Scenario reduction methodologies under uncertainties for reservoir development purposes: distance-based clustering and metaheuristic algorithm

Seyed Kourosh Mahjour¹ · Antonio Alberto Souza Santos¹ · Manuel Gomes Correia¹ · Denis José Schiozer¹

Received: 25 February 2021 / Accepted: 7 June 2021 / Published online: 19 June 2021
© The Author(s) 2021

Abstract

The simulation process under uncertainty needs numerous reservoir models that can be very time-consuming. Hence, selecting representative models (RMs) that show the uncertainty space of the full ensemble is required. In this work, we compare two scenario reduction techniques: (1) Distance-based Clustering with Simple Matching Coefficient (DCSMC) applied before the simulation process using reservoir static data, and (2) metaheuristic algorithm (RMFinder technique) applied after the simulation process using reservoir dynamic data. We use these two methods as samples to investigate the effect of static and dynamic data usage on the accuracy and rate of the scenario reduction process focusing field development purposes. In this work, a synthetic benchmark case named UNISIM-II-D considering the flow unit modelling is used. The results showed both scenario reduction methods are reliable in selecting the RMs from a specific production strategy. However, the obtained RMs from a defined strategy using the DCSMC method can be applied to other strategies preserving the representativeness of the models, while the role of the strategy types to select the RMs using the metaheuristic method is substantial so that each strategy has its own set of RMs. Due to the field development workflow in which the metaheuristic algorithm is used, the number of required flow simulation models and the computational time are greater than the workflow in which the DCSMC method is applied. Hence, it can be concluded that static reservoir data usage on the scenario reduction process can be more reliable during the field development phase.

Keywords Petroleum field development · Scenario reduction · Uncertainty quantification · Cluster analysis · And metaheuristic algorithm

Introduction

Defining a reliable production strategy during the decision-making process and analysing it by computational simulators is normally challenging due to different types of uncertainty, including: (1) reservoir uncertainties related to geological and flow properties, (2) operational uncertainties associated with production system accessibility, (3) economic uncertainties linked to the oil price, investments and operational costs (Santos et al. 2018). Hence, it is possible to generate hundreds to thousands of reservoir models to capture all defined uncertainties. However, the numerical simulation

effort for a high number of models is not technically cost-effective and can be very time-consuming. Therefore, petroleum engineers use scenario reduction methods to choose a small subset of models (scenarios) that roughly reflect the features of the full ensemble, and these can also be applied in defining a production strategy. Consistent with this, two main groups of frameworks were introduced in the literature to reduce the number of scenarios: (1) scenario reduction using static features, (2) scenario reduction using simulation-based (dynamic) features (Kang et al. 2019).

The scenario reduction methods using static features are typically applied over geostatistical models before the flow simulation and data assimilation processes. In reservoir characterisation, a large number of petrophysical and facies models are generated using a geostatistical sampling (Scheidt and Caers 2009) but in practice, a high percentage of models are spatially and statistically similar to each other (Meira et al. 2016). Hence, the main assumption for

✉ Seyed Kourosh Mahjour
s.k.mahjour@cepetro.unicamp.br

¹ CEPETRO/FEM, University of Campinas (UNICAMP),
Postal Code 6052 13083-970, Campinas, São Paulo, Brazil

reducing the geostatistical models is that similar models have similar responses, and one is not required to consider all models in the ensemble for the simulation step. Instead, by grouping the similar models into different clusters and selecting a model as a representative from each cluster, the uncertainty range can be preserved (Alzraiee et al. 2012). There are several studies in the literature that present different ways of defining and evaluating the similarity or dissimilarity between geostatistical models. In most cases, a (static) distance matrix is determined by using static reservoir properties for measuring the similarity or dissimilarity of the models. It is noteworthy that the distance must be defined based on the type of objective function. The better the relationship between distance and objective function, the more efficient the distance approach will be. Suzuki and Caers (2006) calculated the Hausdorff distance, and Park and Caers (2007) computed the connectivity-based distance to measure the similarity for every pair of initial facies and petrophysical models. Mahjour et al. (2020a) also introduced another pairwise similarity distance for flow unit models using the simple matching coefficient. Furthermore, the scenario reduction techniques using static features are accepted as an applicable and user-friendly method in other research works (Mahjour et al. 2020b; Torrado et al. 2015; Armstrong et al. 2013).

Meanwhile, there are other approaches to select the subset of scenarios from simulation-based outputs or dynamic features. In these methods, the output variables of the simulator are used in reducing the number of scenarios. Therefore, we apply these methods on the simulation models after the flow simulation and data assimilation processes. According to the literature, the variety of techniques to reduce the scenarios using dynamic features are more common than static-based techniques. Shirangi and Mukerji (2012) used the k-medoids clustering technique to decrease the number of models using flow-based trends. Yasari et al. (2013) applied the ranking of Net present Values (NPV) obtained from an initial strategy to select a number of representative models from a full-set. (Sarma et al. 2013) introduced the minimax algorithm to select the closest subset to the three predefined percentiles P10, P50, P90 (quantities of cumulative probability), with respect to the NPV and Oil Recovery Factor (ORF) value. Meira et al. (2020) proposed the RMFinder tool as an optimization-based technique to select a set of RMs, automatically. RMFinder uses a metaheuristic algorithm with respect to the cross-plots and risk curves of the main simulation outputs and the probability distribution of the uncertain attributes. Furthermore, other studies based on different techniques have been applied for reducing scenarios using dynamic features (Schiozer et al. 2004; Caers and Park 2008).

