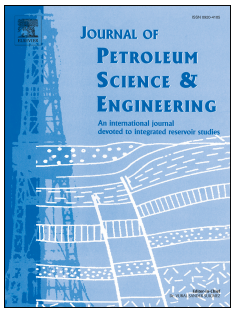
Accepted Manuscript



Emulation of reservoir production forecast considering variation in petrophysical properties

R. Moreno, G. Avansi, D. Schiozer, I. Vernon, M. Goldstein, C. Caiado

PII: S0920-4105(18)30162-1

DOI: 10.1016/j.petrol.2018.02.056

Reference: PETROL 4726

To appear in: *Journal of Petroleum Science and Engineering*

Received Date: 7 June 2017

Revised Date: 25 January 2018

Accepted Date: 23 February 2018

Please cite this article as: Moreno, R., Avansi, G., Schiozer, D., Vernon, I., Goldstein, M., Caiado, C., Emulation of reservoir production forecast considering variation in petrophysical properties, *Journal of Petroleum Science and Engineering* (2018), doi: 10.1016/j.petrol.2018.02.056.

Emulation of Reservoir Production Forecast Considering Variation in Petrophysical Properties

Moreno, R. [1,*]; Avansi, G. [1]; Schiozer, D. [1]; Vernon, I. [2]; Goldstein, M. [2]; Caiado, C. [2]

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

[2] Department of Mathematical Sciences, Durham University, United Kingdom

[*] corresponding author: rahdezm@fem.unicamp.br

Submitted to: Journal of Petroleum Science and Engineering

Abstract

Implementation of proxy models, such as emulators might reduce the computational time required in a variety of reservoir simulation studies. By definition, an emulator uses reservoir properties as input parameters in a statistical model constructed from simulator outputs. However, incorporation of petrophysical properties distributions in all model grid-blocks implies too many input parameters for direct emulation. Currently, most employments of emulation only consider single-value parameterization of reservoir properties.

In this work, we propose a methodology to consider spatially-distributed properties, such as porosity and permeability, in reservoir emulation technique. First, we present the process of finding a procedure to deal with geostatistical realizations in the emulator and then implement it in a risk quantification application. Construction of an emulator in a probabilistic approach involved: selection of a base model, definition of uncertain inputs, selection of outputs to be emulated, sampling inputs to generate scenarios, simulation of scenarios, and building the emulator. As an application, we used emulators to generate risk curves at the final production time of a synthetic reservoir model.

By implementing the proposed procedure, we showed that emulators can provide reliable results during risk analysis in oilfield development. Furthermore, with emulators it is possible to generate risk curves that reproduce simulations results at a lower computational cost.

It can be expected that parameterization of petrophysical properties will boost the applicability of the reservoir emulation technique. For instance, emulators can significantly reduce both the time and computational resources demanded in various reservoir studies for high heterogeneity and complex reservoir models such as found in the Brazilian pre-salt area.

Keywords: Risk, Petrophysical uncertainty, Proxy model, Reservoir, Simulation.

1. Introduction

During the initial stage of oilfield development, as described by Schiozer et al. (2015), a reservoir characterization under uncertainties is required to build possible scenarios. Reservoir

petrophysical properties distributions are among the numerous features that must be described at this point.

From well, core and seismic data it is possible to model spatial distributions for properties like porosity and permeability, which constitute the reservoir numerical model. So, under uncertainties, and, in a probabilistic approach, several geostatistical realizations are possible for a reservoir model. Depending on the purpose of the study, we can generate from hundreds to thousands of equiprobable geo-realizations. Combinations of these realizations with other structural, technical and/or economic uncertainties compose the different reservoir model scenarios.

This inherent uncertainty about reservoir features and behavior translates into a necessity of quantifying the associated risk to this lack of knowledge. Among the available tools for risk appraisal we have production risk curves. In petroleum studies context, these curves might correspond to cumulative oil, gas, or water, prospect net-present-value, among other objective functions.

For a thorough generation process of risk curves, the uncertain solution space must be covered with a representative sample of all possible reservoir scenarios. Depending on the complexity of the model and available computational resources, reservoir studies that implement the numerical simulator can demand an excessive computational effort and CPU time, i.e., the amount of time used for processing reservoir numerical models.

Among the alternatives to circumvent this issue we find: (1) simplifications and variations of the statistical treatment (Schiozer et al., 2016), (2) sophisticated selection of representative models (Meira et al., 2015) and (3) use of low fidelity models such as proxy models (Zubarev, 2009).

Proxies, also known as surrogates, are mathematical representations (e.g. regression, kriging, neural networks, Bayesian emulators etc.) that try to mimic reservoir numerical simulator outputs at a lower computational cost. The inputs of a proxy model are reservoir model attributes and its outputs can be observables such as fluid production rates, bottom-hole pressures, fluid saturation, pressure distributions and so forth.

Therefore, as a substitute of the simulator that can be used to survey the uncertain space, proxy models might be applied in diverse applications within reservoir studies such as history matching (Craig et al., 1996), sensitivity analysis (Cullick et al., 2006), uncertainty assessment (Slotte et al., 2008; Mohaghegh et al., 2006), production strategy selection (Avansi et al., 2009), production forecasting and risk analysis (Amorim et al., 2012; Polizel et al., 2017).

Furthermore, given the role of uncertainty in reservoir studies, the Bayesian framework represents a natural approach in the proxy-building context (Craig et al., 1996; Cumming et al., 2009). Some previous works in petroleum studies have been carried out involving reservoir Bayesian emulation. For instance, Cumming and Goldstein (2009) used emulation technique to

history-match reservoir models, which were generated by parameterizing reservoir properties maps with multipliers. Ferreira et al. (2014) used emulators in uncertainty reduction quantification given availability of production data. Later, Ferreira et al. (2015) showed a methodology to use 4D seismic data to improve uncertainty reduction by using emulation of water saturation maps.

These works demonstrate the applicability of emulation but they are characterized by single-value parameterizations of reservoir properties. For instance, Cumming and Goldstein (2009) accounted for porosity and permeability maps by using multipliers in pre-defined regions. In fact, most of employments of proxy models (Cullick et al., 2006; Slotte et al., 2008; Zubarev, 2009; He et al., 2016) have been restrained to single-value parameterizations of spatially-distributed properties. As noticed by Mohaghegh et al. (2006), this restriction of proxy models is mainly due to "curse of dimensionality" given the high number of parameters that define a reservoir geological model. Besides, single-value parameterizations do not preserve geological consistency (spatial covariance model) required in a thorough treatment of petrophysical uncertainty (Chambers et al., 2000). An attempt to solve the issue was proposed by Zabalza-Mezghani et al., (2004). They introduced a joint-model method (JMM) that combines geo-realizations and proxy-models to account for geological uncertainty in computationally-expensive applications such as risk analysis. As shown by Santos et. al, (2017), implementation of JMM is difficult for complex cases and present technical and practical disadvantages when compared with other methods such as DLHG proposed by Schiozer et al., (2016). Discretized Latin Hypercube combined with geo-realizations (DLHG), represents well the treatment of geological uncertainty and reduces the computational cost in some reservoir studies. Still, because of their low computational cost, proxy-models show promise in applications where evaluation of a high number of reservoir scenarios is required.

Geostatistical uncertainty, represented by geo-realizations, is not trivial to be consistently captured by single-value parameterizations. Also, using property values at each grid cell as inputs in the proxy construction is unfeasible because of the high number of blocks of a typical model. Thus, there is a need to pre-process reservoir properties distributions to be considered as inputs in emulation procedure. This would allow dealing with petrophysical uncertainty in a variety of reservoir studies where emulation can be implemented and computational and human effort might be reduced.

2. Objective

The main goal of this work is to present a procedure that considers uncertainty of spatially-distributed reservoir properties, such as porosity and permeability, in emulation of reservoir model behavior.

Besides, we build emulators for chosen objective functions with different sizes of training dataset and then generate production risk curves to compare with simulation results. Based on those results, we establish quality criteria to evaluate emulators that can reproduce risk curves obtained with simulation.

Finally, we assess the implementation of emulator in risk analysis in terms of error and total computational cost in comparison with the simulator.

3. Methodology

The work proposal concerns the incorporation of petrophysical uncertainty, represented by geo-realizations, as inputs in the building of emulators. Several attempts were made to solve this issue along the development of this research. The main difficulties rely on the high number of parameters that define a realization and the non-trivial relationship between the set of petrophysical properties at each grid-block and well responses. For a typical simulation model, realizations are characterized by the values of porosity, permeability in the three spatial directions and net-to-gross ratio (NTG). On the other hand, responses of a given well may depend upon the characteristics of its region of influence along the production period and this dependency can be difficult to describe in mathematical terms.

To overcome these challenges and design a procedure that allowed us to build and validate emulators from realizations, we tested combinations of division of reservoir by zones and selection of grid points (random, evenly spaced and dimension reduction by Principal Variables (PV)).

At the end, the procedure with better performance consisted in implementation of dimension reduction of the number of inputs by selecting variables using the PV method which is based on principal component analysis (PCA), in combination with flow-based zonation and direct emulation of objective functions. This allowed us to pick representative points within flux regions and petrophysical properties for the chosen points were used as input parameters in the proxy modelling.

3.1. General Methodology

The general methodology used for emulator building and application in reservoir studies is based on the general proxy-modelling framework adapted from Razavi et al. (2012), Ferreira et al. (2014) and He et al. (2016). The workflow is divided in five steps as presented in Figure 1. The main contribution of this work focuses on specific procedure implemented between step 2 and step 3.

