Journal of Petroleum Science and Technology 13(4): 40, 2023, Pages 28-44
DOI:10.22078/jpst.2024.5034.1859

Journal of Petroleum Science and Technology

Research Paper

https://jpst.ripi.ir/

Investigating Proxy Models for a Production System in Integrated Simulations with Oil Reservoir

Joao Carlos von Hohendorff Filho*, Igor Ricardo de Souza Victorino, Marcelo Souza de Castro and Denis José Schiozer

Center for Energy and Petroleum Studies, University of Campinas, Campinas, Brazil

Abstract

This work evaluates the Proxy model application representing the production system for integrated simulation with a reservoir to reduce computational time while preserving the representativeness of financial return and hydrocarbon production behavior relative to a reference model. It includes specific Proxy models for production systems in integrated simulations that include their geometrical parameters, focusing on field production strategy optimization. The production system's Proxy models are developed through response surface methodology (RSM) and artificial neural network (ANN), which are generated and validated from a medium fidelity model (MFM). The validation is performed by cross-checking simulations. The developed RSM-based Proxy model obtained the highest representativeness by combining discrete variables (pipe segment diameters and the gas flow rate for artificial lift) with split continuous variables (lengths of the production column and flowline, liquid rate, and water cut) using several response surfaces. The developed ANN-Based Proxy model enhanced representativeness by combining all variables and increasing the number of MFM samples for ANN training. The RSM-Based Proxy model was selected due to its lower residual value than the ANN-Based Proxy model. The results from the production strategy of the simulated Proxy model in the MFM showed a difference of 4% in net present value compared to the simulation of the reference model, with both strategies obtained inside a production strategy optimization process. The reduction of computational time was close to 30% with the selected Proxy model, which it presents an advantage of using the proposed approach in optimization applications. The developed methodology provides an alternative to replace more robust production system models in integrated simulations with several advantages, such as: reduction of computational times, applications in more complex problems, and better-exploring uncertainties, and thereby, faster decision-making is obtained.

Keywords: Proxy Models, Numerical Simulation, Response Surface Methodology, Artificial Neural Network, Optimization.

Introduction

In the development of current oil production projects, given the high investments involved, adequate decision-making is necessary to maximize the hydrocarbon production performance and the financial return. The evaluation of the production strategy demands the integration between the reservoir and production systems, as this considers dynamic changes in the reservoir, leading to a more realistic analysis of the field production [1-5].

For decision-making, studies involving integrated optimization processes can require many numerical simulations, influencing the way to obtain their results. These studies have been applied in petroleum engineering, where the main objective is to plan production strategies with a convenient financial return for the development and management of oil fields. Objective functions (net present value or oil recovery factor, for instance) have been evaluated in optimization processes to assess long-term decision variables, such as well placement, processing capacities, and schedules for drilling, completion, and interconnection of wells [6]. For the integrated simulation, an integrated model is compounded by the model that represents the reservoir and another model that represents the production system, consisting of wells, gathering systems, flow networks, and surface installations. Simulations of integrated models with reservoir and production systems are time-consuming because of the complexity of these models. Reservoir modeling simplification is usually employed for the feasibility of time-consuming studies, and it can be categorized as physics-based simplifications, data-driven simplifications, and the hybrid approach, in which both simplifications are combined [7].

**Corresponding author:* Joao Carlos von Hohendorff Filho, Center for Energy and Petroleum Studies, University of Campinas, Campinas, Brazil

E-mail addresses: hohendorff@cepetro.unicamp.br

Received 2023-02-24, Received in revised form 2024-01-30, Accepted 2024-02-12, Available online 2024-05-02

Within the physics concept, fidelity is used to categorize simulation models with different degrees of reservoir description [8]. A higher-fidelity model (HFM) is a high grid cell resolution model regarding all possible geological and fluid model details that it may represent and run using the numerical simulator. Normally, HFM is computationally expensive with an expected higher quality of the results (accuracy). Medium-fidelity model (MFM) is a moderate-detailed model with a feasible running time and an expected acceptable quality of results. The low-fidelity models (LFM) refer to models that present a simplified degree of reservoir description related to rock and/or fluid properties derived from a medium-fidelity model (MFM). The advantage of using LFM is that it reduces the computational demand despite the expected lower accuracy.

Data-driven simplifications use data to build models trained (using regression techniques or machine learning) to learn and mimic the forecast behavior of reservoirs. Examples are time-dependent type curve (decline) models [9] and surrogate reservoir models trained with machine learning [10].

The hybrid physics-based and data-driven model takes advantage of physical phenomena (flow in porous media) and data observations (real data from the petroleum field) for standard practices in reservoirs. It is a new field of research that combines the interpretability, robust foundation, and understanding of a physics-based modeling approach with the accuracy, computational efficiency, and pattern-identification capabilities of data-driven. This model generally uses machine learning and artificial intelligence algorithms.

The Proxy model is an analytical function that provides a quick estimate of an objective function of a simulation model. An example of a Proxy model is the polynomial regression model obtained by response surface methodology (RSM). Its quality depends on the mathematical approach, the input used for its construction, and the complexity of the modeled system because this methodology involves the statistical design and different possible proxies that can be obtained, which may significantly affect the results. Another example of a Proxy model is the surrogate reservoir model, a Proxy model based on an artificial intelligence technique (artificial neural network - ANN) to represent the behavior of a history-matched reservoir simulation model [7].

Proxy models have been used in several reservoir engineering applications, including uncertainty modeling, sensitivity analysis, and history correspondence, among others [11-12], risk evaluation and analysis [12-13], performance forecast, upscaling [14], production history study [15] for complete reservoir simulations [10-,16], reduced models [17] and optimization [18-19].

Studies in the literature involve Proxy models for production systems, which focus on design and operation optimization [20-21]. The authors applied multidimensional piecewise-linear models to approximate the nonlinear functions of multiphase-flow simulations for production optimization of gas-lifted oil wells under facility, routing, and pressure constraints.

Some researchers developed a fast and accurate Proxy model for gas pipeline networks based on second-degree polynomial equations to model pressure drop in pipeline segments, maximizing oil production while maintaining safe and sustainable levels of CO_2 content and pressure in the gas stream [22].

Other authors used a methodology based on multidimensional segmented linear (piecewise) regression, obtaining a Proxy model for the integrated production system with a reservoir running explicit integration [17]. The authors proposed a more efficient interaction between the simulators, avoiding repetitions in production system simulation, reducing the total time of the numerical coupling, and accelerating the decision-making process in the field.

A new Proxy model based on neural networks, which integrates reservoir and surface behavior, was proposed recently [23]. All proxies provide production curves for the complete production period, replacing an integrated reservoir and surface system simulators for well placement optimization. In the simplified, integrated model, the representation of the entire reservoir external system is accomplished with multiphase flow tables. These tables are generated before reservoir simulation and subsequently included in the reservoir simulation file (during the optimization process). However, to our best knowledge, no studies have evaluated specific Proxy models for production systems in integrated simulations that include their geometrical parameters, focusing on field production strategy optimization. The dynamics of the reservoir and production systems are different, and there are situations in which a particular production system response is very similar or well-known over time, indicating the possibility of incorporating simplified models as Proxy models.

The current work develops a Proxy model (RSM and ANN-based) for the well and gathering system. It compares the selected one with a medium fidelity model (MFM) for production systems through integrated optimization of a production strategy based on [6, 24].

The evaluation considers the performance of the financial return, the fluid production and injection behavior, and the computational time of the total simulations. In this way, it is possible to verify if the Proxy model presents reliable results similar to that obtained by the MFM in this case study, providing an alternative for other studies and production projects.

Motivation

The use of integrated Proxy models for production systems over complex reservoir models can be representative depending on the application. In field development optimization studies, evaluating an oil field production using a model that enables finding fast and reliable solutions is important. The optimization process involved in this type of problem has a high computational cost due to the intensive use of physical simulators. Also, a simplified model can show good fidelity compared to a more complex one. Thus, shorter computational times can be achieved similarly to real situations, promoting efficient decision-making.

In this way, the influence of important stages in optimizing production strategies to evaluate the production performance of a field could be determined more quickly and easily, and the choice of the best strategy can be more effective.

Objective

This work aims to develop a Proxy model (RSM- and ANN-based) for a production system representative of integrated simulations with reservoirs derived from the MFM for the same production system in decision-making based on production strategies. Based on this developed model, a comparison is made between the financial and production performances of the MFM and the Proxy model integrated with a reservoir model through the optimization of the production strategy to verify similarities between the models and the feasibility of using the Proxy model to replace the MFM.