The obtained subset of scenarios from different techniques can then be applied in the field development and

management phases during the decision-making process. Figure 1 shows the five summarized phases of the decision-making process extracted from the study of (Schiozer et al. 2019). During phase 1, all data obtained from a variety of sources such as core descriptions, well logs, and well testing are gathered for model construction. In phase 2, we perform building models, scenario development, and probability estimation. This important phase requires a comprehensive approach to consider all types of static and dynamic uncertainties. Phase 3 includes the prediction of reservoir performance as a key aspect of field development planning. Its success relies on the precise investigation of the rock and fluid properties using Pressure–Volume–Temperature (PVT) analysis. A field development plan in phase 4 is a long and intricate process that begins with the evaluation of urgent variables required to develop a field optimally. The selection and design of these variables such as recovery methods, well positions, artificial lifting, among others, are mandatory in this phase (Gaspar et al. 2016). Phase 5 involves the field management decisions. This phase represents the operational circumstances of the equipment over time. These conditions can be shifted at any moment, while the expense of changes is negligible. For example, changing the flow rates in each well over a specific period can affect the long-term assessments of design variables. Normally, to this end, some operation variables such as control valve choke in well and platform level and well re-completion are considered (Gaspar et al. 2016).

It is noteworthy that a proper reduction scenario method can be different along the phases of the process. For instance, the selected models that are most suitable for the well placement optimization during the field development plan (phase 4) may not be appropriate models for the well control optimization (with the fixed well locations) during

Fig. 1 Phases of the decision-making process (Schiozer et al. 2019)

field management decision (phase 5), since these procedures are very sensitive to different geological factors (Shirangi and Durlowsky 2016). Accordingly, comparative studies are essential to allow the user to select the most suitable and precise technique according to the different objectives and contexts (Mahjour et al. 2020a).

In this work, we compare two scenario reduction techniques: (1) Distance-based Clustering with Simple Matching Coefficient (DCSMC) as a data-mining approach applied before the simulation process using reservoir static data, (Mahjour et al. 2020a), and (2) metaheuristic algorithm (RMFinder technique) as an optimization-based approach applied after the simulation process using reservoir dynamic (simulation based) data (Meira et al. 2020). We used these two methods as samples to investigate the effect of static and dynamic data usage on the accuracy and rate of the scenario reduction process focusing field development purposes. Hence, other scenario reduction techniques which are working with dynamic and static data can be considered in this comparative work. To do so, we perform a set of statistical experiments to study and evaluate the outcomes of the apply methods. We apply this study to a synthetic fractured benchmark case named UNISIM-II-D, considering the Flow Unit (FU) modelling approach (Mahjour et al. 2019).

Methodology

The methodology consists in studying: (1) Distance-based Clustering with Simple Matching Coefficient (DCSMC) method Fig. 2a; and (2) metaheuristic algorithm (RMFinder technique) to reduce the scenarios for the field development process Fig. 2b. Towards this end, we provide a brief overview of each method and then introduce the comparison experiments used in this study. Each method is used to a set of static (DCSMC method) and dynamic (metaheuristic algorithm) attributes for the field development process.

Distance-based clustering with simple matching coefficient (DCSMC) method

Figure 2a presents the workflow to apply the Distance-based Clustering with Simple Matching Coefficient (DCSMC) method, as a deterministic method, for scenario reduction in the field development process (workflow A). The scenario reduction step is shown by the red hexagon shape in the flow chart.

In step 1, a set of “m” geostatistical models are generated under geological uncertainties. To do so, we employ a probabilistic geostatistical approach named Latin Hypercube

Fig. 2 field development based on **a** DCSMC method-Workflow A, **b** RMFinder method-Workflow B

Fig. 3 An example of model smoothing: **a** before smoothing and **b** after smoothing

Sampling (LHS) for uncertainty quantification based on multi-source data (Mahjour et al. 2020b). Note that this step is not the main part of this work and is independent of the technique used to reduce the scenarios. Additional details of this step can be found in (Mahjour et al. 2019) and (Correia et al. 2015). Next, the DCSMC method (as one of the main focuses of this study) is used for the geostatistical models reduction. The outlines of the method are described below (Mahjour et al. 2020a):

- 1 Calculate similarity distance: The similarity distance of any pair of geostatistical models is calculated. To this end, the 3D facies, or flow unit models, are firstly smoothed to eliminate short-scale dispersed data. In some cases, such variations can change the flow response and predicted reserves; therefore, smoothing algorithms are needed (Mahjour et al. 2020a). Note that during the smoothing process, the reference distribution must be preserved. Different smoothing algorithms can be used for this goal. Here, we apply median filtering to remove the noise maintaining the main patterns Fig. 3. Median filtering smooths the short-scale grid cells whose values differ notably from their surrounding grid cells without impacting on other grid cells (Mahjour et al. 2020a). Next, the “m” 3D smoothed models with “n” grid cells are converted into the “m” 1D arrays in a way that a matrix is built in that each column contains a model, and each row represents a categorical value of the variable in the grid cell Fig. 4. Ultimately, the similarity index between any two 1D models is calculated by the Simple Matching Coefficient (SMC) method (Sokal and Michener 1958) as Eq. 1:

$$\text{SMC} = \frac{\text{number of matching grid cells between two models}}{\text{number of total grid cells for a model}} \quad (1)$$

Fig. 4 The “m” 3D models with “n” cells converted into “m” 1D arrays

Models	1	2	Matching attributes
Grid cells			
1	1	1	Match
2	2	2	Match
3	2	4	Mismatch
4	3	2	Mismatch
5	3	3	Match
6	1	1	Match
7	4	3	Mismatch

Fig. 5 An example of SMC method

where SMC is the number of grid cells that are matched between two models divided by the total number of grid cells. If SMC equals 1, the models are the same, whereas if SMC equals 0, the models are completely different. Figure 5 represents an example of the SMC method. Based on Eq. 1, the matching grid cell number is 4 and the total grid cell number is 7; therefore, the SMC value between these two models is 0.57.

2. Generate dissimilarity matrix: After defining the similarity index between any two 1D models, a “m × m” dissimilarity matrix “d” is built given the “m” num-

ber of models and the dissimilarity distance index S_{ij} ($S_{ij} = 1 - SMC$) between model i and model j . Figure 6 represents an example of distance matrix for four models. The Multi-Dimensional Scaling (MDS) method is then implemented using the matrix “d”, as an input data, to map high-dimensional data into the 2D Euclidean space R (Borg and Groenen 2005). Each point in the 2D Euclidean space indicates each model. Figure 7 shows an example of 2D Euclidean space that is created by 21×21 distance matrix. In the 2D Euclidean space, the distance calculated between two points is considered as a similarity index in a way that: the closer the points, more similar the models will be (Haghighat Sefat et al. 2016).