3.1.1. Reservoir Characterization under Uncertainties

The first step of the general methodology consists in the definition of reservoir properties together with their correspondent uncertainty ranges. For the purposes of this work, we only consider uncertainties in properties of the geological model represented by geo-realizations. As referred, a realization is numerically characterized by the spatial distribution values of porosity, permeability in three spatial directions and net-to-gross ratio. Therefore, the number of parameters (order of 10^5 for a typical simulation model) that characterize a realization depends on the number of gridblocks of the reservoir numerical model.

3.1.2. Inputs Sampling

A sampling method is required to generate scenarios for the uncertain reservoir model. In this specific work, we consider only petrophysical uncertainty in our model. Therefore, we do not require a sampling method to combine uncertainties. Instead, equiprobable geo-realizations define each possible scenario for the reservoir simulation model. The outputs of simulation runs are used for proxy model building. Moreover, because our final goal is to construct a tool which is faster than the simulator for applications such as risk analysis, we evaluate prediction power of emulators for different sample sizes (training dataset).

3.1.3. Emulator Building

The idea of using reservoir emulation technique consists in estimating proxy models (PM) with outputs corresponding to some observable of the reservoir dynamics such as cumulative oil production for reservoirs. Craig et al. (1996) proposed a framework to build emulators. This consists in building a stochastic representation (emulator) of the computer model (simulator) outputs for input combinations that were not evaluated. Thus, an emulator takes system properties (x) as inputs and returns outputs (f_i) that correspond to selected observables of the problem. The contribution of this work relies on the manner of pre-processing a high-dimensional input space that is represented by geostatistical realizations in the reservoir simulation problems. For the purposes of this work, the objective functions to be emulated are cumulative oil, water and gas for a future production date. For each selected objective function we want to emulate, we represent the function as:

$$f_i(x_A) = \sum_j \beta_{ij}\, g_{ij}(x_A) + u_i(x_A) \qquad (1)$$

In Equation 1, x_A is the subset of input parameters considered in the estimation, β_{ij} are scalars, g_{ij} are deterministic functions and u_i represents a Gaussian process. In particular, the

deterministic functions and scalars can be estimated by a step-by-step regression model selection (Venables & Ripley, 2002) based on Aikake Information Criteria (AIC). In principle, the Gaussian process is optionally implemented to interpolate residuals, whereas the most of model output variation is explained by the regression (O' Hagan, 2006). The AIC-based modelling used for construction of mathematical models is a linear regression where the terms are selected by a stepwise algorithm that implements Aikake Information Criteria in Equation 2. Given a set of possible predictors the stepwise regression runs backward by dropping terms from the model and looking at improvements of the AIC measure. The selected input variables that are in the final model are called *active* variables. In Equation (2), each model likelihood L is computed from the model deviance and the variable *e.d.f.* corresponds to equivalent degrees of freedom.

$$\mathrm{AIC} = -2\log L + 2 \times e.d.f \quad (2)$$

3.1.4. Emulator Validation

To guarantee that a built emulator can reproduce reservoir numerical simulator outputs in any specific part of an application, we must assess the prediction quality of each component f_i, i.e., objective functions (OF). The purpose of this procedure is to confirm that emulator can encompass simulator results for a random sampled scenario. The first diagnostic criterion considered is the statistical fit measure Adjusted-R^2. This measure is calculated by Equation 3, where R is the coefficient of determination, n the sample size and k the number of predictors. Therefore, Adjusted-R^2 penalizes the use of spurious variables in the model.

$$\mathrm{R}^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right] \quad (3)$$

Then, to verify emulator prediction power, a cross-validation test is performed. This process involves a qualitative analysis (cross-plots) of simulator against emulator outputs for sampled scenarios (validation data) that are not used in the emulator building process.

Besides, to quantify prediction quality of emulators, we use a measure of discrepancy between simulation and emulation results known as normalized root mean square error ($RMSE_n$) defined by Equation 4. RMSE is a common measure (Chen et al. 2016) of difference between predictions of a model (emulator output) and the actual or observed values (simulator output).

$$\mathrm{RMSE_n} = \frac{\sqrt{\sum_1^{\mathrm{N}}(\hat{\mathrm{y}} - \mathrm{y})^2}}{\sqrt{\sum_1^{\mathrm{N}}(\bar{\mathrm{y}} - \mathrm{y})^2}} \quad (4)$$

In this case, normalized RMSE is a function of proxy outputs ($\hat{y}$), simulator outputs (y) and mean ($\bar{y}$) of predictions from the training dataset. The normalization is performed due to the different orders of magnitude for objective functions. Values of normalized $\mathrm{RMSE_n}$ near one represents a prediction no better than the average of outputs used as training data, and $\mathrm{RMSE_n}$ near zero represents an ideal match between the predicted and actual results.

Therefore, we have adjusted-R^2 (related to training data) and $RMSE_n$ (related to validation data) as measures for diagnostic and emulator quality assessment, respectively (See Table 1). Emulator errors that can be tolerated may well depend upon the application and the purpose of the study. As stated in the objectives section, we aim to set quality criteria for validation of emulators based on the results of our specific application.

Table 1: Summary of indicators used along this work.

Measure	Abbreviation	Related to...
Adjusted Coefficient of determination	$\boldsymbol{R^2}$	Training data
Root mean square error (normalized)	$\boldsymbol{RMSE_n}$	Validation data
Mean average percentage error	***MAPE***	Risk curves

3.1.5. Application

Reservoir emulation can be implemented in several applications within reservoir studies. The interest relies on using emulators to substitute the reservoir numerical simulator in procedures that demand a high number of scenario evaluations and therefore an extensive computational effort and time. As such, emulators can be used in several steps within methodologies for history matching, sensitivity analysis, uncertainty reduction, strategy optimization, risk analysis, among other applications. In our particular case, we use emulators to generate production risk curves using several sizes of training dataset. The idea is to find the cheaper (least number of scenarios for estimation) validated emulator to reproduce simulator results. To do that, we assess the accuracy of emulator at reproducing risk curve shapes by using an appropriate error measure, and then we establish quality criteria for emulator validation. Finally, we evaluate the error and computational cost for implementation of emulator in risk analysis.

To measure the computational cost of implementation of emulation in generation of production risk curves, we define the implementation time as a sum of total time of simulation

of training models, the time spent in building the emulator and the simulation time of validation data.

The error between risk curves is calculated using the mean absolute percentage error (*MAPE*). This gives us a quantification of the accuracy of emulator at reproducing the risk curve obtained with simulation. For a general case where we have a reference risk curve with points R_i and a predicted risk curve with points P_i, the *MAPE* is defined in Equation 5. There are no hard rules for tolerated *MAPE* ranges. Accepted intervals may depend upon the specific study case and purpose. In this case, we define *MAPE* tolerance based on the results for selected reference risk curves obtained with simulation for benchmark cases (*MAPE* between risk curves obtained with simulation of 500 and 1000 scenarios). For illustration of *MAPE* measure refer to Figure 2.

$$MAPE = \frac{100}{\mathrm{N}} \times \sum_{1}^{\mathrm{N}} \left| \frac{\mathrm{P_i} - \mathrm{R_i}}{\mathrm{R_i}} \right| \quad (5)$$

3.2. Consideration of variation in petrophysical properties for emulation

This work concerns the incorporation of petrophysical uncertainty, represented by geo-realizations, as inputs in building emulators. This means bridging the gap between steps 2 and 3 of the general workflow (Figure 1) when we consider variation in reservoir spatially-distributed properties.

The strategy for approaching the problem consists in the selection of representative points within flux regions, which petrophysical properties could explain the variability of the corresponding well responses.

To devise a procedure that allows us to build emulators from realization inputs, we test specific workflows. All workflows can be separated in two core components: 1) Variable selection and 2) Zonation. These two components relate to parameterization of geo-realizations for use as inputs in emulation. We present the two components separately and then we explain how we used them for the different tests.

3.2.1. Variable Selection

Given that geo-realizations have the same source data (well logs, sampling, etc.), property values at each grid cell are correlated involving a stochastic process. For instance, in the model used in this work, a Sequential Gaussian Simulation (SGS) process is implemented to generate porosity and permeability spatial property distributions. The high number of parameters that define a realization is one of the main difficulties to include geological uncertainty in proxy modeling. For instance, it is unfeasible to estimate regression models by taking information at

all grid-blocks as inputs because of the high number of observations that would be required to correctly estimate all regression parameters. Besides, there is a lack of efficient computational techniques to tackle the challenge (Shan & Wang, 2010).

Hence, the proposal is to use a dimension reduction technique for the input parameter space to decrease the number of parameters that allow us to distinguish a realization from another. In the context of statistical inference (Guyon & Elisseeff, 2003; Boukouvalas et al., 2007), dimension reduction methods can be classified in *projection* and *screening* methods. If the belief is that there exists a smaller dimension representation, projective methods transform inputs into a manifold spanned by functions of original input values. On the other hand, screening methods consists in selection of relevant inputs (or disregarding spurious ones) than can act as predictors for modelling.

In this work, we implemented a selection (screening) of representative points in porosity and permeability maps for the training set of realizations. Three different procedures for variable selection are tested:

- *Random points:* We select random points in the grid to act as a representative sample of the whole realization. The idea behind this procedure is to select that an arbitrary collection of points that does not consider distribution of reservoir properties.
- *Evenly-spaced points:* Spaced points are chosen in the reservoir simulation model to reduce the number of total grid information in the realization. As in the previous approach, this procedure does not consider variability of petrophysical properties over realizations, but attempts to select a homogeneously located sample of points.
- *Principal variables:* The PV approach is a dimension reduction methodology based on Principal Component Analysis (PCA) that selects variables that most represent a problem in a statistical experiment. This method uses a criterion that combines correlation among variables and loadings on the Principal Components (For more details, see Cumming and Wooff, 2007). For our problem, this technique ranks grid points by using the variances and correlation matrix of property values among the set of realizations, allowing the selection of representative grid points for each property by their positions in the ranking.