Materials and methods

Proxy Model Development and Validation

Developing a Proxy model representing the production system was proposed as an alternative to the MFM and considered a reference for this study for developing an integrated oil field. Two different approaches for Proxy modeling were evaluated: RSM-based and ANN-Based.

RSM-Based Proxy Model

Response surface models are commonly referred to in statistical literature. Even though this model does not exactly approximate the experimental data, response surface models have been widely adopted in the petroleum industry due to their ease of understanding, flexibility, and computational efficiency. Their general formulation for quadratic polynomial regression can be given by Equation 1:

$$y(x)=\beta_0\sum_{i=1}^{n_d}\beta_i x_i+\sum_{i=1}^{n_d}\sum_{\substack{j=1\\ j>1}}^{n_d}\beta_{ij}x_i x_j+\sum_{i=1}^{n_d}\beta_{ii}x_i^{\,2} \quad (1)$$

where x is a vector of input variables of length nd, x_i is a linear term, $x_i x_j$ is a cross term, x_i^2 is a quadratic term, and β_0, β_i, β_{ij}, βii represent unknown regression coefficients for constant, linear, cross, and quadratic terms respectively. Beta terms are estimated using the least squares approach.

Estimation of the response surface model includes (1) selection of the terms to be included in the model, (2) selection of an experimental design for building a second-order (quadratic) model, and (3) calculation of the regression coefficients. This Proxy model performs poorly for highly nonlinear multidimensional spaces [16].

The evaluations for obtaining the RSM-Based Proxy for the production system are based on [25] using response surfaces, and, in this study, three phases are considered. Response surfaces are obtained through experimental statistical planning (factorial design) to represent the Proxy model.

Variables considered important in the production system are selected, including those that are operational (e.g., injection gas flow for the artificial method of lifting by a gas lift and flow of produced liquid), design (e.g., internal pipe diameters, lengths of production columns and marine pipes, valve position for gas lift), and uncertainties (e.g, water fraction - BSW and gas-oil ratio - GOR) which in various combinations between them generate models for different simulations to obtain the respective responses, as example, well head pressure (WHP) or bottom-hole pressure (BHP).

In generating each response surface, an adjustment value is obtained - R^2 (coefficient of determination) and compared to a minimum pre-established value considered in the first analysis as part of having an accepted response surface.

The definitive approval of the RSM-Based Proxy model (composed of one or more response surfaces) occurs by a validation test with random data of the variables included in the respective operating ranges that can confirm the representativeness of the model. Thus, new simulations are carried out (with the random data on the response surfaces), and the results are correlated with those obtained by generating each response surface that is initially developed. Another correlation coefficient between the new evaluations must be evaluated to validate the response surfaces.

The development of an RSM-Based Proxy model initially represented by a single response surface is generated through all selected variables and regarded as important in the production system, considering the operational limits adopted by each one. Each combination is simulated, generating the response adopted at work. The factorial design produces several combinations and responses, generating a response surface model representing the Proxy model, denominated as Phase 1.

If the correlations between models do not satisfy the minimum adjustment values stipulated, Phase 2 is started by performing the same procedure, fixing some variable values, and considering the rest with their respective operational limits. A response surface is defined for each combination of fixed variables, and it must answer the same validation criteria as in Phase 1. If it does not satisfy the minimum stipulated values, Phase 3 is started. This last phase evaluates impacting variables on production performance and selects them for news analysis with sectioned limits. The initial evaluation considers splitting the operating limit ranges in half, and the procedure for each range is carried out in the same way as the previous phases. At some point in the evaluation, with a certain operating range for one or more variables, the related criteria must be satisfied.

The final RSM-Based Proxy model is the composition of all response surfaces referring to the operational ranges evaluated and validated for the impacting variable in production with fixed and non-fixed variables.

ANN-Based Proxy Model

An artificial neural network is an emulation of a biological neural system. ANN consists of base elements – nodes, analogous to neurons in biological systems. Any node receives signals from neighboring nodes and processes them to provide a single output. To construct an artificial neural network, the user must define its topology, including the number of hidden layers, nodes per hidden layer, and activation function. Then, input weights should be estimated for every node. The number of hidden layers and nodes influences the ability of the neural network to reproduce different degrees of non-linearity. However, the number of hidden nodes is restricted by several experiments used in ANN construction [16].

The evaluations for obtaining the ANN-Based Proxy for the production system are based on [26] using nonlinear regression approaches based on the Group Method of Data Handling (GMDH).

GMDH is a family of mathematical modeling and nonlinear regression algorithms [27]. This approach, also known as

Polynomial Neural Network, can be assumed as a specific type of supervised Artificial Neural Network (ANN). In addition to modeling specifications, GMDH uses the idea of Natural Selection to control the network size, complexity, and accuracy. The main application of GMDH is the modeling of (1) complex systems, (2) function approximation, (3) nonlinear regression, and (4) pattern recognition.

A complex multidimensional decision hypersurface can be approximated by a set of polynomials in the input signals containing information about the hypersurface of interest. Using a multilayered perceptron-like network structure, the approach fits a high-degree multinomial to the input properties. Thresholds are employed at each layer in the network to identify those polynomials that best fit into the desired hypersurface. Only the best combinations of the input properties are allowed to pass to succeeding layers, where more complex combinations are formed. Each element in each layer in the network implements a nonlinear function of two inputs. The coefficients of each element are determined by a regression technique, which enables each element to approximate the true outputs with minimum mean-square error. For further information, readers can refer to [27].

Variables considered important in the production system are the same as previously selected, including those that are operation, design, and uncertainties, which in various combinations between them generate models for different simulations to obtain the respective responses, for example, well head pressure (WHP) or bottom-hole pressure (BHP). Simulations are carried out with the random data for MFM representing the production system to generate an original dataset (dataset the model uses to learn from data).

In generating the artificial neural network, the model is trained over a subset of the original dataset. For this set, an adjustment value is obtained - R^2 (coefficient of determination) and compared to a minimum pre-established value considered in the first analysis as part of having an accepted ANN.

The definitive approval of the ANN-Based Proxy model occurs by a validation test with another subset of the original dataset of the variables included in the respective operating ranges that can confirm the representativeness of the model. Another correlation coefficient must be evaluated so that the ANN can be validated. If the correlation coefficients are unacceptable, the dataset size is increased to allow a new ANN training.

Proxy Model Selection

RMSE (Root Mean-Square Error) is used to measure the total residuals between modeled values (Proxy outputs) and the observed values (simulator outputs). Equally, normalized RMSE (NRMSE) is used to measure the discrepancy between Proxy and simulation results and quantify the forecasting quality of proxies. Equation 2 calculates this indicator:

$$NRMSE = \frac{\sqrt{\sum_1^N (\hat{y} - y)^2}}{\sqrt{\sum_1^N (\bar{y} - y)^2}} \tag{2}$$

where $\hat{y}$ is the Proxy output, y is the simulator output, and $\bar{y}$ is the mean of forecasts from the training dataset [28].

The Proxy model with the lowest RMSE and NRMSE, mostly near zero, will be selected for application for the following items.

Proxy Model Application and Comparison with MFM through the Optimization of the Production Strategy in an Integrated System

Evaluation and Comparison Between the MFM and the Proxy Model through a Production Strategy

Before the evaluation in stages of the production strategy, the validated Proxy model is submitted to a simulation integrated into the reservoir to verify its financial and production performance. Using the same production strategy, it is then compared with the same performances obtained from the MFM integrated to the reservoir. Suppose there is a considerable divergence between the results. In that case, the process is re-evaluated and restarted, returning to proxy model development and validation. If the results are similar and within certain criteria adopted, the production strategy can be optimized.

Evaluation Description in Stages of the Production Strategy

The evaluation using sequential optimization in stages is performed according to [29]. The financial performance, the evaluation requirement, is calculated by the objective function of the NPV (net present value). In five stages, the resulting configurations with the highest NPV are selected. Stage 1 evaluates the number of producer and injector wells. Stage 2 evaluates configurations with the coordinates of all wells involved. Stage 3 analyzes the pipe diameters and gas injection rate. Stage 4 seeks to coordinate for the platform in the field. Through potential areas, coordinates are then arbitrated. Stage 5 verifies the production and injection capacities of the platform.