- Classify models: As the last step of DCSMC method, similar models (the points with the closest distance to each other in the 2D Euclidean space) are classified into different groups via clustering algorithms. In this procedure, the number of groups can be defined by the user (Armstrong et al. 2013). The number of clusters can be adjusted based on further analysis by uncertainty quantification. Eventually, a single model from each cluster is chosen as a representative geostatistical model using the centroid sampling method. Figure 8 represents an example of the clustering and sampling procedures to select four representative models from 24 models.

After selecting the representative geostatistical models, the Discretized Latin Hypercube sampling with Geostatistical realizations (DLHG) technique (Schiozer et al. 2016) is applied to: (1) combine other types of uncertainty with the selected subset of geostatistical models, and (2) generate simulation models during step 4 (Flow simulation of scenarios). More details regarding DLHG technique are presented by (Schiozer et al. 2016). Then, all simulation models are simulated and the models are prepared for the data assimilation step. To carry out the data assimilation in step 5, several

Fig. 6 Example of distance matrix for four models (4×4 matrix)

Fig. 7 An Example of 2D Euclidean space with 21 models: An example of model smoothing: **a** before smoothing and **b** after smoothing

methods are available, according to the complexity of data (Costa et al. 2018; Davolio and Schiozer 2018; Maschio and Schiozer 2016; Oliveira et al. 2018; Mahjour et al. 2020c). During this step, a subset of generated simulation models is selected based on the past reservoir performance (Gaspar et al. 2016). In the end, the subset of scenarios can be used for the field development purposes.

According to the workflow, we observe that the scenario reduction is applied over the geostatistical models before the flow simulation and a deterministic RM set is generated.

Metaheuristic algorithm (RMfinder technique)

Figure 2b shows the workflow to apply the metaheuristic algorithm (RMfinder method) as a probabilistic technique for field development process (workflow B). The red hexagon shape in the flow chart shows the scenario reduction step.

Fig. 8 Example of clustering and sampling procedures. Each oval shows a cluster, and each red dot illustrates a representative model of each cluster

In the first step, “m” geostatistical models are generated under geological uncertainties. In step 2, we apply the DLHG sampling to integrate other types of uncertainty with all generated geostatistical models and create simulation models. Then, in steps 3 and 4, we simulate the models and select a subset of these models during the data assimilation step, respectively. In step 5, the RMFinder method is applied to decrease the number of approved simulation models for the field development goals (Meira et al. 2020). A number of visualization techniques are employed simultaneously in RMFinder method to check the representativeness of the models using the weight of the dynamic features (NPV, ORF, oil production (NP), water production (WP), etc.). The weight of dynamic features is defined by the users’ experience, based on the relative importance of each variable. For instance, if in some study the representativeness of the NPV is more relevant and has more priority in comparison with other dynamic parameters, the weight of NPV will be greater than the others. NPV is calculated by Eq. 2 (Gaspar et al. 2016):

$$NPV = \sum_{k=1}^{n_t} \frac{NCF_{t_k}}{(1+r)^{t_k}} \quad (2)$$

where NCF_{t_k} is the net cash flow at time t_k , k is the time index, n_t is the total number of time periods, r is the rate of discount, and t_k is the time in the step k considering the date of analysis $t=0$.

In this study, we determine NCF for each period by Eq. 3 according to the Brazilian Royalty & Tax fiscal regime:

$$NCF = [(R - Roy - ST - OC) \times (1 - T)] - I - AC \quad (3)$$

where R is gross revenues obtained from oil and gas sales, Roy is the total amount of royalties, ST is the total amount of Social Taxes, OC is the costs of operational production, T is the rate of corporate tax, I is the investments in equipment and facilities, and AC is the abandonment cost.

One of the visualization techniques used in RMFinder technique is the cross-plot of dynamic data obtained from the simulation process. The normalized NPV, ORF, NP, and WP are considered to be important outputs of the simulator to make the cross-plot curves (NPV vs ORF and NP vs WP), but others can also be applied, as for example Oil in Place (OIP). Figure 9 represents the high- and low-quality of RM sets for cross-plot NPV vs ORF. A high-quality RM set must be well spread over all the points in each cross-plot Fig. 9a. Instead, a low-quality RM set is not well spread over all points Fig. 9b.

Another visualization method is the Cumulative Distribution Function (CDF) curve (or risk curve) of objective functions. The CDF curve is used to check and evaluate the distribution of data sets (Kaplan and Garrick 1981). The goal is to obtain the CDF curves of RMs as similar to the CDF curves of reference (full-set). Figure 10 represents two sets of RMs with the similar and dissimilar CDFs of the reference (full-set) based on NPV data. A high-quality set of RMs must have a close distribution to the reference.

If the generated RMs are not properly representing the reference models in a cross-plot and risk curve, the weight related to such a cross-plot or risk curve must be increased. According to the workflow, we observe that the scenario reduction is applied on the simulation models after the numerical simulation and a probabilistic RM set is generated to automatically balance each individual criteria.

Eventually, in step 6, each obtained subset of scenarios from RMFinder method can be used for the field development purposes.

Comparison experiments

We compare both methodologies in four steps:

1. The accuracy of the results is assessed by comparing the Cumulative Distribution Function (CDF) curve of RMs and reference (full-set) for the objective functions NPV, ORF,

Fig. 9 High-quality (left) and low-quality (right) RM sets according to the cross-plot of NP vs WP. A high-quality set of RMs must be well distributed over all the points

Fig. 10 Similar (left) and dissimilar (right) RM sets to the reference sets according to the CDF curves of NPV. A high-quality set of RM must have a close distribution to the reference

Fig. 11 A flow unit map of UNISIM-II-D flow unit-based model

2. The cross-plot of RMs and reference is evaluated for the objective functions NPV vs ORF and NP vs WP,
3. The dispersion of the RMs in Q_o and Q_w curves is checked by visual inspection,
4. We compare the computational time of each field development process performed by different scenario reduction methods.