The objective in this component is to represent the geostatistical realizations with a lower number of parameters. Property values at selected points for porosity and permeability maps are then used as inputs to emulate well responses.

3.2.2. Zonation

This component aims to define the region of interest for variable selection procedure. Because of the nature of fluid movements in reservoir, it is expected that well responses are more

correlated with petrophysical properties of regions where the fluids flow along the production period. Based on that premise, we tested two different approaches for defining those regions:

- *Location-based:* In this method, we correlate well responses with properties of grid-blocks near each well by dividing the reservoir in separate regions in accordance with well locations in the reservoir model. This procedure reduces the number of inputs parameters that must be treated in tandem.
- *Flow-based:* In this approach, first we evaluate fluids behavior along the production period within each well production zone and then define the regions by distinguishing draining areas. In this case, we can obtain overlapping regions for different wells.

In both approaches, we look forward to relating input parameters and simulation outputs for wells corresponding to the same region.

Then, combinations of both components described above configure procedures for "pre-processing" geostatistical realizations as inputs in emulation. The selection of the appropriate procedure is based on the model performance in accordance to diagnostics and validation described for step 4 of the general workflow of Figure 1. In Table 2 we present a summary of tested workflows.

Table 2: Combinations of tested workflows to parameterize geo-realizations.

Procedure	Variable Selection	Zonation
1	Random	Location-based
2	Spaced	Location-based
3	PV	Location-based
4	Random	Flow-based
5	Spaced	Flow-based
6	PV	Flow-based

3.3. Proposed procedure

In this section, we outline the generalization for random case studies of the procedure (Procedure 6 in Table 2) to consider variation of spatially-distributed properties in reservoir behavior emulation. The procedure consists in the implementation of a flow-based zonation plus a selection of variables that considers distribution and variability of petrophysical properties over a set of realizations, such as Principal Variables.

Thus, the proposed procedure to parameterize the spatial properties distributions as inputs in the emulator building can be summarized as: *selection of representative grid-block properties within each well drainage region*. As part of the workflow depicted in Figure 1, this is an intermediate step between the inputs sampling and emulator building that can be considered as a "pre-process" of inputs as illustrated in Figure 3.

Various approaches can be used for the implementation of the proposal. We present a procedure (See Figure 4) that was used in the development of this work, but alternatives exist for each step.

Once the inputs space is sampled in step 2 of the general workflow, the training dataset is used twice: On one hand, a small set of scenarios is used for zonation of the reservoir model. On the other hand, the complete set of training scenarios is used in the variable selection after zones are defined for each well. The suggested procedure is divided in three main steps:

a) *Selection of representative models (RMs):* As reservoir flow characteristics depend on the specific scenario, we first propose a selection of representative models for identification of drainage areas per well. For instance, we can use the method by Meira et al., (2015) which is based on simulation outputs for the training dataset: oil recovery factor and cumulative production for oil, water and gas. This method is based on Equation 6 and it consists in the selection of a set of scenarios ($\boldsymbol{\mathcal{R}}$), which minimizes a cross-plot function **F** based on Euclidean distances between objective function for subsets of training data. (See details in Meira et al., (2015)). The number of RMs can vary depending on the available resources for analysis. In this study, we recommend ten representative models, which is a reasonable number of scenarios to analyze (Figure 5).

$$\mathrm{F}_{cross}(\mathcal{R}) = \sum_{f,g} \mathrm{F}_{f,g}^{cross}(\mathcal{R}) = \sum_{f,g} \sum_{s \epsilon \mathcal{J}}^{\mathrm{N}} \Delta_{f,g}(s, \mathcal{R}) \quad (6)$$

b) Reservoir zonation: This step consists in a flow analysis for the selected representative models to identify drainage regions per well. This procedure can be done, for instance, by phase-velocities streamlines analysis. The analysis consists in assessing the flux lines along the production period of each well and highlighting the zones where these lines lie.

c) *Input variable selection*: Once the drainage regions per well and the representative models are defined, we implement a variable selection method such as Principal Variables for the inputs of the whole dataset of training scenarios. After Principal Components decomposition, this method classifies grid point data by $\mathbf{h_j}$ values calculated by Equation 7,

selecting variables based on eigenvalues ($\boldsymbol{\lambda_i}$) of the decomposition and variables with high loadings ($\mathbf{a_{ji}}$) on important PCs (See details in Cumming and Wooff, 2007). In this manner, we obtain the inputs variables per zone that will be used in the emulation of the corresponding well response.

$$h_j = \sum_{i=1}^{P} \left(\lambda_i a_{ji}\right)^2 \quad (7)$$

4. Case Study

A reference 3D geological model was built based on data from Namorado Field, Campos Basin, Brazil. It has been used to test and compare different proxy methodologies. In summary, to build a consistent geological model, we followed the creation of structural, facies and petrophysical models.

Facies modeling was defined using a Sequential Indicator Simulation (SIS) with vertical trend (Ravenne et al., 2002). In a general context of applying SIS, it provides 3D realistic images of the reservoir heterogeneities and is useful for controlling fluid flow and assessing final uncertainties in production (Seifert & Jensen, 1999).

Petrophysical modeling of porosity was defined using a 3D stochastic modeling, SGS, to perform the petrophysical modeling of porosity; combining well logs, distribution values for omni-directional variograms and 3D facies model to control and condition the porosity distribution (Dubrule, 1998; Kelkar, M., & Perez, 2002). This is a kriging-based method in which un-sampled locations are visited in a random order until all are visited. Porosity was then simulated, reproducing per-facies distribution as derived from the blocked well data. The same SGS algorithm was used to model permeability distribution.

Following the structural and properties modeling, it was necessary to define the rock and fluid properties. The rock fluid properties, represented by oil and water relative permeability curves and capillary pressure, were created based on real dataset of four different rock types. The fluid properties were also modelled through a real PVT data sample. The oil density of the model is 881.81 kg/m³ (28.97 ºAPI) at stock tank conditions (101.32 kPa and 15.6 ºC). The bubble point pressure is 20,909.73 kPa and reservoir temperature is 85°C. The oil viscosity (μo), gas viscosity (μg), the oil (Bo) and gas (Bg) formation volume factor and the solubility ratio (Rs) are coupled to the PVT curves as shown in Figure 6. Then, in our studies, we used the results of the black-oil fluid model.

For the purpose of this work in considering the variation of petrophysical properties in emulation, we selected a two-dimensional representation of the full-field fluid-flow numerical

simulation model to test and validate the proposed methodology. This model was named as META-2D (Figure 7).

META-2D comprises a black oil fluid model and reservoir with four vertical producers and one injector, arranged in a five-spot configuration as shown in Figure 7. This 2D model is composed of a 400 blocks (20x20x1) in a regularized corner-point grid with mean block dimensions of 92x92x150 m. The rock compressibility is 5.3×10^{-5} kPa^{-1} and bubble point pressure is 20,909.7 kPa. The total production time for the model is 20 years under the following operating and monitoring well conditions:

- Liquid rates are produced with the maximum possible rate for the field, 2,000 m³/day;
- Minimum production pressure is 18,633 kPa (190kgf/cm²);
- Water cut is 90%, maximum gas-oil ratio is 200 m³/m³ and minimum oil rate is 20 m³/day for monitoring and closing conditions for producers, if the condition is reached;
- Water is injected at the maximum possible rate for the field, 5,000 m³/day;
- Maximum injection pressure is 34,323 kPa (350 kgf/cm²).

Geo-realizations that represent each scenario of the simulation model are characterized by spatial distributions of effective porosity and permeability (totaling 800 parameters). Considering that it is a representative model of the full field, the average simulation running time for a single scenario is 30 seconds. Despite being a fast model, the preliminary goal is to validate the proposed procedure and then implement it in more complex cases with high execution time in subsequent studies.

5. Results

In this section, we present the results of implementation of the methodology described above. First, we show the process of emulator building. Then, we evaluate models obtained with different training dataset sizes in terms of prediction quality. Next, we use them to generate production risk curves and compare them with simulation results. Finally, we evaluate the implementation of emulator in risk analysis in terms of the computational cost and accurateness respect to simulation results.

Reservoir characterization and sampling

This section describes steps 1-2 of the methodology described. Simulation results for subsets of 1,000 scenarios (training data), where only petrophysical uncertainty is considered, are used

to build the proxy models for cumulative oil (N_p), gas (G_p) and water (W_p) production. In Figure 8, we have the characterization of permeability for the training dataset.

Emulator building and validation

This section comprehends the steps 3-4 of the general methodology. In the first part we present a description of the process of finding an appropriate procedure to parameterize geo-realizations in order to construct emulators for the chosen objective functions. In the second part of the section we show the assessment of emulators obtained with the selected procedure for different sizes of training dataset.

Procedures for emulator building

We selected N_p, G_p and W_p at final production time as output variables whose behavior we try to emulate (See Figure 9). On the other hand, input parameters selected from each procedure (See Table 2) are used in the estimation of regression models for objective functions at each well. The active variables (subset of the initial selected inputs) are chosen by a stepwise algorithm based on Aikake Information Criteria used to build regression models.

We tested various procedures to build emulators by selecting random points, evenly-spaced and using PV for regions defined by location and drainage area for each well. The first attempt consisted in dividing the reservoir in zones by location (procedures 1-3 in Table 2).