The process is iterative in optimization cycles, and the number of cycles can vary depending on an established stopping criterion. A following cycle is performed considering the best result from Stage 5 of the previous cycle, and the same procedure is performed. If the final NPV of the next cycle is equal, or the difference between NPVs is within the value of the stopping criterion, the process is ended; otherwise, another cycle is performed.

The simulation times are compared in both optimizations of integrated models. The computational times of the Proxy model stages are verified as the total time involved in the entire cycle analysis.

The stages of sensitivity of the production strategy are evaluated, and in this analysis, the stages that most impact the increase in financial return are verified.

In addition, the final production strategies and the behavior between models involving oil and water production in the field are compared. The results are verified if there is a similarity or divergence among the results and if the Proxy model can be representative in the integrated simulations.

Application

The study was developed using data from the UNISIM-II-D benchmark [30-31], which represents a carbonate model of the reservoir with several characteristics, such as a depth of 5,000 m from the surface, reservoir temperature of 58 ºC and the initial reservoir pressure of 450 kgf/cm². More details are found in the cited publications.

The characteristics of the fluids used are the same as those

mentioned in [29]: the relative densities of oil, gas, and water of 0.862, 0.863, and 1.01, respectively; GOR (gas-oil ratio) of 233.8 m³/m³; oil viscosity at roughly 1.14 cP, and CO_2 molar fraction of 8.24%.
Empirical correlations of multiphase flow by Beggs and Brill [32] were used for the MFM of the production system. The correlation of Standing [33] was used for fluid. The production system consisted of 12 producer and 8 injector wells (all satellite wells), composed of a production/injection column with a gas lift valve, submarine flowline, and riser connected to a platform represented by nominal production and injection capacities. The base scenario of the production system was the same as [4, 29].
The wells had distinct conditions (different productivity index - PI), and the geometry considers the scenario of satellite wells (similar for all wells) with variation in pipe lengths (production column and maritime flowline), which depends on the depth of each well and their distance to the platform. Thermal calculation was not considered, assuming a linear temperature gradient in pipes.
The economic model considered for calculating the objective function NPV is described in sequence through considerations based on [4, 29].
Equation 3 calculates the platform cost (Invplat):

$$Inv_{plat} = 0.84(417 + 16.4Cp_o + 3.15Cp_w + 3.15Ci_w + 0.1nw) \quad (3)$$

Cp_o, Cp_w, and Ci_w are platform capacities for oil and water production and injection, respectively. The nw is the number of wells connected to the platform.
To calculate the objective function (NPV), the equations and parameters remained like those of [6]. The objective function considered net cash flow over a field's lifetime. In this project, the net cash flow for each period was calculated based on the Brazilian R&T fiscal regime considering gross revenues from oil and gas sales, total amount paid in royalties (charged over gross revenue), total amount paid in special taxes on gross revenue, operational production costs associated to the oil and water production and water injection, corporate tax rate, investments on equipment and facilities, and abandonment cost.
The main fiscal assumptions and the economic scenario are described in Table 1. Economic parameters for the production system are described in Table 2. The value ID is related to piping inner diameter in inches.

Table 1 Economic scenario of parameters [24].

Parameters	Value	Unit
Oil price	257.9	USD/m3
Oil production cost	48.57	USD/m3
Gas price	0.026	USD/m3
Gas production cost	0.013	USD/m3
Gas injection cost	0.014	USD/m3
Water production cost	4.86	USD/m3
Water injection cost	4.86	USD/m3
Investment on platform	Eq. (1)	USD millions
Abandonment cost (% investment in drilling and completion)	8.2%	-
Annual discount rate	9%	-
Corporate tax rate	34%	-
Social tax rate—charged over gross revenue	9.25%	-
Royalties rate—charged over gross revenue	10%	-

Table 2 Economic parameters for the production system [24].

Economic parameters	Technical parameter/ decision variable	Value ID (in)	Cost	Unit
Investment in connection (well-platform) of vertical wells	Flowline	6	647	USD/m
		8	1,666	USD/m
	Riser	6	1,276	USD/m
		8	2,189	USD/m
	Riser and flowline installation		9.86	USD millions
Investment in drilling and completion of vertical wells	Production column	5	228	USD/m
	Drilling and completion		18.35	USD millions
Additional investment in connection for artificial lift	Injection flowline	4	346	USD/m
	Riser	4	742	USD/m

Proxy Model Development and Validation

RSM-Based Proxy Model

A satellite well model was considered the basis for constructing the response surface [4, 29].

Phase 1 considers all the variables selected for the study and their operational limits adopted. For this situation, a single response surface was generated. Factorial design (Central composite–centered) was chosen to obtain the response surface using the Cougar Software (Beicip).

The variables for designing experiments and obtaining the response surface are described in Table 3; these are the lengths of the pipes (production column - PC, flowline - FL, riser - RI with a fixed length of 166 m), liquid flow rate (QL), water fraction (BSW), gas injection flow rate for the artificial lift method by gas lift (Qgi), and internal pipe diameters (ID) [4, 29]. The design of experiments considered the response to bottom-hole pressure (BHP) calculated in the MFM simulator for each combination. For the generation of the response surface, 65 simulations were performed considering the operational ranges of the variables.

The Marlim multiphase flow simulator developed by Petrobras S/A was used for MFM simulations. The simulator was prepared for BHP generation and used to develop the response surface in the Cougar Software.

The validity test selected random values within the variables' limits on the generated response surface. The number of simulations carried out was 2,000 in the MFM to obtain the respective answers (BHP).

Phase 2 (Table 4) considered fixed values for certain variables (pipe ID and Qgi) as possible discretization values. The other variables' respective operating ranges were considered.

A specific response surface was generated using the same experiment design in each combination of fixed and non-fixed variables. There were 27 combinations of the variables, each generating 25 simulations, totaling 675 simulations for generating response surfaces.

The combinations of the fixed variables that involve the diameters of the pipes are denominated as standard abc; with (a) for the production column (PC), (b) for the flowline (FL), and (c) for the riser (RI). All diameters are in inches. The following includes the injection gas flow rate for the gas lift (Qgi) in m³/day, maintaining the format [*abc*–Qgi].

In this step, only the 5-inch diameter for PC was used because of the increase in the number of demanded response surfaces, and this value presents a better financial return combined with greater oil production [4, 29].

For each of the 27 response surfaces, random variables were selected within the operational ranges of the non-fixed variables for verification of responses (BHP) and correlated with the values obtained by the response surfaces with the operational ranges (maximum and minimum limits) initially considered in the construction of the response surfaces. The validation tests required 64 more simulations for each combination, running a total of 1728 simulations for the validation criterion to be attended.

For Phase 3, the variable QL was selected since it had the greatest impact among the variables analyzed (QL, BSW, and lengths of the production column and flowline).

A total of 96 response surfaces were created, with 25 simulations totaling 2,400 simulations. The validation test for these combinations was performed with 64 simulations totaling 6,144 simulations.

The response surface is accepted with a coefficient of determination $R^2 \geq 0.99$. For validation, the valid determination correlation is $R^2 \geq 0.95$.

Table 3 Values of the parameters used in the factorial design experiments to obtain the response surface.

PC (m)	FL (m)	QL (m³/day)	BSW (%)	Qgi (m³/day)	ID-RI (in)	ID-FL (in)	ID-PC (in)
4,624	1,000	1,200	0.0	0	4	4	3
4,657	2,099	1,850	47.5	100,000	6	6	4
4,690	3,197	2,500	95.0	200,000	8	8	5

Table 4 Combinations of fixed and non-fixed variables with their operational ranges for generating response surfaces representing the RSM-Based Proxy model.

Fixed Variables	Variables with Operational Ranges
ID pipes, abc (in) – Qgi (m³/day)	PC, FL, QL and BSW
544-0; 544-100,000 and 544-200,000	$4500 \leq PC \leq 4700$ (m) $1000 \leq FL \leq 6400$ (m) $100 \leq QL \leq 2700$ (m³/day) $0 \leq BSW \leq 95$ (%)
546-0; 546-100,000 and 546-200,000	
548-0; 548-100,000 and 548-200,000	
564-0; 564-100,000 and 564-200,000	
566-0; 566-100,000 and 566-200,000	
568-0; 568-100,000 and 568-200,000	
584-0; 584-100,000 and 584-200,000	
586-0; 586-100,000 and 586-200,000	
588-0; 588-100,000 and 588-200,000	

ANN-Based Proxy Model

The same satellite well from the previous field item was considered the basis for the construction of the ANN.

