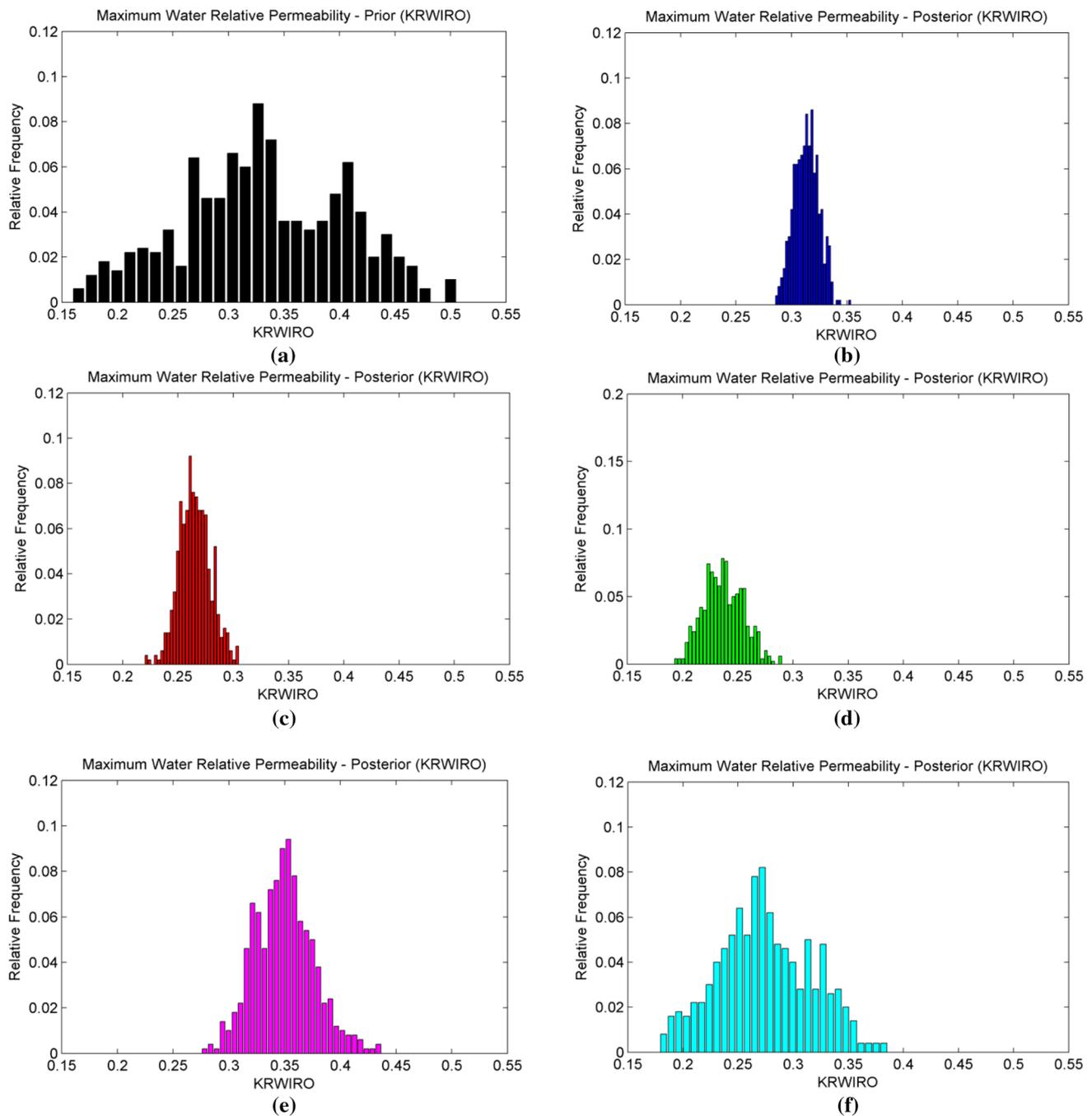
Fig. 6 k_{rwiro} histograms: Prior (**a**); b through f, posterior: BC (**b**); BCC-C$_{MD}$ (**c**); BCC–R ($\lambda = 0.1$) (**d**); BCC–R ($\lambda = 0.2$) (**e**); and FCC (**f**)

most of the responses from other approaches. Thus, the higher variability of the scalar uncertainties contributed to generating more different scenarios and the lower data mismatch shown in part A demonstrated that these models represent more adequate the reference model compared to BCC.