Application

The study is applied to a synthetic fracture model named the UNISIM-II-D flow unit-based model. Flow units divide reservoir into layers for simulation studies (Mahjour et al. 2019). Mahjour et al. (2019) fully explained how the model was built. Having a known answer, the synthetic model is preferred since the aim is to compare different approaches, not study the answer itself. The synthetic model shows the complexity of a real fracture reservoir (Correia et al. 2015). The UNISIM-II-D flow unit-based model is made for the primary stage of the field-development phase with 516 days of initial production history for one vertical producing well (Fig. 11). The simulation model includes a corner point grid with an average size of $100 \times 100 \times 8$ m, in roughly 95,000

grid cells (41,000 active cells). To generate the geostatistical models, we consider the geological uncertainty variables mentioned in Table 1. Other types of uncertainty applied to generate the simulation models are shown in Table 2. The probability of incidence for each uncertainty level is identical. We define an initial production strategy (strategy D) to generate the objective functions. This strategy is simulated by using dual-porosity/dual-permeability approach with 11 vertical production wells and 9 injector wells, as shown in Fig. 12. All vertical wells are drilled from the first to the last layer of the reservoir, and all horizontal wells are drilled in different layers of the reservoir. The final simulation time is of 10,957 days, and the well's operational conditions and platform data for production are illustrated in Tables 3 and 4, respectively. Table 5 represents the deterministic most-probable economic parameters used to calculate NPV.

Results

The results described in the following subsections are based on generated objective functions from each scenario reduction method.

Quality check of the obtained RMs from the DCSMC and RMFinder methods based on strategy D

Figure 13 depicts the obtained number of generated models from each step of both workflows. In the scheme figure and the first workflow (based on the DCSMC method), 18 representative geostatistical models are selected from the 200 available models (9% of the total geostatistical models), while in the second workflow, the RMFinder tool is used to select 16 representative simulation models from 182 approved models (9% of the total approved models from data assimilation step). Note that selecting the number of representative geostatistical and simulation models depends

Table 1 Geological uncertainty elements to build the geostatistical models

Property	Attribute	Probability distribution	Geostatistical method
Horizons	Height	Normal	Minimum curvature
Flow units (six flow units except flow unit 7)	Stochastic seed	Seed variable	Sequential indicator simulation (SIS)
	Spatial variability	Normal	
Flow unit 7 (Super-K)	Stochastic seed	Seed variable	Object modelling
	Height	Normal	
Porosity	Stochastic seed	Seed variable	Sequential Gaussian Simulation (SGS)
	Average	Normal	
	Spatial variability	Normal	
Permeability	Stochastic seed	Seed variable	
	Correlation factor with porosity	Lognormal	
	Average	Lognormal	
Discrete fractures	Aperture	Lognormal	Stochastic
	Intensity	Normal	
	Length	Normal	
	Azimuth	Normal	
	Dip	Normal	

Table 2 All types of uncertainties to build the simulation models

Properties	Levels
Geological properties	200 geostatistical models (all models)
K _{rw} for flow unit 7 (SK), unit less	6 levels
K _{rw} for other flow units, unit less	10 levels
PVT	3 levels
C _{por} , (106 kgf/cm ²⁻¹) ⁻¹	3 levels
K _z multiplier, unit less	20 levels [0.2, 10]
K _x and K _y multiplier, unit less	20 levels [0.2, 10]
POR multiplier, unit less	10 levels [0.9, 1.1]

on what uncertainty space the user wants (Mahjour et al. 2020a).

In both workflows, the ratio of the number of simulation models and geostatistical models during the DLHG sampling step is three (54 simulation models are generated from 18 geostatistical models during the DCSMS process, and 600 simulation models are generated from 200 geostatistical models during the RMFinder process). (Mahjour et al. 2020a) proved the number ratio three between simulation and geostatistical models is sufficient to cover the uncertainty range.

In this part, we check the quality of obtained RMs from the tools using the comparison experiments. We initially applied strategy D to obtain the defined objective functions. According to the DCSMC method, the statistical data of objective functions for 18 representative geostatistical

Fig. 12 Initial strategy (strategy D) with 11 vertical production wells and 9 horizontal injection wells

models should be compared with the total 200 models (reference model). Hence, we only consider the geological uncertainties (Table 1) to generate the simulation models and obtain the objective functions. We evaluate the accuracy of the DCSMC method to select suitable representative geostatistical models, and we can ignore the role of dynamic uncertainties in this analysis. Based on RMFinder tool, the

Table 3 Well data

Type	Producer well	Injector well	Unit
Water rate	–	Max 5000	m3/day
Liquid rate	Max 3000	–	m3/day
Gas lift rate	Max 240,000	–	m3/day
BHP	Min 275	Max 480	Kgf/cm2
Radius	0.108	0.108	m
Geometric factor	0.37	0.37	–
Angular well fraction	1	1	–
Well skin factor	0	0	–

Table 4 Platform data

Type	Platform pro- duction	Platform injec- tion	Unit *10 ³
Max water rate	120	240	bbl/day
Max liquid rate	180	–	bbl/day
Max oil rate	180	–	bbl/day
Max gas rate	8000	8000	m3/day

distribution of objective functions for 16 representative simulation models should be compared with the total of 182 approved models from the data assimilation step.

Figures. 14 and 15 show the CDF curve obtained from the DCSMC and RMFinder methods, respectively. They are plotted for the objective functions of NPV and ORF. In the

figures, we observe that the representative models have similar characteristics to the reference set of models.

Analysing the distribution of RMs in the cross-plots of NPV vs ORF and NP vs WP for both methods (Figures. 16 and 17), it is clear that the RMs are well-distributed over all the models in each cross-plot.

Figures. 18 and 19 show the oil and water rate curves based on the DCSMC and RMFinder methods, respectively. In the figures, the RMs and the full-set of models are checked, showing that the uncertainty spread range is well captured by the RMs.