Table 5 presents each variable with minimum and maximum values in each column (representing the range of validity for each variable) used to generate the complete dataset for GMDH training and validation. Qgi, ID-RI, ID-FL, and ID-PC were considered fixed as described in the RSM-based proxy model.

The Marlim multiphase flow simulator was used for simulations involving the MFM. The simulator was prepared for BHP generation and was used to develop the GMDH in Matlab Software (Mathworks) based on [26].

The GMDH network was set with a maximum number of neurons in a layer of 15, a maximum number of layers of 4, a selection pressure in layers of 0.6, and a train ration (versus validation ratio) for an original dataset of 0.28 (similar to RSM-Based Proxy model). A full dataset of 12,000 MFM samples was generated.

The ANN training is accepted with a coefficient of determination $R^2 \geq 0.99$. The valid determination correlation for validation is $R^2 \geq 0.95$, the same values applied to the previous Proxy model.

Evaluation and Comparison between Models by Production Strategy

As an initial case for the problem, a production strategy was determined through a static quality map (NTG-Phi-So and effective permeability) with 12 producer and eight injector wells, focusing on the carbonate matrix, alongside the verification of the streamlines of original wells to determine the radius of influence of the wells. Structural tops, hydraulic compartments, potential of gravitational oil, minimum spacing among wells, and initial oil saturation were also verified. The bottom-hole position of the wells in this carbonate field was also considered, including the platform's initial allocation (X = 355,400; Y = 7,516,500), to model the production system simulation. From [29], the pipe diameters used were 5in for the production column and 6in and 8in for the flowline and riser with varying combinations and evaluations because, with these configurations, the results showed greater performance in the financial return. The values used for the initial Qgi of 100,000 m³/day and the platform capacity of 85% were selected as initial values based on the same work.

The evaluation of the production strategy verified the proximity of the results in terms of NPV (differences up to 10% would be accepted) of the results between integrated models.

The optimization of the production strategy included the following evaluation stages: number of wells (E1), allocation of wells (E2), sensitivity analysis of the pipe diameters and gas injection flow rates for the artificial lift (E3), platform allocation (E4), and platform capacity (E5). Part of the assumed values, mainly related to pipe diameters and platform capacity, were considered based on [29].

The sequence of these stages represents one cycle, and the number of cycles to be performed depends on the final values of each cycle, that is, the configuration that best presents the financial return by NPV. For NPV values equal to or above 7% relative to the previous cycle, another cycle of analyses was performed. If it is less, the optimization was finished without new cycles.

The computational times of each stage of the production strategy and the total evaluation time were analyzed and compared.

The explicit integration technique used in the simulations between the reservoir and production system was based on [3]. This technique poses an advantage, as it allows the coupling of several simulators to model the system properly and promotes flexibility to include well management alternatives and external software. The reservoir model was simulated by the IMEX™ software (CMG), coupled to the Proxy model via the internal coupler program.

Analysis of the Number of Wells

This assessment (E1) occurred according to Table 6, which describes the sequence of simulations up to the total number of wells in this field (method adopted in this work). The number of wells initially adopted, and the combination form may differ depending on the study's objectives. Forty-six simulations (S) were performed with different combinations of wells and the configuration containing a certain number of wells.

Analysis of Well Allocation

For the best result obtained in E1, coverage areas for each well without overlap were selected. The areas were different because of the heterogeneous distribution of the wells in the field. With the areas defined, simulations assigned several coordinates for each well. The criteria for choosing the areas contemplated regions that did not overlap the wells initially considered with a distance of each area of each well of at least 100 meters in the possible spaces between the wells.

Because the areas are vast, the optimization methodology used was the Iterative Discrete Latin Hypercube (IDLHC). IDLHC is a population-based optimization method that uses an iterative process based on the discrete Latin hypercube sampling method to maximize the objective function. At each iteration, the method selects the best samples and gradually reduces the search space, treating the frequency distribution as a posteriori of the levels of each variable. The advance of the method depends on three parameters only: (1) the number of iterations, (2) the number of samples, and (3) F, the cut percentage [34].

Fig. 1 shows the position of the wells in the field in the initial stage, with the chosen areas not overlapping and different from each other, according to the arrangement of the wells used to evaluate Cycle 1. The same procedure was used to select new areas for the disposal of wells in Cycle 2.

Table 5 Values of the parameters used in the original dataset to obtain the GHMD.

PC (m)	FL (m)	QL (m³/day)	BSW (%)	Qgi (m³/day)	ID-RI (in)	ID-FL (in)	ID-PC (in)
4,500	1,000	100	0.0	0	4	4	3
4,700	6,400	2,900	95.0	200,000	8	8	5

Table 6 Chosen configurations for simulations (S) with the number of wells (NW) evaluated and which wells were involved in the analysis (E1).

S	NW	Wells	S	NW	Wells
1	2	Prod-1,Prod-2	24	11	Prod-1,...,Prod-10,Inj-1
2	3	Prod-1,...,Prod-3	25	12	Prod-1,...,Prod-10,Inj-1,Inj-2
3	4	Prod-1,...,Prod-3,Inj-1	26	13	Prod-1,...,Prod-10,Inj-1,...,Inj-3
4	5	Prod-1,...,Prod-4,Inj-1	27	14	Prod-1,...,Prod-10,Inj-1,...,Inj-4
5	6	Prod-1,...,Pros-4,Inj-1,Inj-2	28	15	Prod-1,...,Prod-10,Inj-1,...,Inj-5
6	6	Prod-1,...,Prod-5,Inj-1	29	16	Prod-1,...,Prod-10,Inj-1,...,Inj-6
7	7	Prod-1,...,Prod-5,Inj-1,Inj-2	30	17	Prod-1,...,Prod-10,Inj-1,...,Inj-7
8	7	Prod-1,...,Prod-6,Inj-1	31	12	Prod-1,...,Prod-11,Inj-1
9	8	Prod-1,...,Prod-6,Inj-1,Inj-2	32	13	Prod-1,...,Prod-11,Inj-1,Inj-2
10	9	Prod-1,...,Prod-6,Inj-1,...,Inj-3	33	14	Prod-1,...,Prod-11,Inj-1,...,Inj-3
11	8	Prod-1,...,Prod-7,Inj-1	34	15	Prod-1,...,Prod-11,Inj-1,...,Inj-4
12	9	Prod-1,...,Prod-7,Inj-1,Inj-2	35	16	Prod-1,...,Prod-11,Inj-1,...,Inj-5
13	10	Prod-1,...,Prod-7,Inj-1,...,Inj-3	36	17	Prod-1,...,Prod-11,Inj-1,...,Inj-6
14	11	Prod-1,...,Prod-7,Inj-1,...,Inj-4	37	18	Prod-1,...,Prod-11,Inj-1,...,Inj-7
15	9	Prod-1,...,Prod-8,Inj-1	38	19	Prod-1,...,Prod-11,Inj-1,...,Inj-8
16	10	Prod-1,...,Prod-8,Inj-1,Inj-2	39	13	Prod-1,...,Prod-12,Inj-1
17	11	Prod-1,...,Prod-8,Inj-1,...,Inj-3	40	14	Prod-1,...,Prod-12,Inj-1,Inj-2
18	12	Prod-1,...,Prod-8,Inj-1,...,Inj-4	41	15	Prod-1,...,Prod-12,Inj-1,...,Inj-3
19	10	Prod-1,...,Prod-9,Inj-1	42	16	Prod-1,...,Prod-12,Inj-1,...,Inj-4
20	11	Prod-1,...,Prod-9,Inj-1,Inj-2	43	17	Prod-1,...,Prod-12,Inj-1,...,Inj-5
21	12	Prod-1,...,Prod-9,Inj-1,...,Inj-3	44	18	Prod-1,...,Prod-12,Inj-1,...,Inj-6
22	13	Prod-1,...,Prod-9,Inj-1,...,Inj-4	45	19	Prod-1,...,Prod-12,Inj-1,...,Inj-7
23	14	Prod-1,...,Prod-9,Inj-1,...,Inj-5	46	20	Prod-1,...,Prod-12,Inj-1,...,Inj-8