Figure 6 illustrates the final distribution of the scalar parameters as in Fig. 5, but using histograms (relative frequency) for k_{rwiro}, where we can see the higher variability for the FCC approach (Fig. 6f).

In addition, despite different responses for the scalar uncertainties, petrophysical uncertain attributes showed very similar patterns for the mean of ln (k_x) (Fig. 7).
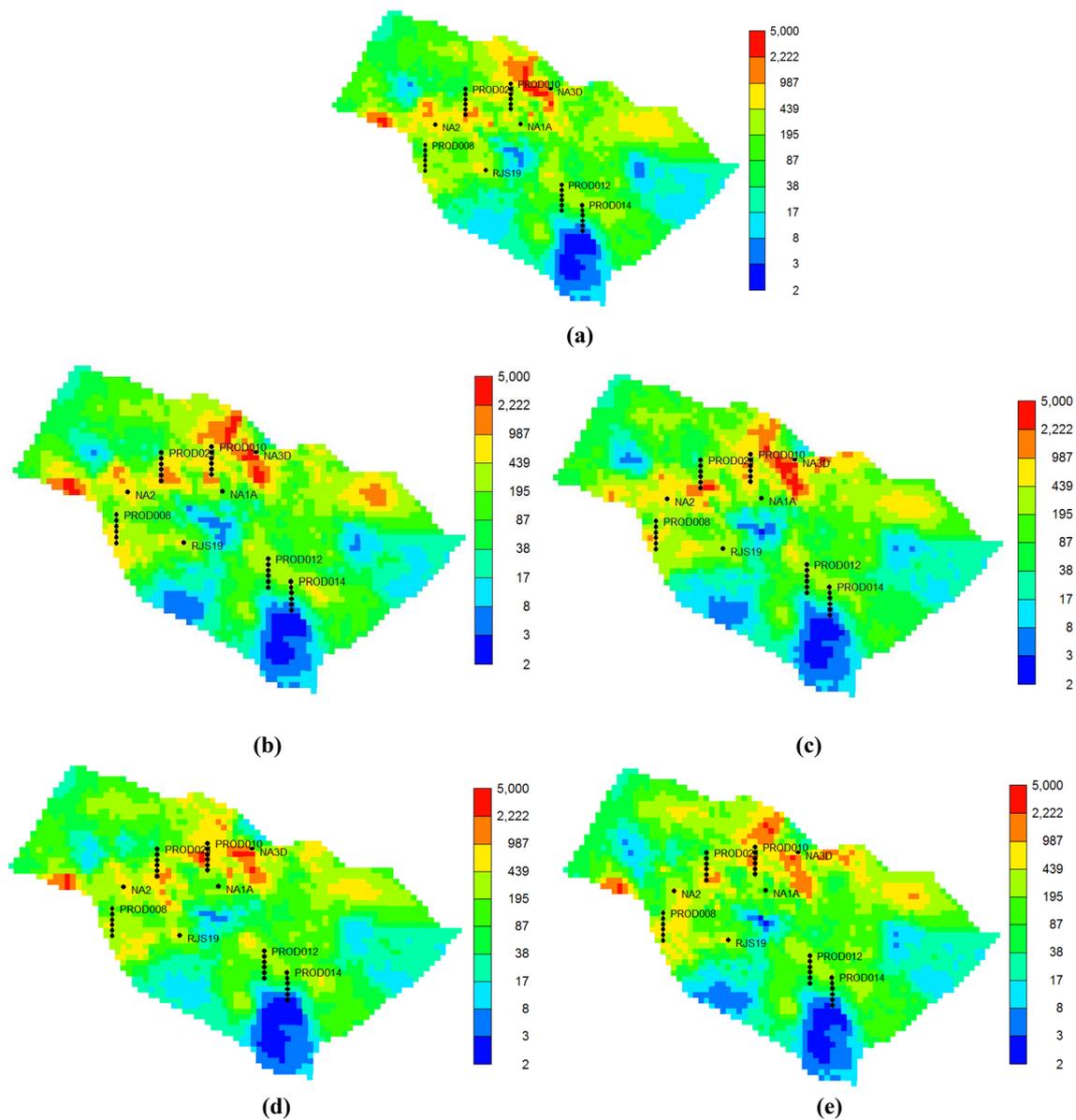
Fig. 7 Mean of ln (k_x): BC (**a**); BCC-C_{MD} (**b**); BCC–R ($\lambda = 0.1$) (**c**); BCC–R ($\lambda = 0.2$) (**d**); and FCC (**e**)