The location-based zonation procedure consisted in dividing the reservoir in four proportional regions (each with 100 grid-blocks) in accordance with the location of the four producers in the model. In Figure 10a, we illustrate the active variables selected for the quadrant corresponding to the zone of producer 2.

For this approach, we selected 40 grid-blocks for permeability and 40 for effective porosity (defined as porosity times net-to-gross, which is used as input in the simulator calculations) per region, using each one of the three variable-selection methods. The premise was that most of variability of each well response could be explained by the petrophysical properties of grid blocks around the well. For the first tested procedures (with location-based zonation), this turned out to be true for N_p and G_p. We obtained models with acceptable prediction quality for those objective functions. However, behavior of cumulative water seemed complex and its variability could not be explained by this location-based zonation and selection of points with any of the three approaches. Further tests by taking points outside each region indicated that behavior of well responses, in particular W_p, was better represented by points spread over the whole reservoir model. For this reason, we proposed a zonation approach that was based on drainage area for wells.

For the flow-based zonation approach (procedures 4-6 in Table 2), streamline analysis showed that drainage area for each well comprised the whole reservoir extension. Then, we selected 160 values for permeability and 160 for grid-block effective porosity in the whole reservoir, using the three variable selection methods. In Figure 10b, we illustrate the active variables chosen by AIC in the reservoir to explain N_p behavior of well 2. As shown, this automatically selected gridblocks (explainable variables) are more concentrated around the corresponding well.

In total, 320 property values (inputs) represented a realization in this approach. This is a very large number of inputs parameters for the AIC regression algorithm. The strategy was to build "partial" models for subsets of the 320 and combine the selected *active* variables by the step-wise algorithm to build a single proxy-model that represented the behavior of each objective function.

To compare the performance of the proposed procedures we sampled 400 scenarios and quantified the prediction quality of built emulators by using the $RMSE_n$. Results are presented in Table 3.

Table 3: Comparison of performance for tested procedures. Average $RMSE_n$ for 10 trials.

	Procedures					
	1	**2**	**3**	**4**	**5**	**6**
	Cumulative Oil Np					
PROD1	0.46	0.51	0.44	0.32	0.34	0.25
PROD2	0.38	0.41	0.37	0.30	0.26	0.23
PROD3	0.41	0.47	0.40	0.32	0.31	0.25
PROD4	0.46	0.49	0.44	0.36	0.39	0.29
	Cumulative Gas Gp					
PROD1	0.45	0.51	0.44	0.32	0.33	0.24
PROD2	0.39	0.41	0.37	0.29	0.25	0.23
PROD3	0.41	0.48	0.41	0.32	0.32	0.24
PROD4	0.43	0.46	0.42	0.35	0.39	0.28
	Cumulative Water Wp					
PROD1	0.54	0.54	0.50	0.50	0.47	0.41
PROD2	0.83	0.83	0.81	0.52	0.44	0.42
PROD3	0.49	0.53	0.48	0.41	0.39	0.34
PROD4	0.69	0.70	0.69	0.65	0.57	0.58

According to results of tests presented in Table 3, Procedure 6 (described in Section 3.3) was the best performing (lower prediction error measured by $RMSE_n$) approach that allowed us to build models explaining the observables behavior as a function of the properties of selected grid-points within the reservoir. For the purposes of the present work, results and application are obtained by implementing Procedure 6 to represent geo-realizations in emulation.

Emulation for selected procedure

In Figure 11, we present the Adjusted-R^2 for built emulators with procedure 6 as a function of the number of scenarios used as training data. From this data, we observe that there is not best case for Adjusted-R^2, so we must look at the predictive power of those models. Overfitting cases where Adjusted-R^2 is high but prediction quality is poor, must not be disregarded. Therefore, we are treating Adjusted-R^2 as an indicator (diagnostics) but not as definitive criterion for model assessment. Non-monotonic trends (Figure 11) for models built with less than 300 scenarios correspond to smaller number of principal variables selected as predictors for these cases, given that number of sample size limits the number of predictors for proper regression.

A cross-validation test was performed to obtain a qualitative evaluation of regression models of each objective function. For this process, 200 scenarios were sampled and simulated (Validation data). Figure 12 presents a comparison of cross-validation plots for cumulative water in well 2 emulators built with 100 and 300 scenarios. As observed, regression models with higher Adjusted-R^2 do not perform better at reproducing simulator results than regression models with smaller coefficient of determination. This result implies an over-fitted regression for emulators built with a small training dataset that does not work well for validation scenarios. We are then compelled to assess the prediction power of the built emulators using RMSE. In summary, we can say Adjusted-R^2 is a good indicator for emulator prediction power, but it is not definitive.

Prediction quality assessment

As proposed, we implement $\mathrm{RMSE_n}$ to evaluate prediction power of built proxy-models. For this case we build emulators for 10 different training dataset samples of equal size. Then, we calculated the average of normalized $\mathrm{RMSE_n}$ for each case using a validation dataset. Results are plotted in Figure 13 as a function of size of training dataset.

A reference $\mathrm{RMSE_n}$ curve is established from the training data used in each case. This prediction error for training data represents a minimum for $\mathrm{RMSE_n}$ of validation data given that emulator is fitted for the training scenarios. From normalized RMSE values found in Figure 13, we observe that results obtained for validation data are above reference values obtained from training data, as expected. The superposition of $\mathrm{RMSE_n}$ curves for N_p and G_p is reflecting a consistency of the procedure since reservoir pressure is above the fluid saturation pressure.

In addition, there is an indication that more training points does not necessarily translate into more prediction power. The $\mathrm{RMSE_n}$ reached a specific plateau for models at all wells for N_p, G_p and some wells for W_p. In the case of W_p, $\mathrm{RMSE_n}$ values obtained for PROD4 are above the reference value in comparison with other wells. This implies a more complex variability of the

493 objective function and confirms a lower prediction power as indicated by smaller Adjusted-R^2
494 values.

495 Being a proxy model, we expect emulator do not reproduce exactly simulation results. Then,
496 the issue is how much discrepancy we can tolerate. The answer may depend on the application
497 we consider. For instance, for production strategy optimization studies we might demand better
498 emulator prediction quality than for uncertainty reduction studies in an initial field development
499 plan. In this study, we use emulators to substitute simulation in generation production risk
500 curves at an early phase of oilfield development. Consequently, based on the error estimation
501 (MAPE) of risk curves obtained for emulators in comparison with simulation results, we
502 establish a "rule of thumb" criterion that might be used to discern whether a specific emulator
503 can substitute a simulation study in such application.

504 **Application: Production risk curves**

505 We implement emulators in a risk analysis procedure for oilfield in early stage of production.
506 We use emulators to generate production risk curves results for the final production time (7,305
507 days) and compare the results with those obtained by using the reservoir numerical simulator for
508 a medium fidelity model. For this purpose, we select a risk curve constructed with 1000
509 simulated scenarios as reference risk curve.

510 In order to compare risk curves we compute simulator/emulator discrepancy using the mean
511 absolute percentage error (MAPE). In our specific study case, it was noticed that for MAPE
512 values close or larger than 0.5%, dissimilarity between risk curves is visually significant. This
513 means we can use MAPE=0.5% as the tolerated cut-off value for dissimilarity between risk
514 curves obtained with validated emulator and reference result.

Table 4: Mean absolute percentage error (MAPE %) for production risk curves. We highlight the case for accepted MAPE with smaller training dataset size.

Training dataset size	PROD1			PROD2			PROD3			PROD4		
	Np	Wp	Gp	Np	Wp	Gp	Np	Wp	Gp	Np	Wp	Gp
100	0.15	0.46	0.10	0.15	0.83	0.20	0.16	0.53	0.14	0.10	0.73	0.15
150	0.11	0.23	0.17	0.11	0.64	0.12	0.11	0.14	0.12	0.11	0.45	0.13
200	0.23	0.20	0.22	0.14	0.94	0.09	0.11	0.27	0.11	0.13	0.49	0.15
250	0.15	0.29	0.15	0.08	1.05	0.12	0.11	0.10	0.09	0.15	0.21	0.16
300	0.13	0.21	0.14	0.10	0.20	0.10	0.12	0.12	0.11	0.18	0.29	0.19
350	0.11	0.23	0.11	0.09	0.22	0.07	0.09	0.10	0.07	0.13	0.32	0.14
400	0.09	0.28	0.09	0.07	0.15	0.08	0.09	0.12	0.10	0.13	0.30	0.12
450	0.08	0.25	0.10	0.07	0.21	0.08	0.08	0.12	0.07	0.14	0.29	0.12
500	0.07	0.19	0.07	0.07	0.22	0.08	0.08	0.11	0.07	0.13	0.32	0.13
550	0.06	0.16	0.07	0.06	0.21	0.07	0.09	0.14	0.09	0.14	0.35	0.15
600	0.06	0.16	0.07	0.06	0.20	0.08	0.09	0.14	0.09	0.14	0.30	0.14
650	0.06	0.19	0.07	0.07	0.22	0.08	0.09	0.15	0.09	0.12	0.32	0.12
700	0.06	0.20	0.06	0.06	0.21	0.08	0.09	0.16	0.08	0.13	0.30	0.12
750	0.06	0.19	0.06	0.06	0.23	0.08	0.08	0.15	0.07	0.11	0.31	0.11
800	0.06	0.20	0.07	0.06	0.24	0.08	0.08	0.14	0.07	0.11	0.30	0.11
850	0.06	0.16	0.06	0.07	0.26	0.08	0.08	0.13	0.07	0.10	0.32	0.11
900	0.06	0.17	0.06	0.06	0.23	0.07	0.07	0.14	0.07	0.11	0.35	0.10
950	0.05	0.16	0.06	0.07	0.23	0.07	0.07	0.14	0.07	0.11	0.33	0.10
1000	0.06	0.18	0.06	0.07	0.23	0.07	0.07	0.13	0.07	0.10	0.38	0.10

According to MAPE results in Table 4, the case with smaller number of scenarios that meet this criterion is the emulator built with 300 scenarios (Cross-validation plots for N_p emulators are found in Figure 14). This configures the cheapest validated emulator that can reproduce simulator results in this application. Then, according to results in Figure 11 and Figure 13, we can establish the criteria adjusted R-squared greater than 0.8 and normalized RMSE smaller than 0.5 as quality measure for emulators that reproduce simulator results in this application. This represents a *sufficiency* condition based on our specific case, noting that risk curve W_p of PROD4 was reproduced by an emulator outside the recommended criteria ranges. It is also noted that for number of scenarios greater than 300, differences among MAPE values are not significant and no relevant variation of predicted risk curves is observed.