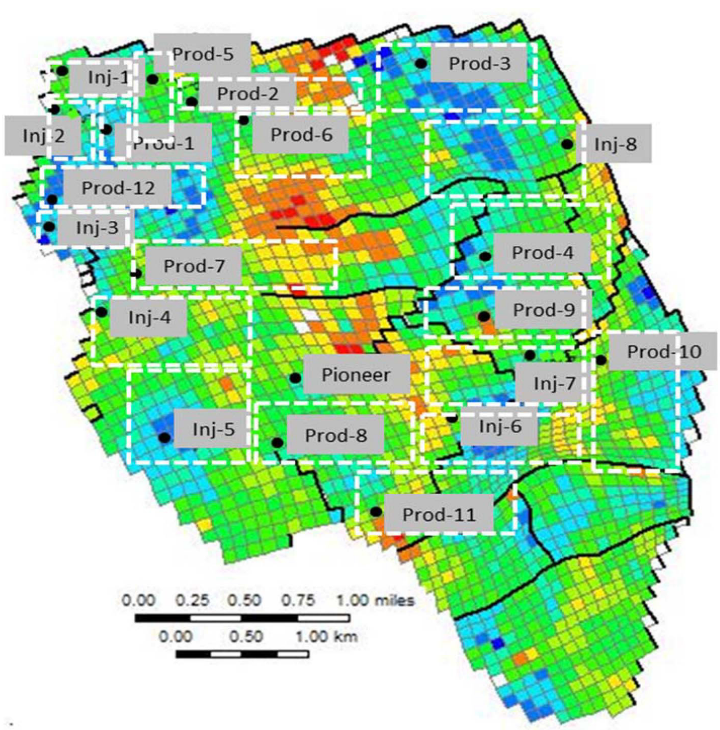

Fig. 1 Exemplification for determining the areas for initial well allocation.

The IDLHC methodology was performed with the parameters used by [24]: number of iterations = 8, number of samples (N) = 75, and cut percentage F = 0.9. A total of 600 simulations were run. The search space discretization considered a spacing of 100 meters for both vertical and horizontal directions within each area of each well.

Sensitivity Analysis Involving Pipeline Diameters and Injection Gas Flow Rate for Gas Lift

Through the best result of E2 or E1, this evaluation (E3) was performed. Table 7 describes the variables used to define pipe diameters (PC, FL, and RI) and gas injection flow rate for the artificial lift method [29]. The total number of combinations for simulations was 12.

Table 7 PC, FL, and RI pipe diameters and gas injection flow rate for the artificial lift gas lift (Qgi) method.

PC (in)	FL (in)	RI (in)	Qgi (m^3/day)
5	6	6	0
	8	8	100,000
			200,000

Platform Allocation Analysis

This stage (E4) evaluated the best coordinate for the platform through the NPV. Coordinates in available areas were arbitrated by the new well configuration obtained in E2. These areas were smaller, irregular, non-continuous, and non-adjacent. The preference was for more central coordinates of most wells if there are not impediments.

A minimum radius of 500 meters between the coordinates of each well and the coordinate of the platform was considered, avoiding overlapping of regions that affect operational interventions. At each coordinate, a sequential search was performed in all (X and Y) directions with a distance of 100 meters, generating four new radial coordinates.

Platform Capacity Analysis

For the final stage (E5), five percentage capabilities were simulated for the platform (100, 85, 65, 45, and 25%) with maximum flow rates for Qo (oil flow rate) = 28,621 m^3/day, Qw (water flow rate) = 22,897 m^3/day, Qwi (injection water flow rate) = 38,162 m^3/day, QL (liquid flow rate) = 28,261 m^3/day, and Qg (gas flow rate) = 6,668,000 m^3/day.

If there was an increase in NPV, the conducting configuration was adopted as the end result of the cycle and selected for stage E1 in the new cycle.

Determination of Computational Simulation Times

The computational time evaluation considers the time consumed for each stage. At the end, the total time of a cycle is the sum of the computational times of each stage, including the assembly time of the Proxy model for this model.

In E1, the total simulation time is the sum of the simulation times for each configuration containing several wells evaluated (Table 3). The fact that each configuration has a different number of wells generates a different simulation time relative to other configurations. In E2, the average time of each simulation round is evaluated and cumulated, resulting in the total simulation time. At E3, the average time for each combination is considered and multiplied by the number of combinations, providing the total simulation time. In E4, the total simulation time considers a representative average time among all simulated cases multiplied by the number of cases. E5 also considers an average time multiplied by the number of cases.

Results and discussion

Proxy model development and validation

Development of an RSM-based Proxy Model Represented by a Single Response Surface

Developing a response surface resulted in a coefficient of determination R^2 of 0.99, which is considered appropriate. However, the validity test had an R^2 of 0.86, below the adopted criterion. Fig. 2 shows the predicted BHP values from the RSM-based Proxy model and the simulation results from MFM in Phase 1 to check model performance. This cross-plot is useful to understand how well the regression model makes predictions for different response values. The Proxy model was not representative, as shown in Fig. 2, and Phase 2 was necessary (multiple response surface).

Development of an RSM-based Proxy Model Represented by Fixed and Non-fixed variables - Multiple Response Surfaces

Fig. 3 shows the predicted BHP values from the RSM-based Proxy model and the simulation results from MFM in Phase 2 to check model performance. All combinations of variables fixed presented coefficients of determination related to 27 suitable response surfaces ($R^2 \geq 0.99$), but for validation between the models, 22 of then presented values below the adopted criterion ($R^2 \geq 0.95$), with no representativeness, as shown in Fig. 3.

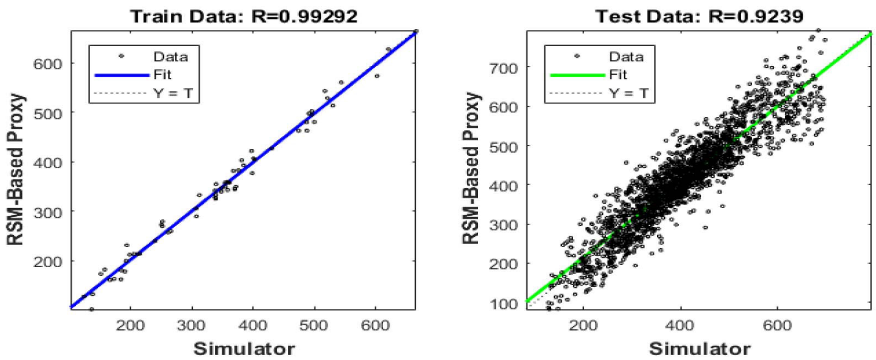

Fig. 2 Performance of the RSM-based Proxy in Phase 1 for BHP (kgf/cm²).

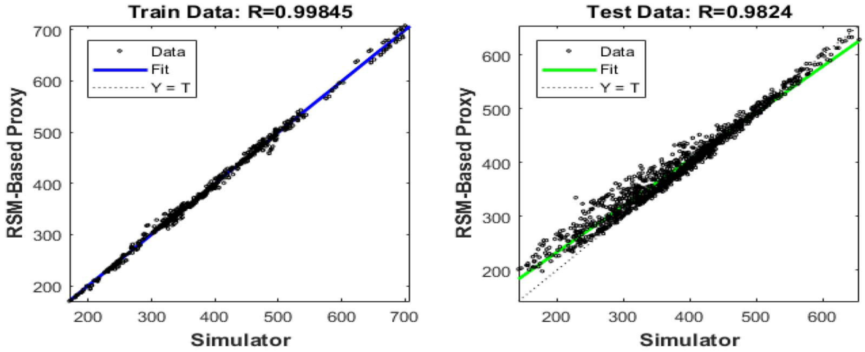

Fig. 3 Performance of the RSM-based Proxy in Phase 2 for BHP (kgf/cm²).

Development of a Proxy Model Represented by Fixed, Non-fixed Variables and by the most Impact Variable - multiple Response Surfaces

Variables (QL, BSW, lengths of the production column, and the flowline) were tested by verifying the difference between the BHP of the Proxy model and the MFM. For each variable, different operational ranges were tested; the other variables had their operating ranges maintained in the initial condition (in the case of creating a single response surface), and the variable QL promoted greater differences in several operational ranges.

Fig. 4 shows the predicted BHP values from the RSM-based Proxy model and the simulation results from MFM in Phase 3 to check model performance. Specific response surfaces were then generated from the choice of the QL variable between 3 and 4 intervals. These response surfaces obtained suitable determination coefficients ($R^2 = 1$) and validation between models ($R^2>0.97$), as shown in Fig. 4. Finally, the valid Proxy model was composed of 96 response surfaces.