Table 4 Filtered models

Approach	% Filtered
BC	91.6
BCC-C_{MD}	97.8
BCC–R ($\lambda = 0.1$)	95.2
BCC–R ($\lambda = 0.2$)	94.2
FCC	97.6

using acceptable models with NQDS values lower than or equal to an absolute value of 10, many models were filtered, as shown in Table 4. Despite presenting a higher variability of the scalar uncertainties, FCC filtered most models (97.6%).

D. Production forecast

Using the acceptable models defined in the previous step (C), we forecasted field behavior and noticed that all approaches were capable of encompassing the reference response for cumulative oil production (N_p) and cumulative water production (W_p). Note that the forecast reference

C. Filter

As already shown by NQDS values, data matches were similar for all approaches and, with the filtration process

Springer
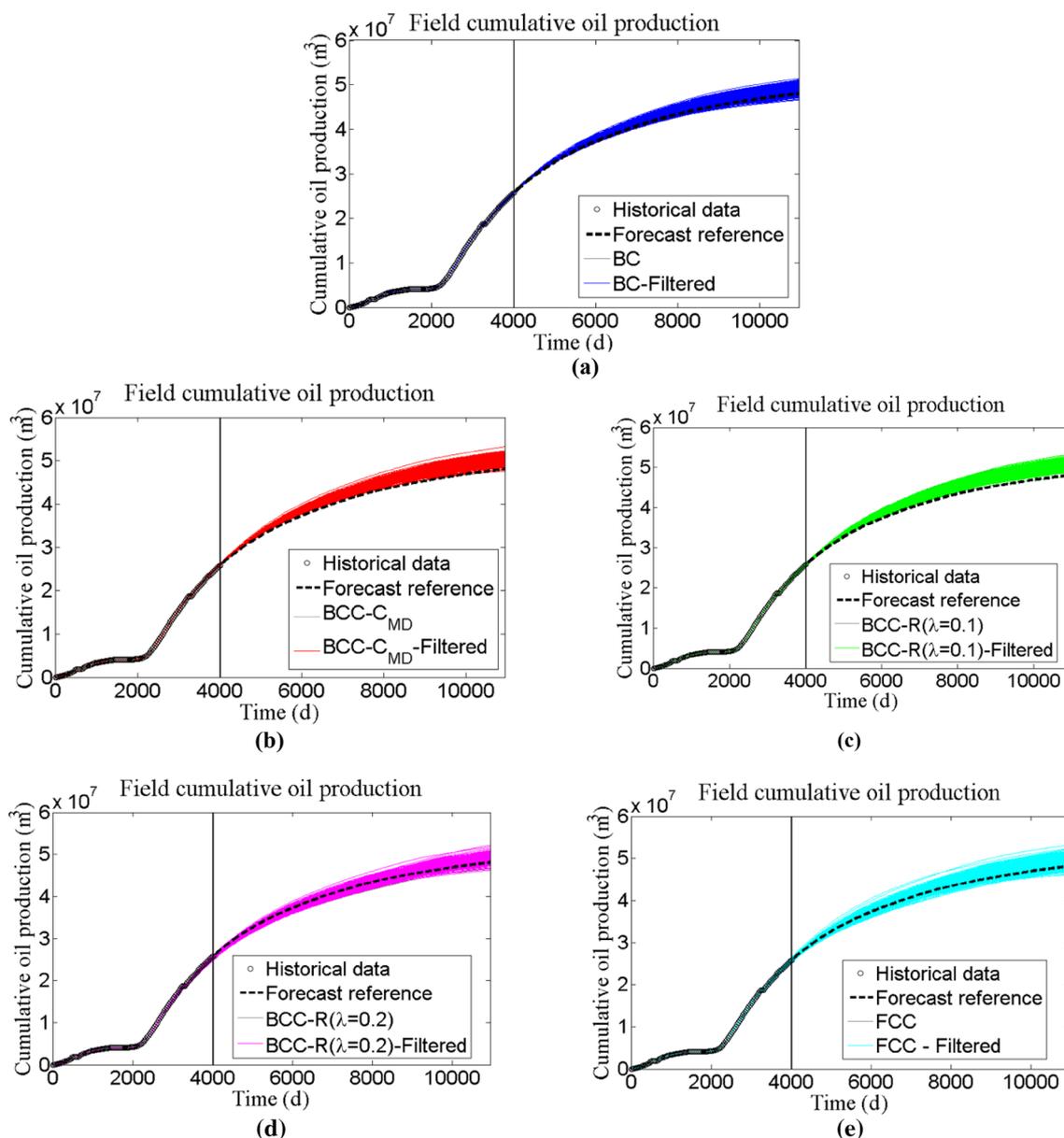
Fig. 8 Field cumulative oil production compared to the forecast reference from the true case: BC (**a**); BCC-C_{MD} (**b**); BCC–R ($\lambda = 0.1$) (**c**); BCC–R ($\lambda = 0.2$) (**d**); and FCC