As indicated from the MAPE assessment, comparison of risk curves obtained in Figure 15 and Figure 16 shows that the emulator built with 300 sample scenarios is capable of reproducing production risk curves (1000 trials for emulator and simulator) for N_p and W_p for all wells at the selected evaluation time. Besides, the curve obtained with the simulation outputs of the 300 scenarios is also plotted. Results show that curve constructed with emulator outperforms the risk curve for 300 simulated scenarios at reproducing the true curve (*Sim 1000*). In these plots, the reference point corresponds to a synthetic reality selected for the study case that derives from a

finer grid model constructed for research purposes. To complement the comparison, Figure 17 shows the results for the field as an integration of individual wells.

Implementation assessment

To compare the computational effort required by using emulation, we record the time spent in the estimation of regression models for each number of scenarios in the training dataset. Based on that, we define implementation time as the total time invested in the simulations used as training data, plus the actual time of estimation of regression models and the time spent in simulation of validation data. In this assessment, we are not including the human resource required to learn and implement the emulation technique.

In Figure 18, we plot the calculated implementation time for each case considered for N_p, W_p and G_p. The threshold time corresponds to simulation of 500 scenarios which is considered as "good enough" case compared to reference case according to a MAPE analysis. We find that the cheapest validated emulator (obtained with 300 scenarios) that reproduces reference risk curves within the established error tolerance is cheaper (20% less time) than the "good enough" case using simulation.

6. Conclusions and remarks

Previous works in reservoir emulation dealing with petrophysical uncertainty treated the problem in a restrictive way. For instance, some of them are characterized by implementation of multipliers or lack of geological consistency. A validated approach to deal with spatially distributed inputs, such as permeability and porosity in emulation was proposed and tested in a risk analysis application. We evaluated the prediction power of emulators built with different number of initial scenarios and built production risk curves that were assessed against simulation results. We showed that the proxy-models constructed are able to reproduce production risk curves for N_p, W_p and G_p obtained through simulation at the selected evaluation time within the tolerated discrepancy. Furthermore, according to our analysis:

- For emulators built with proposed procedure, Adjusted-R^2 greater than 0.8 and normalized RMSE smaller than 0.5 represent an "rule of thumb" sufficiency criteria to validate emulators that can be used to generate production risk curves that match simulation results within a MAPE tolerance cut-off of 0.5%. For our case study, the quality criteria were met for emulators built with 300 scenarios. Small improvement in prediction power is obtained with more training points at the expense of more computational resources.

- For our study case and the established criteria, an emulator constructed with 300 scenarios can reproduce reference risk curves obtained with simulation at a cheaper computational cost (20% less). Despite being a small gain compared to what can be expected from using proxy models, it can be understood because we are using a model that represent a portion of a full complex reservoir and which is fast to run.

In this preliminary work, we have implemented the emulator in a straightforward application because our focus was the development of the procedure for consideration of variability spatially-distributed properties in emulation. The full potential of this tool is expected to be more relevant when working with simulation intensive studies (e.g. history-matching workflows) and complex models such as carbonate reservoirs in Brazilian pre-salts. Because of the difference between emulator and simulation running times, computational cost saving from using emulators can be bigger as complexity, heterogeneity and size of reservoir model increase. Notwithstanding, complex cases also mean more training data for emulators, so the trade-off between model complexity and computational time saving is a crucial issue of further research.

Acknowledgments

This work was carried out in association with the ongoing Project registered as "BG-32 – Análise de Risco para o Desenvolvimento e Gerenciamento de Campos de Petróleo e Potencial uso de Emuladores" (UNICAMP/Shell Brazil/ANP) funded by Shell Brazil, under the ANP R&D levy as "Compromisso de Investimentos com Pesquisa e Desenvolvimento". The authors thank also UNISIM, DE-FEM-UNICAMP, CEPETRO, and CAPES for supporting this work and CMG, Emerson and Schlumberger for software licenses.

Nomenclature

Latin letters		Unit
Adj-R^2	Adjusted coefficient of determination	
a_{ij}	PCA loadings	
B_g	Gas-Formation volume factor	
B_o	Oil-Formation volume factor	
f	Objective function f	
f_i	Emulated output i	
F_{cross}	Cross-plot function	
g	Objective function g	
g_{ij}	Deterministic functions	
Gp	Cumulative gas production	m^3

h_j	PV measure	
k	Number of predictors	
L	Model likelihood	
n	Sample size	
N	Number of points in risk curve	
Np	Cumulative oil production	m^3
P_i	Predicted data	
R^2	Coefficient of determination	
$RMSE_n$	Normalized RMSE	
R_i	Reference data	
R_s	Gas-Oil ratio	
$\mathcal{R}$	Subset of training data	
$\mathcal{T}$	Training data	
u_i	Gaussian Process	
Wp	Cumulative water production	m^3
x	Input vector	
x_A	Active variables	
y	Simulator outputs	
$\hat{y}$	Proxy outputs	
$\bar{y}$	Mean of training data outputs	

Greek letters

β_{ij}	Regression scalars
λ_i	PCA eigenvalues
μ_g	Gas viscosity
μ_o	Oil viscosity

Abbreviations

AIC	Aikake information criteria
CPU	Central Processing Unit
DLHG	Discretized Latin hypercube with geo-realizations
JMM	Joint-model method
MAPE	Mean average percentage error
NTG	Net-to-gross ratio
OF	Objective function
PCA	Principal component analysis
PM	Proxy Model

PV	Principal variables
PVT	Pressure-Volume-Temperature
RM	Representative Model
RMSE	Root mean square error
SGS	Sequential Gaussian simulation
SIS	Sequential indicator simulation

589

590 **References**

591 Amorim, T., & Schiozer, D., 2012. Risk Analysis Speed-up with Surrogate Models. SPE-
592 153477. https://doi.org/10.2118/153477-MS.

593 Avansi, G., & Schiozer, D., 2009. Assisted Procedures for Definition of Production Strategy
594 and Economic Evaluation Using Proxy Models. SPE-122298.
595 http://doi.org/10.2118/122298-MS

596 Boukouvalas, A., Maniyar, D. M., & Cornford, D., 2007. Dimensionality Reduction in Complex
597 Models. Technical Report NCRG, MUCM Project.
598 https://research.aston.ac.uk/portal/files/201518/NCRG_2007_001.pdf

599 Chambers, R. L., Yarus, J. M., & Hird, K. B., 2000. Petroleum geostatistics for
600 nongeostatisticians Part 1. The Leading Edge, 19(5), 474. http://doi.org/10.1190/1.1438630

601 Chen, H., Loeppky, J. L., Sacks, J., & Welch, W. J., 2016. Analysis Methods for Computer
602 Experiments : How to Assess and What Counts?, Statistical Science 31(1), 40–60.
603 http://doi.org/10.1214/15-STS531

604 Craig, P.S., Goldstein, M., Seheult, A.H. and Smith, J. A., 1996. Bayes Linear Strategies for
605 History Matching of Hydrocarbon Reservoirs. Bayesian Statistics 5 (pp. 69–98). Oxford
606 University Press.

607 Cullick, A., Johnson, W., & Shi, G., 2006. Improved and More Rapid History Matching With a
608 Nonlinear Proxy and Global Optimization. SPE-101933. http://doi.org/10.2523/101933-
609 MS

610 Cumming, J. A., & Wooff, D. A., 2007. Dimension reduction via principal variables.
611 Computational Statistics & Data Analysis, 52(1), 550–565.
612 http://doi.org/10.1016/j.csda.2007.02.012

613 Cumming, J., & Goldstein, M., 2009. Bayes Linear Uncertainty Analysis for Oil Reservoirs
614 Based on Multiscale Computer Experiments. The Oxford Handbook of Applied Bayesian
615 Analysis, 241–270.

616 Dubrule, O., 1998. Geostatistics in Petroleum Geology. AAPG Continuing Education Course
617 Note Series #38. Tulsa, U.S.A: The American Association of Petroleum Geologists.

618 Ferreira, C. J., Davolio, A., Schiozer, D. J., Vernon, I., & Goldstein, M., 2015. Use of Emulator
619 and Canonical Correlation to Incorporate 4D Seismic Data in the Reduction of Uncertainty
620 Process. SPE-174387, http://doi.org/10.2118/174387-ms

621 Ferreira, C. J., Vernon, I., Schiozer, D. J., & Goldstein, M., 2014. Use of Emulator

Methodology for Uncertainty Reduction Quantification. SPE-169405. http://doi.org/10.2118/169405-MS

Guyon, I. & Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3, 1157–1182.