Development of an ANN-Based Proxy Model Represented by GHDM

This analysis started with 375 samples for the GHDM network from the original dataset, resulting in approximately 100 samples for training (a similar number used in RMS-Based Proxy model Phase 1 construction), with a total of 275 samples for evaluation. Fig. 5 shows the predicted BHP values from the ANN-based Proxy model with 375 samples and the simulation results from MFM to check model performance. This ANN resulted in coefficients of determination of training dataset $R^2 = 0.75$ and of testing dataset $R^2 = 0.71$. The ANN was not representative, as shown in Fig. 5.

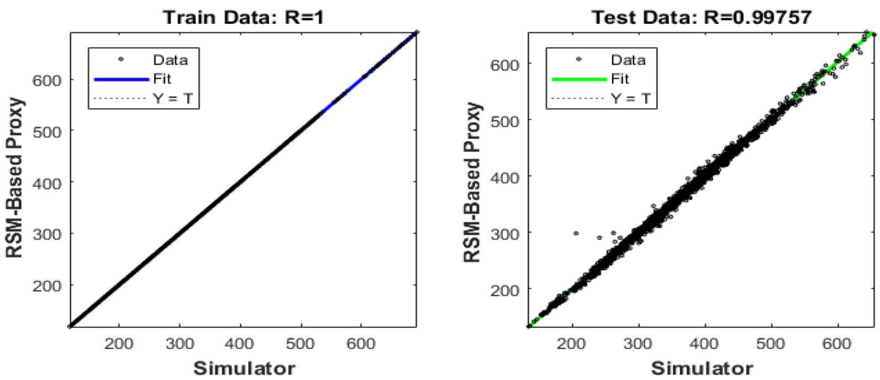

Fig. 4 Performance of the RSM-based Proxy in Phase 3 for BHP (kgf/cm²).

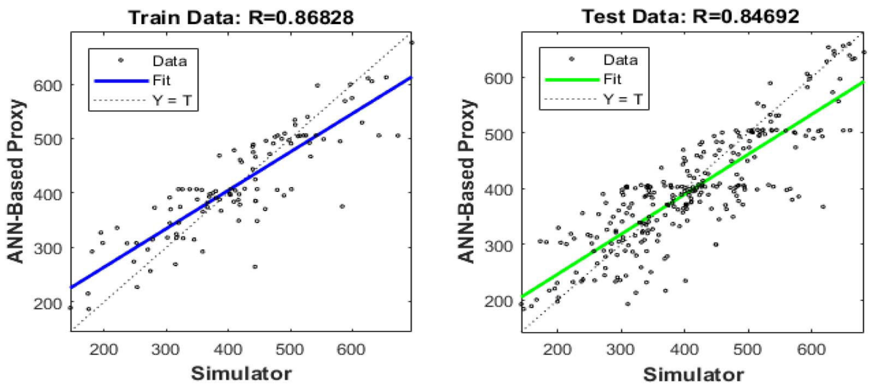

Fig. 5 Performance of the ANN-based Proxy with 375 samples for BHP (kgf/cm²).

Figs. 6 and 7 show the predicted BHP values from the ANN-based Proxy model with 3,000 and 12,000 samples, respectively, and the simulation results from MFM to check model performance. Increasing evaluation dataset size to 750, 1,500, 3,000 (shown in Fig. 6), 6,000, and 12,000 (shown in Fig. 7) samples (the total number of MFM models for production system evaluated in RMS-based Proxy model construction), obtained coefficients of determination of training dataset R^2 and testing dataset R^2 were above 0.96. The ANN was again not representative.

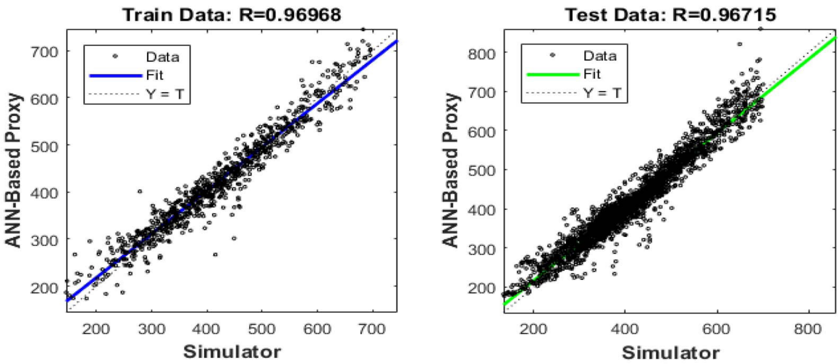

Fig. 6 Performance of the ANN-based Proxy with 3,000 samples for BHP (kgf/cm²).

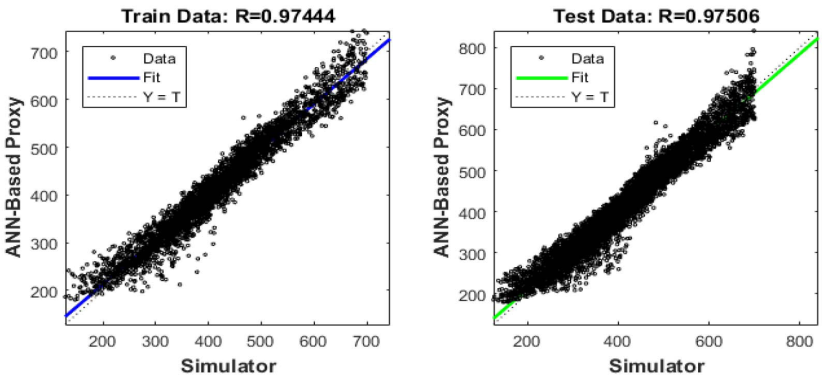

Fig. 7 Performance of the ANN-based Proxy with 12,000 samples for BHP (kgf/cm²).

Proxy Model Selection

Regarding RMSE, RMS-Proxy models obtained 47.1 for Phase 1, 21.7 for Phase 2, and 6.5 for Phase 3. In addition, ANN-Proxy models obtained 64.4 for 375 samples, 33.6 for 750 samples, 31.6 for 1,500 samples, 30.0 for 3,000 samples, 17.0 for 4,430 samples (5-inch diameter filtering), 26.4 for 6,000 samples, and 26.1 for 12,000 samples. Regarding NRMSE, RMS-Proxy models obtained 0.40 for Phase 1, 0.24 for Phase 2, and 0.07 for Phase 3. ANN-Proxy models obtained 0.54 for 375 samples, 0.29 for 750 samples, 0.27 for 1,500 samples, 0.26 for 3,000 samples, 0.11 for 4,430 samples (5-inch diameter filtering), and 0.22 for 6,000 and 12,000 samples. Phase 3 of RMS-Proxy is the best forecasting model.

As the number of MFM samples to obtain the ANN-Proxy model using GHMD was lower than that necessary to obtain the RSM-Based Proxy model, but the quality of results was inferior to acceptable, even with many MFM samples, the ANN-Based Proxy model was discarded for the continuation of this study. The RSM-Based Proxy model was selected.

Evaluation and comparison between Models by Production Strategy

Evaluation and Comparison between the MFM and the Proxy Model through a Production Strategy

In the initial test, the similarity in the NPV value obtained by each model and the production behavior (oil production) was verified alongside the BHP of the wells of the models considered. The only considerable variation occurred with water production in the Proxy model, which was superior to that obtained in the MFM. Oil production was slightly higher than that obtained with the Proxy model.

From the results in Phase 3 (proxy model represented by fixed, non-fixed variables and by the most impact variable - QL), the generated Proxy model satisfied the pre-established criteria (10%), meaning it was representative, and the comparison was made by the production strategy selected as the base case for both models. The validation parameter was the NPV, which produced a difference of 94 US$ million, or 9%, over the Proxy model (US$ 1,191 million) relative to the MFM (1,097 US$ million).

Evaluation Description in Stages of the Production Strategy

Two cycles (C1 and C2) were evaluated for MFM and Proxy model. Only the Proxy model evaluation is shown in this work. Figs. 8 and 9 show the results of evaluations of all stages of cycle 1 and 2 for the Proxy model, respectively. For E1C1 (Fig. 8), the NPV increased as the number of wells was raised (the highest NPV was 1,191 US$ million). The best result occurred with all wells in the field, that is, 12 producer and eight injector wells initially considered.