From the above comparison experiments, we show that both methods have appropriate functions to select RMs based on strategy D.

Checking the representativeness of the selected RMs in other strategies

In this part, we check that the obtained RMs from strategy D can still maintain the representativeness of the models if the strategy is changed. To analyse the representativeness of the RMs, different production strategies (different number of wells and different locations) are defined considering the heterogeneity of the reservoir. Figure 20 shows the characteristics of each strategy below:

- 1- Strategy A (vertical five-spot pattern): 56 vertical wells that are completed from top to bottom of the reservoir (28 producers and 28 injectors),

Table 5 Deterministic economic variables

Parameters	Value	Unit
Oil price	257.9	USD/m ³
Gas price	0.026	USD/m ³
Cost of oil production	48.57	USD/m ³
Cost of gas production	0.013	USD/m ³
Cost of water production	4.86	USD/m ³
Cost of water injection	4.86	USD/m ³
Drilling and completion of horizontal well (fixed cost)	73.75	10 ⁶ USD
Drilling and completion of horizontal well (variable cost)	0.032	10 ⁶ USD/m
Horizontal well connection (well-platform)	13.30	10 ⁶ USD
Drilling of vertical well	23.40	10 ⁶ USD
Vertical well completion	26.94	10 ⁶ USD
Vertical well connection (well-platform)	13.30	10 ⁶ USD
Horizontal well re-completion	11.02	10 ⁶ USD
Vertical well re-completion	10.97	10 ⁶ USD
Well conversion	11.02	10 ⁶ USD
1st Inflow Control Valve (ICV) (for each well)	1.00	10 ⁶ USD
2nd or more ICV (for each well)	0.30	10 ⁶ USD/ICV
Investment percentage in drilling and completion	8.2%	–
Annual discount rate	9%	–

Fig. 13 Number of models generated from each step of the workflows

Fig. 14 CDF of NPV **a** and ORF **b** for all 200 models (dotted red line) and 18 representative models (solid blue line) obtained from the DCSMC method at 10,957 days

- 2- Strategy B (horizontal five-spot pattern): 56 horizontal wells that are completed in different layers of the reservoir (28 producers and 28 injectors),
- 3- Strategy C (optimized strategy): 17 vertical wells (9 producers and 7 injectors).

The prediction period, the well's operational conditions, and platform data for defining the strategies are the same as strategy D. The number of generated models is based on Fig. 13.

Figures. 21, 22, 23 show the different comparison experiments for the same obtained RMs from strategy D and the full-ensemble of models (reference), from other strategies

based on the DCSMC method. As can be seen from the figures, the RMs have similar characteristics to the reference set, meaning that the selected RMs from strategy D can maintain the representativeness of the total models in other defined strategies, given the defined objective functions. Regardless of the different strategies defined, the RMs obtained from one strategy can be used in other strategies and the type does not play a significant role on the selection of the RMs.

Figures. 24 - 26 represent the results of the experiments for the same obtained RMs from strategy D and the full-ensemble of models from other strategies based on the RMFinder method. According to the CDFs plotted in

Fig. 15 CDF of NPV **a** and ORF **b** for all 182 models (dotted red line) and 16 representative models (solid blue line) obtained from the RMFinder method for 10,957 days

Fig. 16 Cross-plots of NPV vs ORF **a** and NP vs WP **b** for all 200 models (yellow points) and 18 representative models (blue points) obtained from the DCSMC method for 10,957 days

Fig. 17 Cross-plots of NPV vs ORF **a** and NP vs WP **b** for all 182 models (yellow points) and 16 representative models (blue points) obtained from the RMFinder method for 10,957 days

Fig. 18 Oil rate curve **a** and water rate curve **b** for all 200 models (blue lines) and 18 representative models (red lines) obtained from the DCSMC method for 10,957 days, based on strategy D

Fig. 19 Oil rate curve **a** and water rate curve **b** for all 182 models (blue lines) and 16 representative models (red lines) obtained from the DCSMC method for 10,957 days based on strategy D

Fig. 20 Strategy A (vertical five-spot pattern), Strategy B (horizontal five-spot pattern) and Strategy C (optimized strategy)

Fig. 24, there are some mismatches between the reference sets and RM sets showing the distributions of these two sets are not identical for different strategies and different objective functions. In Fig. 25, it is clear that RM sets are not well distributed over all the points in each cross-plot. Figure 26 also presents the dispersions of the RMs in Q_o and Q_w curves are not the same for different strategies. As

a result, the RMs do not have similar characteristics to the reference set, meaning that the selected RMs from strategy D cannot preserve the representativeness of the full-set in other defined strategies given the objective functions. Hence, based on the RMFinder tool, it is necessary to select new RMs for each strategy.

Fig. 21 **a** CDFs of NPV and ORF for strategy A, **b** CDFs of NPV and ORF for strategy B, and **c** CDFs of NPV and ORF for strategy C. 200 models (dotted red line) and 18 representative models (solid blue

line) obtained from the DCSMC method. The RM set has sufficiently similar characteristics to the reference set

Comparison of the generated models from each step of the field development process based on different scenario reduction methods

Here, we compare the CDFs of NPV and ORF generated from each step of the field development workflow based on different scenario reduction methods (DCSMC in static feature-based workflow or workflow A, and RMFinder in dynamic feature-based workflow or workflow B). According to Fig. 27, we compare the model sets obtained from the workflows in three steps:

Step 1: The generated models from the flow simulation of scenarios step in both workflows (54 models from workflow A and 600 models from workflow B),

Step 2: The generated models from the data assimilation step in both workflows (16 models from workflow A and 182 models from workflow B),

Step 3: The generated models from the data assimilation step in workflow A (16 models) and the generated models from the scenario reduction step in workflow B (16 models).