He, J., Xie, J., Wen, X. H., & Chen, W., 2016. An alternative proxy for history matching using proxy-for-data approach and reduced order modeling. Journal of Petroleum Science and Engineering, 146, 392–399. http://doi.org/10.1016/j.petrol.2016.05.026

Kelkar, M., & Perez, G., 2002. Applied Geostatistics for Reservoir Characterization. Richardson. Richardson, U.S.A: Society of Petroleum Engineers Inc.

Meira, L. A. A., Coelho, G. P., Santos, A. A. S., & Schiozer, D. J. (2015). Selection of Representative Models for Decision Analysis Under Uncertainty. Computers & Geosciences, 88, 67–82. http://doi.org/10.1016/j.cageo.2015.11.012

Mohaghegh, S. D., Modavi, A., Hafez, H. H., Haajizadeh, M., Kenawy, M. & Guruswamy S., 2006. Development of Surrogate Reservoir Models (SRM) For Fast Track Analysis of Complex Reservoirs. SPE 99667. https://doi.org/10.2118/99667-MS.

O' Hagan, A., 2006. Bayesian analysis of computer code outputs : A tutorial. Reliability Engineering and System Safety, 91, 1290–1300. http://doi.org/10.1016/j.ress.2005.11.025

Polizel, G. A., Avansi, G. D., & Schiozer, D. J., 2017. Use of Proxy Models in Risk Analysis of Petroleum Fields. SPE-185835. https://doi.org/10.2118/185835-MS

Ravenne, C., Galli, A., Doligez, B., Beucher, H., & Eschard, R., 2002. Quantification of Facies Relationships via Proportion Curves. In & A. R. In M. Armstrong, C. Bettini, N. Champigny, A. Galli (Ed.), Geostatistics Rio 2000: Proceedings of the Geostatistics Sessions of the 31st International Geological Congress. Rio de Janeiro: Springer Netherlands. http://doi.org/10.1007/978-94-017-1701-4

Razavi, S., Tolson, B. A., & Burn, D. H., 2012. Review of surrogate modeling in water resources. Water Resources Research, 48. http://doi.org/10.1029/2011WR011527

Santos, S. M. G., Gaspar, A. T. F. S., & Schiozer, D. J., 2017. Comparison of Risk Analysis Methodologies in a Geostatistical Context : Monte Carlo with Joint Proxy Models and Discretized Latin Hypercube. Working Paper.

Sarma, P., Durlofsky, L. J., & Aziz, K., 2008. Kernel Principal Component Analysis for Efficient Differentiable Parameterization of Multipoint Geostatistics. Math Geosci (2008) 40: 3–32. http://doi.org/10.1007/s11004-007-9131-7

Schiozer, D. J., Avansi, G. D., & de Souza dos Santos, A. A., 2016. Risk quantification combining geostatistical realizations and discretized Latin Hypercube. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 39(2), 1–13. http://doi.org/10.1007/s40430-016-0576-9

Schiozer, D. J., Santos, A. A. de S. dos, & Drumond, P. S., 2015. Integrated Model Based Decision Analysis in Twelve Steps Applied to Petroleum Fields Development and Management. SPE-174370 . https://doi.org/10.2118/174370-MS

Seifert, D., & Jensen, J. L., 1999. Using Sequential Indicator Simulation as a Tool in Reservoir Description : Issues and Uncertainties. Mathematical Geology, 31(5), 527–550.

Shan, S., & Wang, G. G., 2010. Metamodeling for High Dimensional Simulation-Based Design Problems. Journal of Mechanical Design, Vol 132. http://doi.org/10.1115/1.4001597

Slotte, P. a, Smørgrav, E., & Asa, S., 2008. Response Surface Methodology Approach for History Matching and Uncertainty Assessment of Reservoir Simulation Models. SPE-113390. https://doi.org/10.2118/113390-MS.

Venables, W. N., & Ripley, B. D., 2002. Modern Applied Statistics With S. Springer. New York. http://doi.org/10.1198/tech.2003.s33

Zabalza-Mezghani, I., Manceau, E., Feraille, M., & Jourdan, A., 2004. Uncertainty management: From geological scenarios to production scheme optimization. Journal of Petroleum Science and Engineering, 44(1–2), 11–25. http://doi.org/10.1016/j.petrol.2004.02.002

Zubarev, D. I., 2009. Pros and Cons of Applying Proxy-Models as a Substitute for Full Reservoir Simulations. SPE-124815. http://doi.org/10.2118/124815-MS

Figure Captions

Figure 1: General methodology flowchart for emulator building and application in reservoir studies

Figure 2: Illustration of MAPE. Measure of discrepancy between risk curves obtained with emulation and simulation. a) Good case. b) Bad case.

Figure 3: Diagram for parameterization of inputs from geostatistical realizations

Figure 4: Suggested procedure for parameterization of inputs from geostatistical realizations.

Figure 5: Illustration of RMs selection method by Meira et al. (2015). Selection of 10 models for study case. a) Risk curve for field cumulative oil. b) Cross-plot for field cumulative oil and cumulative water.

Figure 6: META-2D – Fluid modeling. (a) oil viscosity (μo) and gas viscosity (μg), (b) oil (Bo) and gas (Bg)formation volume factor and (c) Gas-oil ratio (Rs). p_b is the bubble point pressure.

Figure 7: Grid-block effective porosity map realization for META-2D model.

Figure 8: Permeability (mD) characterization for META-2D model. a) Random geostatistical realization. b) Mean values for training dataset. c) Standard deviation for training dataset.

Figure 9: Production variables scenarios used as objective functions in emulation at final production time 7,305 days. a) Scenarios for cumulative oil. b) Scenarios for cumulative water.

Figure 10: Illustration of PV + AIC model selection for emulators built with 300 scenarios for Np of PROD2. Red dots correspond to porosity and black dots to permeability active variables. a) Points selected near well location. b) Points selected for the whole zone.

Figure 11: Summary of emulator building. Adjusted-R^2 for regression models

Figure 12: Cross-validation comparison for W_p emulators. Straight black line represents coincidence of emulator (Y) and simulator results (T). Coefficient of determination for the model and prediction error ($RMSE_n$) for validation data are reported. a) Cumulative Water Producer 3. Emulation with 100 scenarios. b) Cumulative Water Producer 3. Emulation with 300 scenarios.

Figure 13: Normalized RMSE values as a function of number of scenarios used to build emulator. Results correspond to average of 10 different samples.

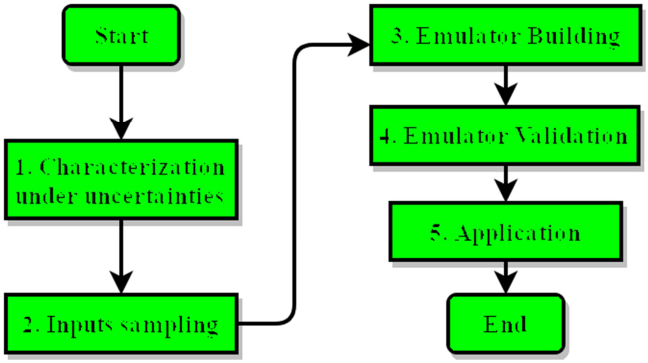
Figure 14: Cross-validation plots between simulator and proxy constructed with 300 scenarios for Cumulative Oil. Straight black line represents coincidence of emulator (Y) and simulator results (T). a) Producer 1. b) Producer 2. c) Producer 3. d) Producer 4.

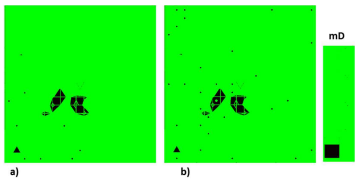
Figure 15: Comparison of Cumulative Oil Np risk curves obtained with 300 scenarios emulator and reference curve. a) Producer 1. b) Producer 2. c) Producer 3. d) Producer 4.

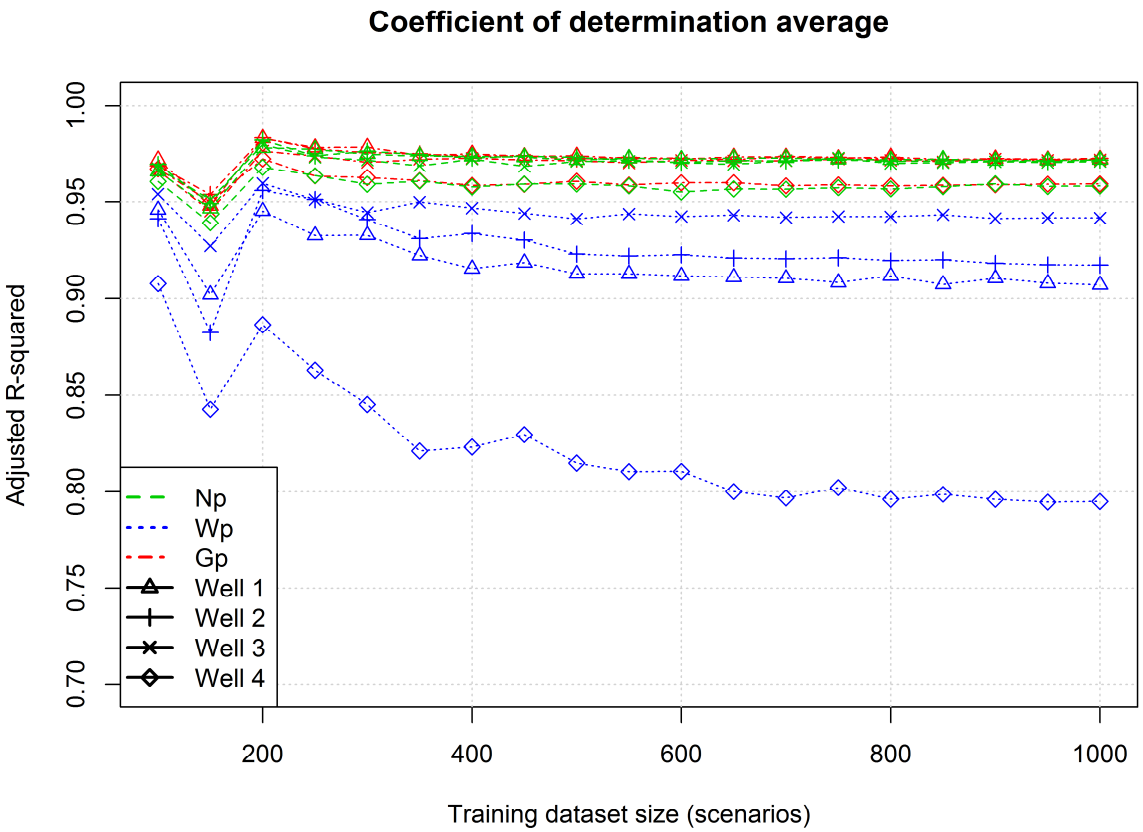
Figure 16: Comparison of cumulative water Wp risk curves obtained with 300 scenarios emulator and reference curve. a) Producer 1. b) Producer 2. c) Producer 3. d) Producer 4.