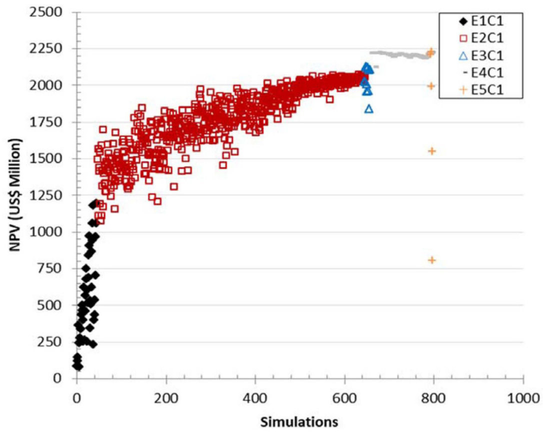

Fig. 8 Results of evaluations of all stages of cycle 1 for the Proxy model.

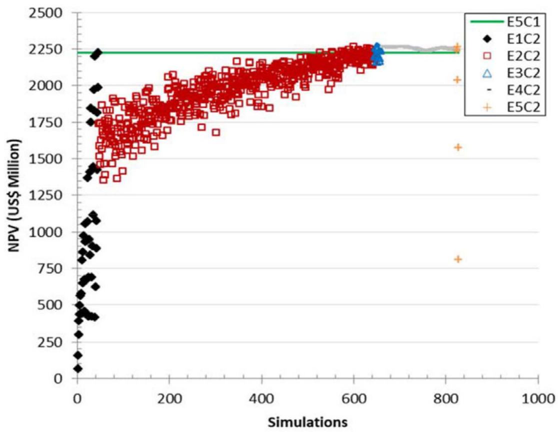

Fig. 9 Results of evaluations of all stages of cycle 2 for the Proxy model.

In E2C1, all wells had their coordinates changed, obtaining an NPV of US$ 2,070 million (74% higher than the previous stage). In E3C1, the financial return was roughly 3% higher than E2 (NPV of 2,125 US$ million).

The configuration was selected with PC, FL, and IR diameters of 5", 8" and 8", respectively, and Qgi = 0 m³/day. In E4C1, the NPV was 2,222 US$ million (5% higher than E3) with a change in the platform coordinate (X=354,800 m and Y=7,512,300 m). In E5C1, there was no increase in NPV, maintaining the capacity of 85% as the best result obtained in the previous stage (E4).

In Cycle 2 (C2) (Fig. 9) in E1, there was no increase in financial return relative to the higher value obtained in Cycle 1 (same value as NPV of 2,222 US$ million). In E2C2, NPV increased to US $ 2,266 million (2% higher than E1). This increase is due to the reorganization of the positioning of the wells, which favored the increase in NPV with changes in pipe lengths, contributing to reduced investments.

In E3C2, there was no increase in NPV. In E4C2, the increase was 0.03% with NPV of US$ 2,266 million compared to E3 (X = 354,900 m and Y = 7,512,700 m), and in E5C2, there was also no increase relative to E4, and the value was maintained.

In Cycle 2, the best result was very close to Cycle 1, indicating that the configuration obtained would have a value close to the optimum value of Cycle 1.

Comparison of Evaluation Stages and Production Behavior of Optimized Models

For the Proxy model, in Cycle 1, there was an increase in NPV of around 3%, maintaining the same diameters of the pipes as in the initial case (PC=5", FL=8" and RI=8") with only the Qgi changing from 100,000 m³/day to 0 m³/day (there is no need for artificial elevation). In Cycle 2, there was no change in NPV and the configuration, preserving what was obtained in Cycle 1. Part of these results occurred due to a partially optimized configuration with the same data used initially (diameters of PC, FL, and RI) and Qgi, only using the platform's capacity that presented better results in the MFM (85%).

Table 8 describes the result between the lowest and highest values obtained in the optimization stages of the production strategy and the difference between them through the NPV involving both cycles for the MFM and Proxy model. These results show the impact of each stage considering the maximum and minimum values obtained in each evaluation of each cycle and the difference between the extremes (D).

Table 8 NPV results of the five stages of the MFM and Proxy model between Cycles 1 and 2.

Cycle	Stage	MFM			Proxy Model		
		US$ million					
		Lowest Value	Highest Value	D	Lowest Value	Highest Value	D
1	E1	-5	1,191	1,196	-30	1,106	1,136
	E2	1,075	2,070	995	773	1,946	1,173
	E3	1,832	2,125	293	1,768	1,957	189
	E4	2,125	2,222	141	2,048	2,073	25
	E5	800	2,222	1,466	743	2,137	1,394
2	E1	64	2,222	2,158	71	2,137	2,066
	E2	1,355	2,265	910	612	2,257	1,645
	E3	2,163	2,265	102	2,114	2,257	143
	E4	2,236	2,266	30	2,195	2,257	62
	E5	810	2,266	1,456	774	2,257	1,483

For the MFM, the best response occurred without a producer well (Prod-12) of all those evaluated, and in the Proxy model, all wells were used. The coordinates resulting from the wells in each model were different. Figs. 10 and 11 show the final maps with the locations of the wells and platform position (Plat) for the MFM and Proxy models, respectively.

In the MFM, the optimized configuration had the following diameters and injection flow values: PC=5”, FL = 8”, RI=8”, and Qgi=200,000 m³/day. The diameters were the same in the optimized Proxy model, but the artificial lift method by gas lift (Qgi=0 m³/day) was unnecessary.

The resulting coordinates of the platform were different in both models (MFM with X=355,000 m and Y=7,512,350 m, and for the Proxy model, X=354,900 m and Y=7,512,700 m). The analysis of the platform capacity confirmed the same value obtained by [29] for both models (85%).

All stages had their importance, contributing to an increase in NPV and configuration changes. The results obtained in Cycle 1 had a greater impact, and in Cycle 2, with configurations closer to the optimal values, they had less impact. Both were important in the final result. There was no need for other cycles as the NPV reached the stopping criterion. Figs. 12 and 13 show the forecasting results for the optimized MFM and the Proxy model (Px) in both cycles, showing the behavior of oil (Qo) and water (Qw) flow rates in the field production, respectively. The behavior of injection water in the field is similar between the optimized MFM and the Proxy model.

The improvement in oil production in the field was similar for both models. The flow rate of water production was slightly higher in the Proxy model related to the MFM.

Cross simulations were analyzed using data from the production strategy related to the values of the input variables of the best result obtained in the Proxy model, inserted into the MFM (resulting in NPV of 2,171 US$ million). Because the MFM is a reference model (more detailed), the value of the cross-simulation with the data from the production strategy of the Proxy model in the MFM is 3.8% lower than the result of the MFM. The oil flow rate is slightly lower for the Proxy model inserted in the MFM, and the water flow rate is also much lower. This behavior affected the objective function.

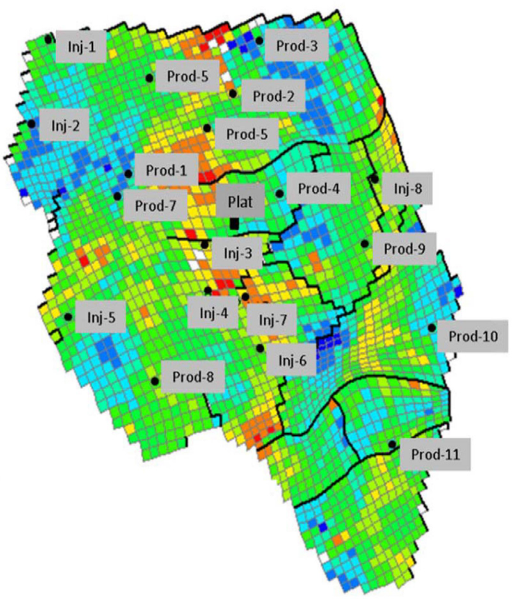

Fig. 10 Final well and platform positioning for the MFM model.

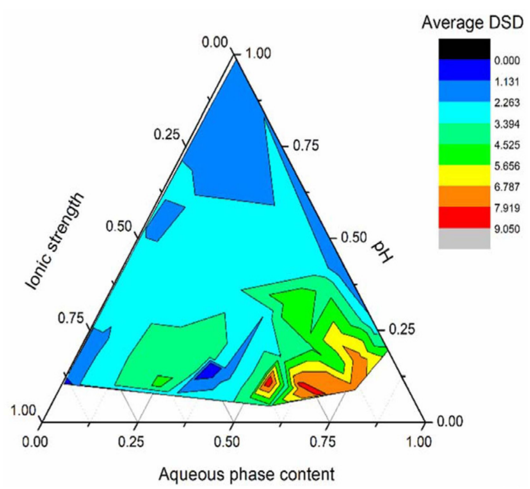

Fig. 11 Final well and platform positioning for the Proxy model.