corresponds to the UNISIM-I-R, which is the reference field treated as our true answer. Figure 8 depicts N_p for all approaches and, because of different combinations of scalar uncertainties, the approaches presented different responses. For BCC–C_{MD} and BCC–R ($\lambda = 0.1$), the reference answer was at the lower limit of the final ensemble while other approaches presented more symmetrical responses compared with the reference response. Figure 9 illustrates W_p and the similar responses for all approaches. Furthermore, FCC showed a slightly higher variability in the forecast when compared with all other approaches. Note that we filtered most of the models, therefore the

curves in gray are hidden by the colorful lines in Figs. 8 and 9.

In conclusion, FCC affected the scalar parameters just as the distance-dependent localization affected the petrophysical images. Therefore, values of ρ ranging from 0 to 1 were assigned, which helped to increase the variability of the scalar uncertainties in a more plausible way, performing a better exploration in the defined search space. However, as all approaches assessed in this work used correlation between production/injection data and uncertain attributes, care must be taken because if there is any spurious correlation among those variables, once it will be carried during
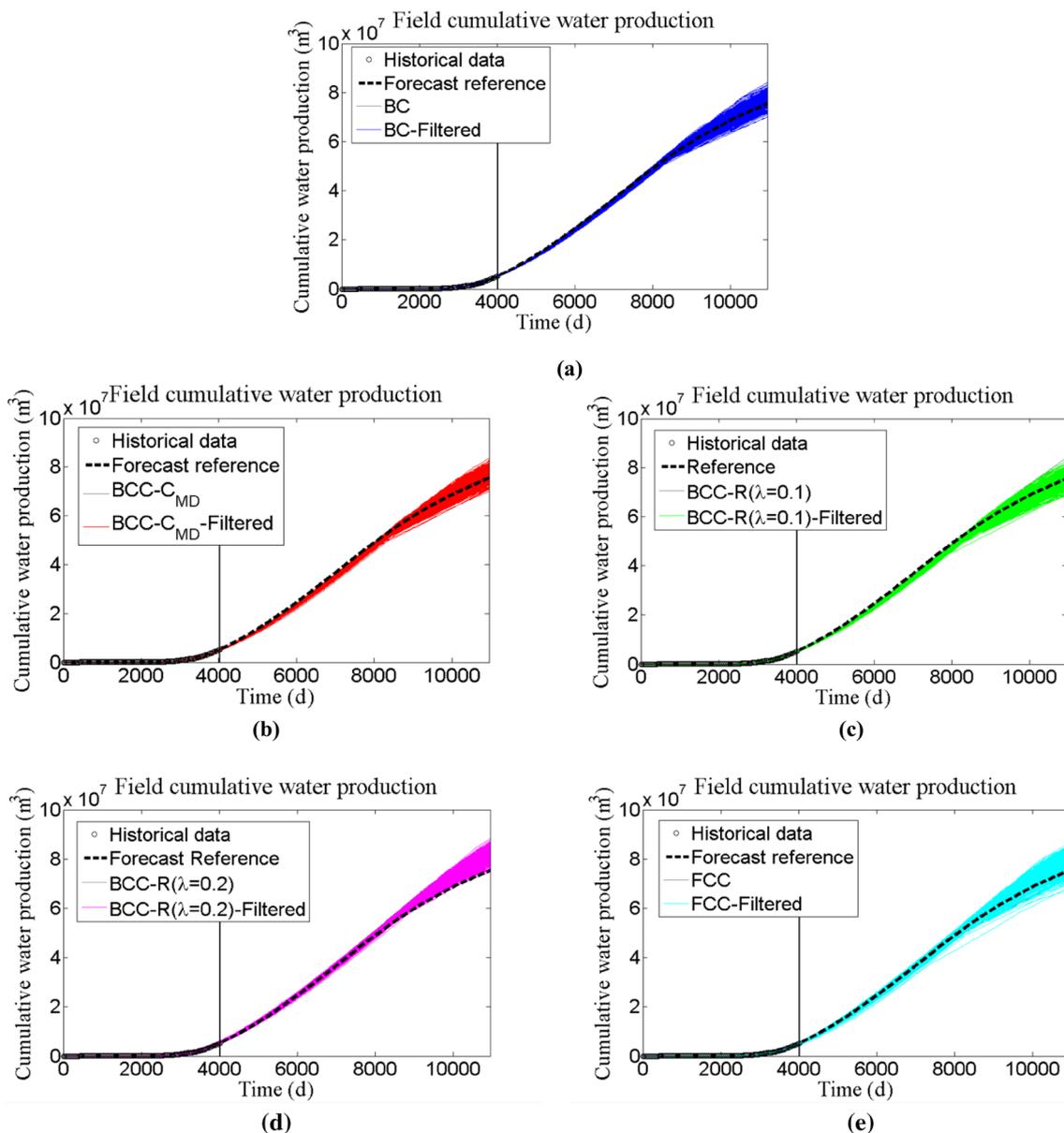
Fig. 9 Field cumulative water production compared to the forecast reference from the true case: BC (**a**); BCC-C_{MD} (**b**); BCC–R ($\lambda = 0.1$) (**c**); BCC–R ($\lambda = 0.2$) (**d**); and FCC (**e**)

the history-matching process. As a result, the process might generate models that poorly represent the real reservoir. To address this, we recommend using a large number of models (around 500), as already suggested by Soares et al. (2018) and Aanonsen (2009).