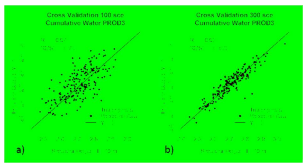
Figure 17: Field results as integration of separate emulators for the four wells. a) Field Cumulative Oil. b) Field Cumulative Water. c) Field Cumulative Gas.

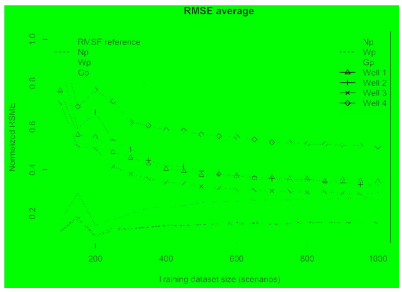
Figure 18: Implementation time for Np, Wp and Gp emulators. We have emulators with $RMSE_n$ smaller than 0.5 with implementation time less than established threshold. Results for 300 scenarios highlighted in violet box.

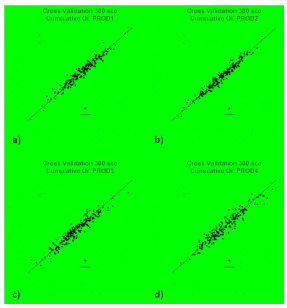


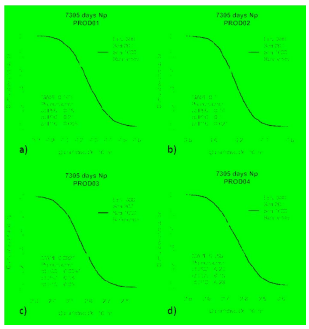


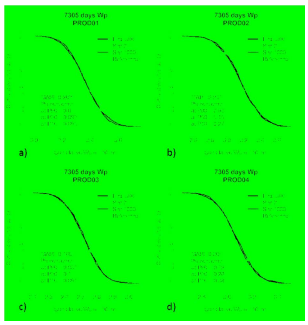


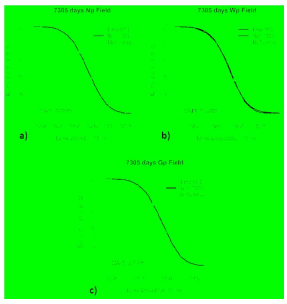


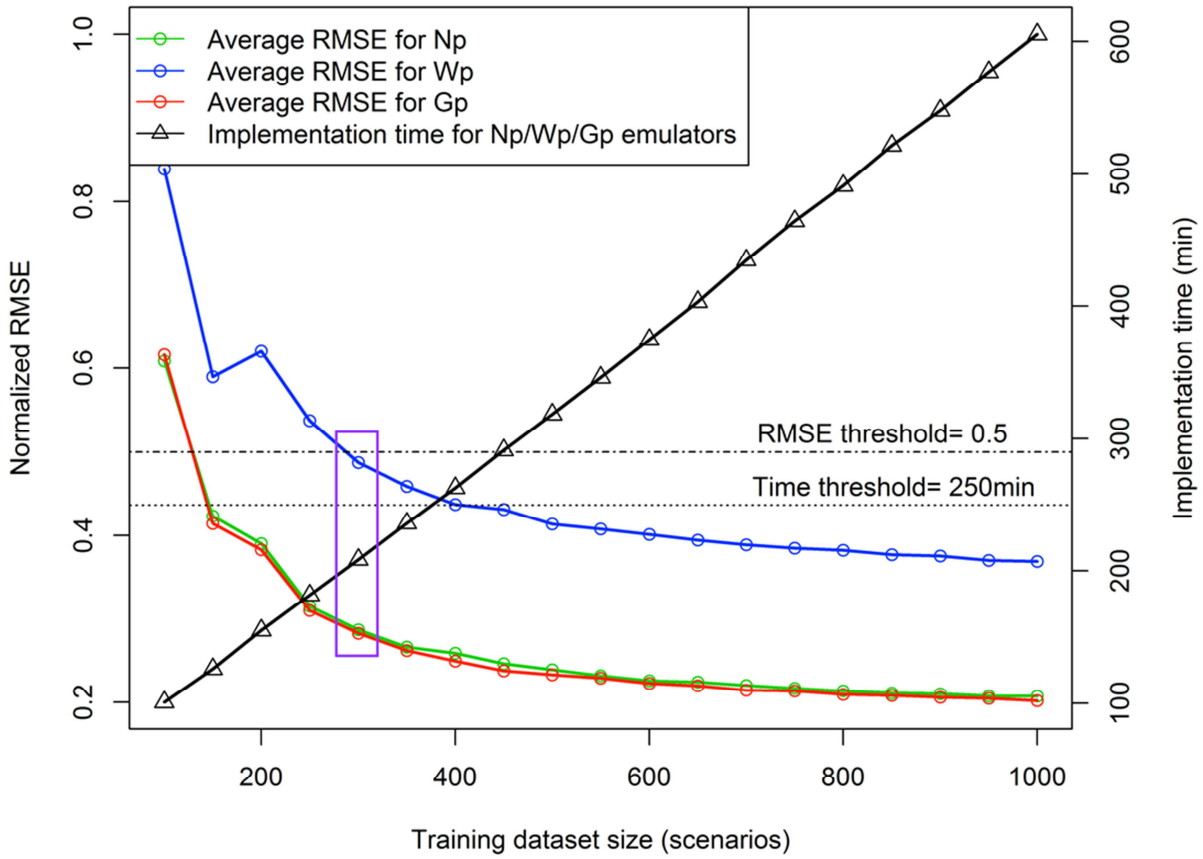


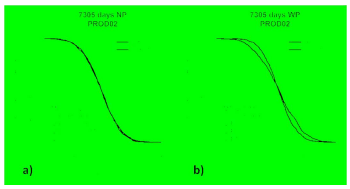


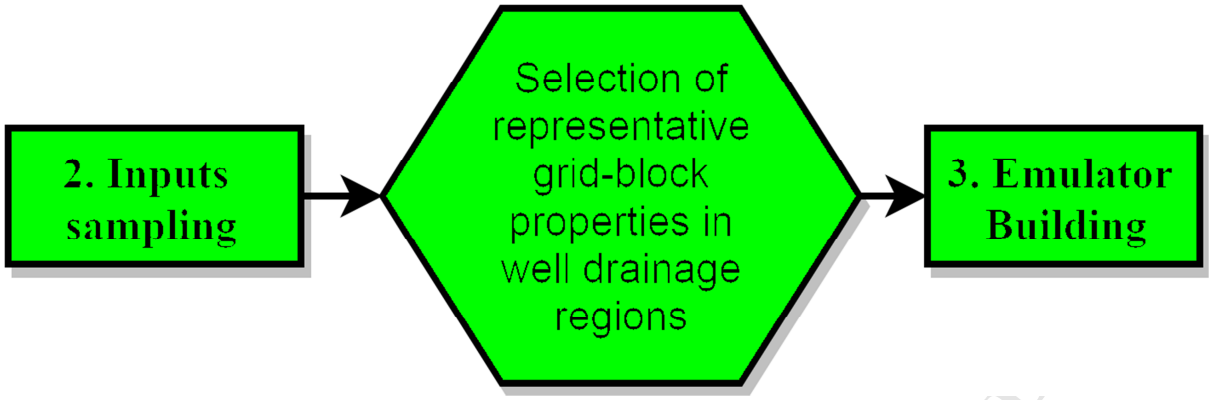