Figures 28 and 29 represent that the CDFs of NPV and ORF obtained from all strategies for step 1, respectively. The results show that the generated models from workflow A (54 models) have sufficiently similar characteristics to those from workflow B (600 models) using the visual inspection. This means that the selected geostatistical models from the DCSMC method (18 models) that are used in the DLHG sampling step pose good representativeness for all geostatistical models (200 models) for all defined strategies (Strategies A, B, C, and D). Therefore, there is no need to consider all geostatistical models for the DLHG sampling process.

Figures 30 and 31 show the CDFs of NPV and ORF for step 2, respectively. The results indicate that the distributions

Fig. 22 **a** Cross-plots of NPV vs ORF and NP vs WP for strategy A, **b** cross-plots of NPV vs ORF and NP vs WP for strategy B, and **c** cross-plots of NPV vs ORF and NP vs WP for strategy C. 200 models

(yellow points) and 18 representative models (blue points) obtained from the DCSMC method. The RMs are well distributed over all the models in each cross-plot

Fig. 23 **a** Oil and water rate curves for strategy A, **b** oil and water rate curves for strategy B, and **c** oil and water rate curves for strategy C. 200 models (blue lines) and 18 representative models (red

lines) obtained from the DCSMC method. The uncertainty spread range is well captured by the RMs

Fig. 24 **a** CDFs of NPV and ORF for strategy A, **b** CDFs of NPV and ORF for strategy B, and **c** CDFs of NPV and ORF for strategy C. 182 models (dotted red line) and 16 representative models (solid

blue line) obtained from the RMFinder method. The RMs have slight differences and bias in characteristics when compared to the reference set

of the model sets obtained from the data assimilation steps in workflows A (16 models) and B (182 models) are close, which means that a small number of simulation models in workflow A are sufficient for the field development process and there is no need to reduce the number of simulation models.

Figures. 32 and 33 indicate that the CDFs of NPV and ORF for step 3, respectively. The results show there are slight mismatches in quantile estimates on the obtained CDFs from the data assimilation step in workflows A (16 models) and the scenario reduction step in workflow B (16 models). The mismatches can be different in each strategy, since the role of strategy types to select RMs by the RMFinder tool in workflow B is significant.

Comparison of the computational time of each workflow

Table 6 presents the computational time for each step of the field development workflows (workflows A and B), based on strategy D. To accelerate the simulation runs, a

parallel computing machine is used. As can be seen from the table, workflow A is twice as fast as the workflow B. Therefore, if we apply scenario reduction on the geostatistical models, before the simulation process and based on static features, we can avoid additional simulation run and excessive computation.

Discussion

In this section, we study and discuss the field development workflows based on different scenario reduction methods in terms of the precision of the results, computational time, and restriction of the workflows.

Precision of the results

The scenario reduction methods in both workflows proved reliable when applied to a unique production strategy. The statistical data of RMs obtained from both scenario

Fig. 25 **a** Cross-plots of NPV vs ORF and NP vs WP for strategy A, **b** cross-plots of NPV vs ORF and NP vs WP for strategy B, and **c** cross-plots of NPV vs ORF and NP vs WP for strategy C. 182 models

(yellow points) and 16 representative models (blue points) obtained from the RMFinder method. The RMs are congested and not well distributed in some part of each cross-plot

Fig. 26 **a** Oil and water rate curves for strategy A, **b** oil and water rate curves for strategy B, and **c** oil and water rate curves for strategy C. 182 models (blue lines) and 16 representative models (red lines)

obtained from the DCSMC method. The uncertainty spread range is not well captured by the RMs

Fig. 27 Comparison of the model sets obtained from each step of the field development workflows: static feature-based workflow (workflow A) uses the DCSMC method for scenario reduction and dynamic

feature-based workflow (workflow B) uses the RMFinder tool for scenario reduction

Fig. 28 CDFs of NPV generated from four strategies for step 1. 600 models obtained from the flow simulation step in workflow B (solid blue line) and 54 models obtained from the flow simulation step in workflow A (dotted orange line)

Fig. 29 CDFs of ORF generated from four strategies for step 1. A total of 600 models obtained from the flow simulation step in workflow B (solid blue line) and 54 models obtained from the flow simulation step in workflow A (dotted orange line)

reduction methods were similar to those from the full-set of models. However, by changing the production strategy, the RMFinder method did not record acceptable results and showed that the RMs obtained from one specific strategy cannot be used for other strategies. While an obtained set of RMs, from a specific strategy, using the DCSMC method, preserved the representativeness in other strategies. We also did not see the need to generate new RMs for each strategy. We also observed that the CDFs plotted from the flow simulation of scenarios and data assimilation steps have similar characteristics for both workflows.

Computational time

Although the scenario reduction methods in both workflows reduced the number of models and computational time, geostatistical model reduction using the DCSMC

method before the flow simulation step can be more effective than reducing the number of simulation models using the RMFinder tool, after the flow simulation step. Workflow A obtained the same number of the subsets as workflow B with substantially lower computational time, while the reliability of outcomes was uncompromised for a different strategy (Sect. 5–1).

Note that the DCSMC method is a deterministic approach, while the RMFinder tool is a probabilistic approach. Hence, the computational time difference between these two workflows can be greater if more RM sets are generated and tested by the RMfinder tool for a single strategy.

Fig. 30 CDFs of NPV generated from four strategies for step 2. 182 models obtained from the data assimilation step in workflow B (solid blue line) and 16 models obtained from the data assimilation step in workflow A (dotted green line)

For workflow B, we can decrease the computational time by reducing the number of simulation models during the DLHG step, but it can compromise the precision of the results; therefore, it is not recommended.

Restriction of the workflows

The results showed a limitation of the RMFinder tool when a single obtained RM set from a defined production strategy is applied to other production strategies. The computational time also increased by considering the full number

of geostatistical models and generating different RMs for each strategy, based on workflow B. The DCSMC method performed well, with the same dataset, and did not present restrictions during this work. However, this method is more suitable and effective for a small set of geostatistical models. Otherwise, the scenario reduction step in workflow A can be time-consuming. In this case, we recommend using, and testing, a combination of these two workflows so that the scenario reduction step is performed before and after the simulation process, as shown in Fig. 34 (red hexagon shape).