Finally, ensemble-based methods have been attracting attention within the oil industry due to its ability to handle large models and its relative easiness of implementation compared to other data assimilation methods, such as gradient-based and MCMC. Several works showed the effectiveness of the method when performing history matching.

As already mentioned by authors such as Aanonsen et al. (2009) and Emerick and Reynolds (2011), for instance, the use of ensemble-based methods with distance-dependent localization helps to generate more geological consistent models. Nevertheless, one drawback of this method is the possible collapse of the response on the final ensemble. Therefore, the methods described in this work (BCC and FCC, for example) can help dealing with this issue, making ensemble-based methods more suitable for data assimilation. In addition, compared to methods as MCMC, ensemble-based methods demands less computational time

and the methodology proposed in this work did not result in any extra significant computational time compared to other ensemble-based method methodologies.

Conclusions

This work presented new approaches based on localization techniques to increase the variability of scalar uncertainties with the application of the ES-MDA method.

The approach BCC-C_{MD} presented a similar response to the Base Case (BC), where no localization technique was used to update scalar uncertainties, with a strong reduction in scalar uncertainties.

BCC–R generated better results when compared with BCC-C_{MD}, and for higher values of λ, only a few data points were used to update scalar uncertainties and this explains the higher variability of BCC–R ($\lambda = 0.2$) compared with BCC–R ($\lambda = 0.1$).