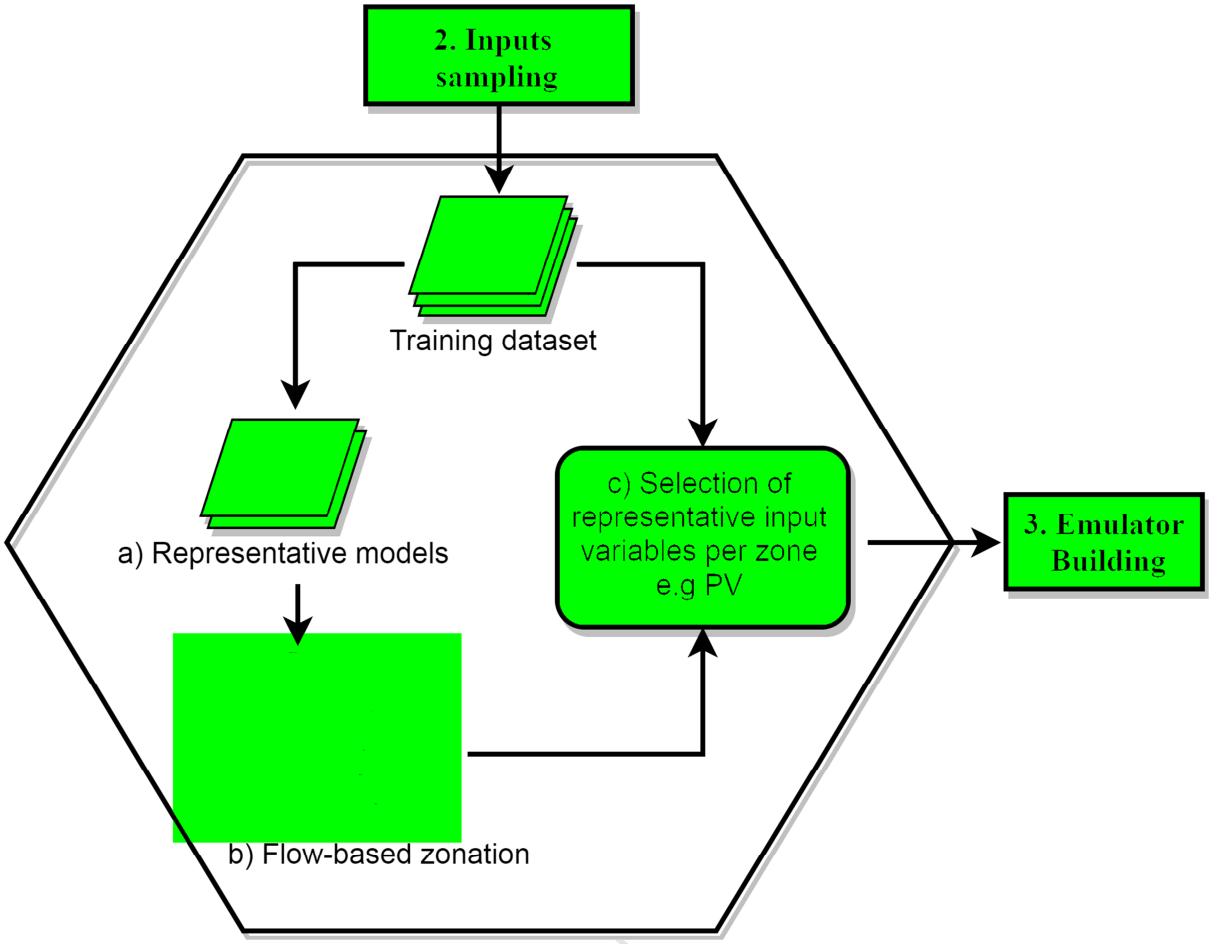


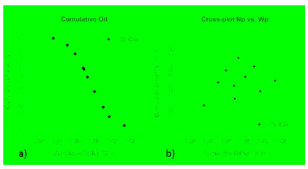


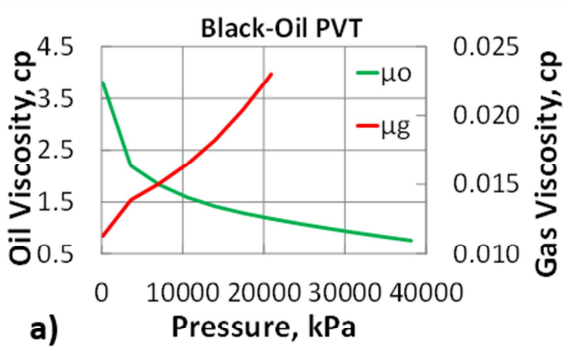


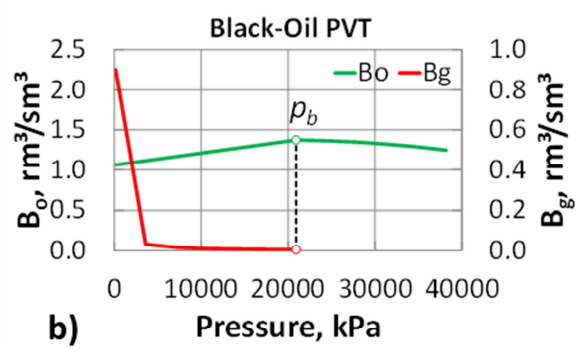


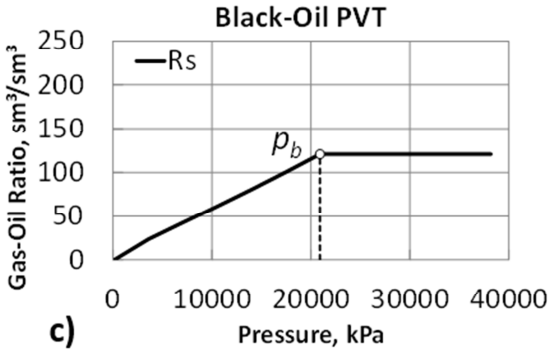


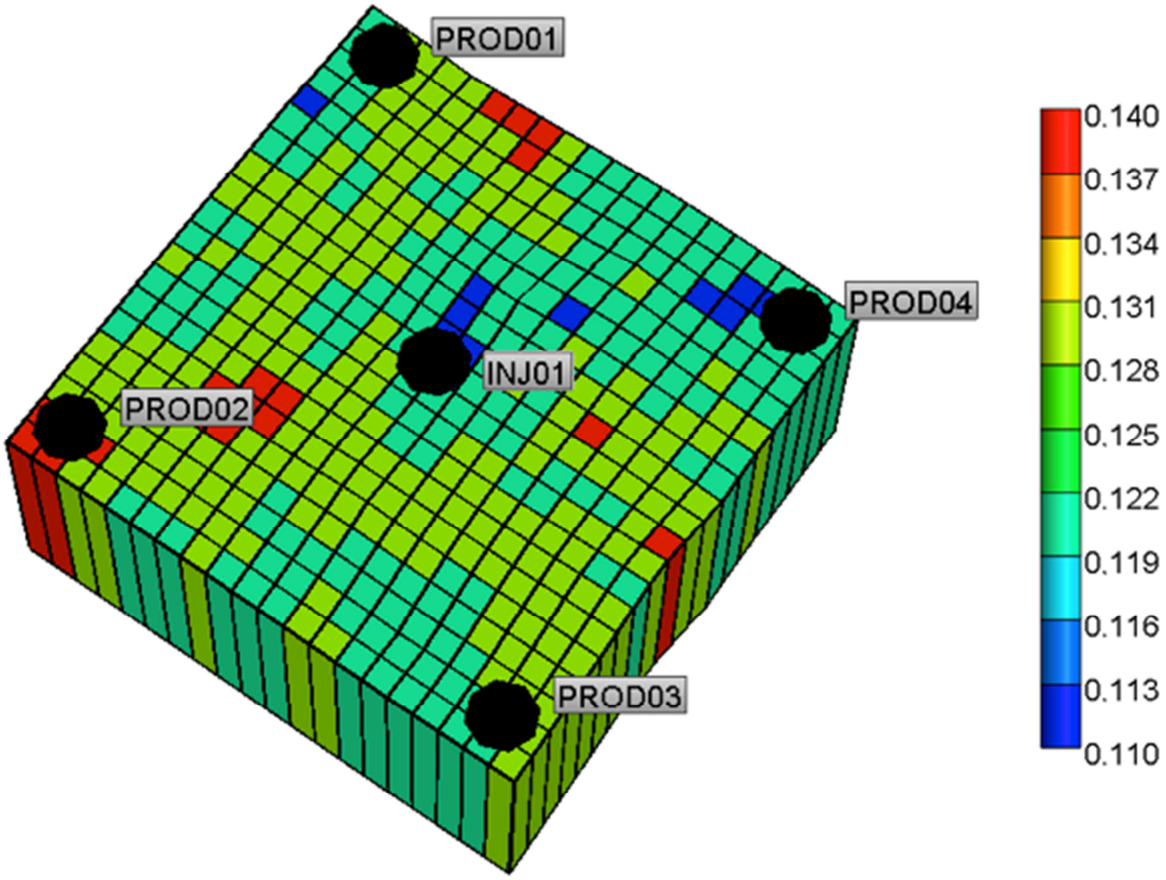


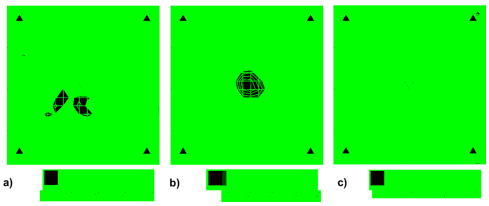


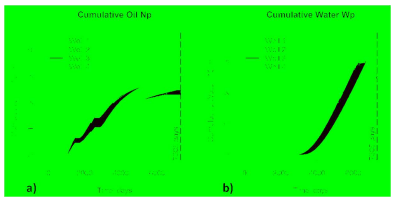












Highlights

- A procedure for consideration of spatially-distributed properties in reservoir behavior emulation is proposed.
- The procedure is based on a selection of representative grid-block properties within well drainage regions.
- Implementation of the proposed procedure in emulator building provides reliable results for risk curves generation in oilfield development.