Fig. 31 CDFs of ORF generated from four strategies for step 2. 182 models obtained from the data assimilation step in workflow B (solid blue line) and 16 models obtained from the data assimilation step in workflow A (dotted green line)

Conclusions

In this work, we selected a subset of reservoir simulation models for the field development phase using (1) distance-based Clustering with Simple Matching Coefficient (DCSMC) based on static data (workflow A), (2) metaheuristic algorithm (RMFinder technique) dynamic data (workflow B). We compare these two methods as samples to check the effect of static and dynamic data usage on the

accuracy of results and rate of scenario reduction process. Our outcomes support the following conclusions:

1. The statistical experiments show that both scenario reduction methods are reliable for selecting RMs from a defined production strategy.
2. We can apply the obtained set of RMs from a specific production strategy, using the DCSMC method, to other different strategies and still preserve the representativeness of the models. While the role of strategy types in

Fig. 32 CDFs of NPV generated from four strategies for step 3. 16 models obtained from the data assimilation step in workflow A (dotted blue line) and 16 models obtained from the scenario reduction step in workflow B (solid grey line)

selecting the RMs, using the RMFinder technique, is significant so that each strategy has its own set of RMs.

3. The CDFs plotted from the flow simulation of scenarios, and data assimilation steps for both field development workflows have similar characteristics. Hence, a number of suitable selected geostatistical models can be used in simulation process instead of using the full-set of geostatistical models.
4. The computational time in workflow B is greater than A because of three reasons: 1—in workflow B, the full-set of geostatistical models is applied for the flow simulation step, which results in increasing the number of simulation models and simulation time, 2—for each strategy in workflow B, the scenario reduction procedure should be repeated to generate RMs while in work-

flow A, the obtained RMs from a single strategy can be applied to other different strategies, 3—the RMFinder tool is a probabilistic approach for selecting the RMs; therefore, the computational time can be excessive for analysing the results if different RM sets are generated for a single strategy. The DCSMC method is a deterministic approach to select the RMs, and only one RM set is generated for a single production strategy.

From the above results, it can be concluded that reservoir static data usage on the scenario reduction process can be more reliable during the field development phase.

For future studies, we recommend to (1) apply and test the combination of workflows A and B for field development purposes (using dynamic and static features), (2) use

Fig. 33 CDFs of ORF generated from four strategies for step 3. 16 models obtained from the data assimilation step in workflow A (dotted blue line) and 16 models obtained from the scenario reduction step in workflow B (solid grey line)

Table 6 Computational time for each step of field development workflows

	Workflow A	Workflow B
Geostatistical model generation under uncertainty	7 h 25 min	7 h 25 min
Scenario reduction (workflow A)	5 h 25 min	–
DLHG sampling	1 min	1 min
Flow simulation of scenarios	4 min	45 min
Data assimilation	1 min	12 min
Scenario reduction (workflow B)	–	20 h 26 min
Total	12 h 56 min	28 h 24 min

and compare other scenario reduction methods in the literature, (3) study and compare the scenario reduction methodologies for field management objectives (step 5 of the decision-making process, as shown in Fig. 1) such as well control optimization with fixed well locations, and (4), use a proxy model instead of the simulation model and compare the decision-making results in terms of computational costs and accuracy.