FCC identified more plausible levels of influence of each data point for uncertain scalar attributes and used this knowledge to update them. This generated higher variability for the scalar uncertainties without losing quality in data matches and production forecasts.

Note that if too few models are used, spurious correlations may be generated. Consequently, this may affect the updates of petrophysical and scalar uncertainties and, therefore, the data matches and production forecasts.

References

Aanonsen SI, Nævdal G, Oliver DS, Reynolds AC, Vallès B (2009) The Ensemble Kalman Filter in reservoir engineering—a review. SPE J 14(3):393–412

Arroyo-Negrete E, Devegowda D, Datta-Gupta A (2008) Streamline–assited Ensemble Kalman Filter for rapid and continuous reservoir model updating. SPE Reserv Eval Eng 11(6):1046–1060

Avansi GD, Schiozer DJ (2015) UNISIM-I: synthetic model for reservoir development and management applications. J Model Simul Pet Ind 9(1):21–30

Avansi GD, Maschio C, Schiozer DJ (2016) Simultaneous history matching approach using reservoir-characterization and reservoir-simulation studies. SPE Reserv Eval Eng 19:694–712

Chen Y, Oliver DS (2014) History matching of the Norne full field model using an iterative ensemble smoother. SPE Reserv Eval Eng 17(2):244–256

Corey AT (1954) The interrelation between gas and oil relative permeabilities. Prod Mon 19:38–41

Emerick AA (2016) Analysis of the performance of ensemble-based assimilation of production and seismic data. J Pet Sci Eng. https://doi.org/10.1016/j.petrol.2016.01.029

Emerick AA, Reynolds AC (2011) Combining sensitivities and prior imformation for covariance localization in the Ensemble Kalman Filter for petroleum reservoir applications. Comput Geosci 15:251–269

Emerick AA, Reynolds AC (2013) Ensemble smoother with multiple data assimilation. Comput Geosci 55:3–15

Evensen G (1994) Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 99(5):10143–10162

Furrer R, Bengtsson T (2007) Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J Multivar Anal 98(2):227–255

Gaspari G, Cohn S (1999) Construction of correlation functions in two and three dimensions. Q J R Meteorol Soc 125(554):723–757

Houtekamer PL, Mitchell HL (1998) Data assimilation using Ensemble Kalman Filter technique. Mon Weather Rev 126:796–811

Lacerda JM, Emerick AA, Pires AP (2019) Methods to mitigate loss of variance due to sampling errors in ensemble data assimilation with non-local model parameters. J Pet Sci Eng 172:690–706

Luo X, Bhakta T, Nævdal G (2016) Data driven adaptive localization with applications to ensemble-Based 4D seismic history matching, SPE-185936-MS. In: SPE Bergen one day seminar, April 5, Bergen, Norway

Maschio C, Schiozer DJ (2016) Probabilistic history matching using discrete Latin Hypercube sampling and nonparametric density estimation. J Pet Sci Eng 147:98–115

Morosov AL, Schiozer DJ (2016) Field development process revealing uncertainty assessment pitfalls, SPE-180094-MS. In: SPE Europec, May 30–Jun 2, Vienna, Austria

Oliver DS, Chen Y (2011) Recent progress on reservoir history matching: a review. Comput Geosci 15:185–221

Silva VLS, Emerick AA, Couto P, Alves JLD (2017) History matching and production optimization under uncertainties—application of closed-loop reservoir management. J Pet Sci Eng 157:860–874

Soares RV, Maschio C, Schiozer DJ (2018) Applying a localization technique to Kalman gain and assessing the influence on the variability of models in history matching. J Pet Sci Eng 169:110–125

Watanabe S, Datta-Gupta A (2011) Use of phase streamlines of covariance localization in Ensemble Kalman Filter for three-phase history matching, SPE-144579-MS. In: SPE Western North American Region Meeting, May 7–11, Anchorage, Alaska

Yustres A, Asensio L, Alonso J, Navarro V (2011) A review of Markov Chain Monte Carlo and information theory tools for inverse problems in subsurface flow. Comput Geosci 16(1):1–20