Fig. 34 Workflow to apply the DCSMC and the RMFinder methods

Acknowledgements We gratefully acknowledge the support of EPIC (Energy Production Innovation Center), hosted by the University of Campinas and sponsored by FAPESP – São Paulo Research Foundation (grant number 2017/15736-3), Equinor Brazil, PETROBRAS, ANP (Brazil's National Agency of Oil, Natural Gas and Biofuels), UNISIM and Energi Simulation. The authors are also grateful to Schlumberger Information Solution for the use of Petrel® and to Computing Modeling Group for the use of Builder CMG®.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alzraiee A, Luis A, Garcia MASCE (2012) Using cluster analysis of hydraulic conductivity realizations to reduce computational time for monte carlo simulations. *J Irrig Drain* 138(5):416–423. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000416](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000416)
- Armstrong M, Ndiaye A, Razanatsimba R, Galli A (2013) Scenario reduction applied to geostatistical simulations. *Math Geosci* 45(2):165–182. <https://doi.org/10.1007/s11004-012-9420-7>
- Borg, I., and Groenen, P., (2005). Modern multidimensional scaling: theory and applications. Springer series in statistics. Springer. URL <http://books.google.com/books?id=duTODIdZzReC>.
- Caers, J., and Park, K. (2008). A distance-based representation of reservoir uncertainty: The Metric EnKF. In: 11th European Conference on the Mathematics of Oil Recovery (ECMOR XI). EAGE. Bergen, Norway, 8 - 11 September.
- Correia, M.G., Hohendorf, J., Gaspar, A.T.F.S., and Schiozer, D. (2015). UNISIM-II-D: Benchmark case proposal based on a carbonate reservoir. SPE Latin American and Caribbean Petroleum Engineering Conference held in Quito, Ecuador, 18–20 November. SPE-177140-MS.
- Costa LAN, Maschio C, Schiozer DJ (2018) A new methodology to reduce uncertainty of global attributes in naturally fractured reservoirs. *Oil Gas Sci Technol Rev IFP Energies nouvelles* 73:41. <https://doi.org/10.2516/ogst/2018038>
- Davolio A, Schiozer DJ (2018) Probabilistic seismic history matching using binary images. *J Geophys Eng* 15(1):261–274. <https://doi.org/10.1088/1742-2140/aa99f4>
- Gaspar ATFS, Barreto CEAG, Schiozer DJ (2016) Assisted process for design optimization of oil exploitation strategy. *J Petrol Sci Eng* 146:473–488. <https://doi.org/10.1016/j.petrol.2016.05.042>
- Haghighat Sefat M, Elsheikh AH, Muradov KM, Davies DR (2016) Reservoir uncertainty tolerant, proactive control of intelligent wells. *Comput Geosci* 20(3):655–676. <https://doi.org/10.1007/s10596-015-9513-8>
- Kang B, Kim S, Jung H, Choe J, Lee K (2019) Efficient assessment of reservoir uncertainty using distance-based clustering: a Review. *Energies* 12(10):1859. <https://doi.org/10.3390/en12101859>
- Kaplan S, Garrick BJ (1981) On the quantitative definition of risk. *Risk Anal* 1(1):11–27. <https://doi.org/10.1111/j.1539-6924.1981.tb01350.x>
- Mahjour SK, Correia MG, Santos AAS, Schiozer DJ (2019) Developing a workflow to represent fractured carbonate reservoirs for simulation models under uncertainties based on flow unit concept. *Oil Gas Sci Technol Rev IFP Energies nouvelles* 74:15. <https://doi.org/10.2516/ogst/2018096>
- Mahjour SK, Santos AAS, Correia MG, Schiozer DJ (2020a) Developing a workflow to select representative reservoir models combining distance based clustering and data assimilation for decision making process. *J Petrol Sci Eng* 190(2020):107078. <https://doi.org/10.1016/j.petrol.2020.107078>
- Mahjour SK, Correia MG, de Souza dos Santos AA, Schiozer, DJ, (2020) Using an integrated multidimensional scaling and clustering method to reduce the number of scenarios based on flow-unit models under geological uncertainties. *ASME J Energy Res Technol* 142(6):063005. <https://doi.org/10.1115/1.4045736>
- Mahjour, S.K., Santos, A.A.S., Correia, M.G., and Schiozer, D.J. (2020). Two-stage Scenario Reduction Process for an Efficient Robust Optimization. ECMOR XVII – 17th European Conference on the Mathematics of Oil Recovery, Online Event, 14–17 September. <https://www.earthdoc.org/content/serial/2214-4609>.
- Maschio C, Schiozer DJ (2016) Probabilistic history matching using discrete latin hypercube sampling and nonparametric density estimation. *J Petrol Sci Eng* 147:98–115. <https://doi.org/10.1016/j.petrol.2016.05.011>
- Meira LA, Coelho GP, Santos AAS, Schiozer DJ (2016) Selection of representative models for decision analysis under uncertainty. *Comput Geosci* 88:67–82. <https://doi.org/10.1016/j.cageo.2015.11.012>
- Meira LA, Coelho GP, Silva CG, Abreu JAL, Santos AAS, Schiozer DJ (2020) Improving representativeness in a scenario reduction process to aid decision making in petroleum fields. *J Petrol Sci Eng* 184(2020):106398. <https://doi.org/10.1016/j.petrol.2019.106398>
- Oliveira GS, Maschio C, Schiozer DJ (2018) A new approach with multiple realizations for image perturbation using co-simulation and probability perturbation method. *Oil Gas Sci Technol Rev IFP Energies nouvelles* 73:68. <https://doi.org/10.2516/ogst/2018065>
- Park, K., and Caers, J. (2007). History matching in low-dimensional connectivity-vector space. EAGE Petroleum Geostatistics Conference, 10–14 September, Cascais, Portugal
- Santos SMG, Gaspar ATFS, Schiozer DJ (2018) Managing reservoir uncertainty in petroleum field development: defining a flexible production strategy from a set of rigid candidate strategies. *J Petrol Sci Eng* 171:516–528. <https://doi.org/10.1016/j.petrol.2018.07.048>
- Sarma P, Chen WH, Xie J (2013) Selecting representative models from a large set of models. SPE Reservoir Simulation Symposium held in The Woodlands, Texas, USA, pp 18–20
- Scheidt C, Caers J (2009) Representing spatial uncertainty using distances and Kernels. *Math Geosci* 41(4):397–419. <https://doi.org/10.1007/s11004-008-9186-0>
- Schiozer D, Ligero E, Suslick S, Costa A, Santos J (2004) Use of representative models in the integration of risk analysis and production strategy definition. *J Petrol Sci Eng* 44(1–2):131–141. <https://doi.org/10.1016/j.petrol.2004.02.010>
- Schiozer DJ, Avansi GD, Santos AAS (2016) Risk quantification combining geostatistical realizations and discretized latin hypercube. *J Braz Soc Mech Sci Eng* 39(2):575–587. <https://doi.org/10.1007/s40430-016-0576-9>
- Schiozer DJ, Santos AAS, Santos SMG, Filho JCH (2019) Model-based decision analysis applied to petroleum field development and management. *Oil Gas Sci Technol Rev IFP Energies* 74:46. <https://doi.org/10.2516/ogst/2019019>
- Shirangi MG, Durlafsky LJ (2016) A General method to select representative models for decision making and optimization under

- uncertainty. *Comput Geosci* 96:109–123. <https://doi.org/10.1016/j.cageo.2016.08.002>
- Shirangi, M.G., and Mukerji, T. (2012). Retrospective Optimization of well controls under uncertainty using kernel clustering. Technical Report Presented at the 25th Annual Meeting of the Stanford Center for Reservoir Forecasting, Monterey, CA. <https://pangea.stanford.edu/researchgroups/scrf/resources/publications>.
- Sokal RR, Michener CD (1958) A statistical methods for evaluating relationships. *Sci Res* 38:1409–1448
- Suzuki, S., and Caers, J. (2006). History matching with an uncertain geological scenario. SPE Annual Technical Conference and Exhibition, 24–27 September, San Antonio, Texas, USA. <https://doi.org/10.2118/102154-MS>.
- Torrado RR, Echeverría-Ciaurri D, Mello U, Embid Droz S (2015) Opening new opportunities with fast reservoir-performance evaluation under uncertainty: brugge field case study. *SPE Econom Manag* 7(3):84–99. <https://doi.org/10.2118/166392-PA>
- Yasari E, Pishvaie MR, Khorasheh F, Salahshoor K, Kharrat R (2013) Application of multi-criterion robust optimization in water-flooding of oil reservoir. *J Petrol Sci Eng* 109:1–11. <https://doi.org/10.1016/j.petrol.2013.07.008>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.