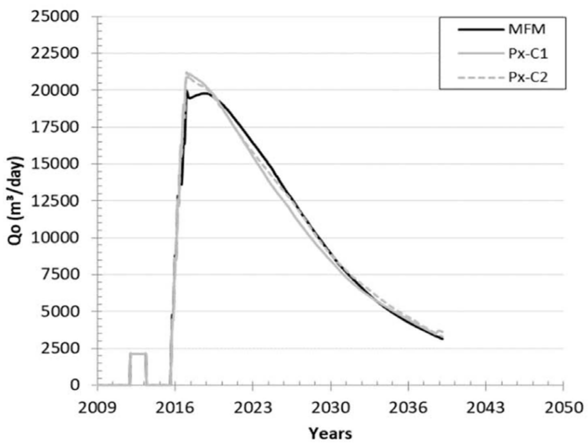

Fig. 12 Oil production flow rates (Qo) in the field after cycles of the MFM and Proxy model (Px).

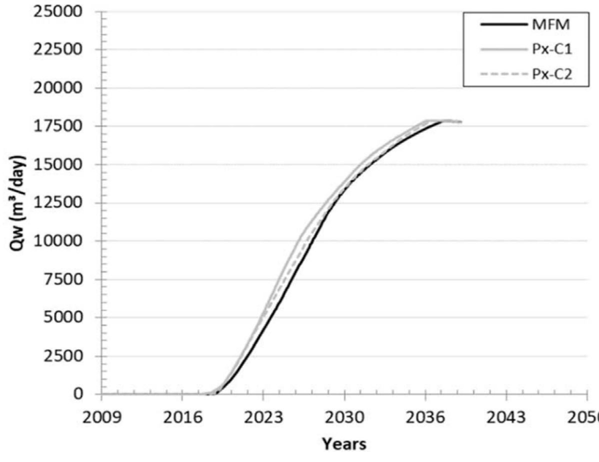

Fig. 13 Flow rates of water production (Qw) in the field after cycles of the MFM and Proxy model (Px).

Evaluation and Comparison between Computational Times between Models

Table 9 groups the average times obtained in a cycle for the Proxy model. The nomenclature for each stage of the production strategy evaluation, for a specific case, is TST, the total simulation average time for each stage; TSR, reservoir simulation average time for each stage; TSA, simulation time spent by the coupling; NSim, number of simulations; TST*, total simulation time. Ei is related to Stage i. Furthermore, each simulation in the assembly of the Proxy model took an average of 0.2 seconds.

The average simulation time of the production system (related to the Proxy model) for each stage is zero, and therefore, it is not placed in Table 7. The total optimization time of the integrated reservoir with the MFM model was 742 hours (with 1,514 integrated simulations). The integrated reservoir with the Proxy model was 518.6 hours (with two cycles plus 0.6 hours to assemble the Proxy model). The decrease in total simulated time was approximately 30%.

Table 9 Average computational times for each stage, the sum for the complete evaluation of each stage, and the production strategy of a cycle for the Proxy model.

Stages	TST (s)	TSR (s)	TSA (s)	NSim	TST* (h)
E1	1031	1011	34	46	14
E2	1210	1168	34	600	207
E3	1166	1149	35	12	4
E4	1200	1180	35	94	32
E5	1048	1029	35	5	2
Total				757	259

Discussion

In developing the Proxy model, the BHP behavior is a nonlinear function with complex behavior. This situation required going through the three stages of the development phase.

Concerning the Proxy model with a response surface, the behavior of BHP as a function of QL is very different between the evaluated models when the same wells are compared. The validation coefficient also reinforces this situation because it has a value below the adopted criterion. It thus shows that the Proxy model is not representative and does not reproduce a BHP profile similar to the reference model (MFM).

The applicability of the RSM-Based Proxy model was confirmed by the adopted criteria, which considers the adjustment of the model by the coefficient of determination ($R^2 \geq 0.99$) and the validation ($R^2 \geq 0.95$). In Phase 3, all requirements were satisfied for the various combinations evaluated, and the ranges of the impacting variables were analyzed. It allowed us to consider the set of response surfaces in the representative and possible Proxy model to be used instead of the MFM.

The residual, defined as the difference between the observed value and the estimated value of BHP, decreased from Phase 1 (all variables combined) to Phase 2 (continuous variables separated from discrete variables), as shown in Figs. 2 and 3. Moreover, variable separation enhanced the results for the RSM-Proxy model but increased its complexity with more response surfaces. From Phase 2 to Phase 3 (splitting ranges for the most important variable), the obtained lowest residual was adequate. The splitting evaluation introduced more complexity to the RSM-Based Proxy model development.

Concerning the ANN-Based Proxy model, the structured MATLAB implementation of GMDH could be used more easily to perform Proxy modeling than the RSM-Based Proxy model structure. It proved to be a promising technique. GMDH was useful for time-series prediction, which can be assumed as a special case of nonlinear regression and function approximation [26].

Using the GMDH network to find the production curve for each well over time, re-training the ANN after human interventions seems ideal. However, our Proxy modeling tried to represent the overall behavior of the production system along with field exploitation. The ANN-Based Proxy could not fit the quality check imposed for both Proxy models, even when the number of layers and neurons per layer of Proxy model was increased. The results showed that the number of MFM samples needed to obtain the adequate GHMD might be greater than that necessary to obtain the RSM-Based Proxy model for an acceptable Proxy model for integration. The residual for the ANN-Based Proxy model decreased with increasingse in dataset size from 375 to 12,000 samples but converged for an intermediary value (Figs. 5, 6, and 7). Even filtering the original dataset to include only 5-inch diameter samples, the GMDH network could not enhance the ANN-Proxy model quality to fit Phase 3 of the RSM-Proxy model. Changing ANN parameters also did not contribute to improving the model.

In [16], the author commented that Proxy modeling techniques strongly depended on the model's complexity, design space dimension, and input dataset quality. But, with the further increase in the non-linearity of a simulation model response, some Proxy models have outperformed others. The choice of the Proxy model type should be problem-specific. In our case, further study is needed to improve the utilization of the ANN-Based Proxy model for production systems with reservoir-integrated simulations.

In the initial test before the optimization of the production strategy, the NPV observed between the models was within the established criteria (maximum difference of 10%), which enabled the evaluation of the Proxy model in the study. This first analysis served as a basis to verify the possible applicability of the Proxy model, which was confirmed in an optimization of the production strategy, and then compare it with the reference model MFM, where the similarity between the results, both in the financial return and in the production behavior was verified.

Pipe sizing and the insertion of injection gas for the artificial lift method gas lift are important for optimizing the gathering system regarding production and financial return, mainly with little system information, low availability of materials, and system undersized [4]. According to the observed results, optimizing well placement was the most crucial stage of the process.

Regarding computational time, the Proxy model showed a decrease of around 30%, which is significant datum for evaluations with many variables and stages of a production strategy that normally require long simulation times. It is important to note that the integration methodology be used in optimization of the MFM already presented a significant reduction of time compared to the direct integration between reservoirs and production systems [17].

The use of more robust models compromises the computational time for the simulations. In this way, the Proxy model can replace complex models when they do not alter the integrity of the results relative to the reference model. Furthermore, developing the Proxy model for the production system satisfied the requirements adopted (determination and validity coefficients). Its application integrated into the reservoir model presents results close to the reference case (based on MFM) when different values of the model components are inserted and simulated.

What is important when constructing the Proxy model to replace the MFM is that it can present a financial return and the production behavior related to MFM that only poses a slight difference or with an acceptable value, depending on the desired conditions of the analyzed problem (through pre-established criteria). In addition, the reduced model should help reduce computational time and make the studies feasible. For this, the development of the Proxy model must be rigorous and representative, seeking to evaluate the variables that have the greatest influence on a response in the production system that is used to construct the response surfaces. Thus, more reliable models can be developed and used in developing oil field production and decision-making.

Conclusions

The results indicate that Proxy models can be an alternative to represent multiphase flow behavior for wells and gathering systems, including their geometrical parameters, because they presented results similar to those of the MFM (medium fidelity model) currently used. The financial return and fluid production behavior obtained from integrated simulation applied in an intensive production strategy optimization process present little difference (4%), showing that the Proxy model evaluated has good applicability. In addition, there was a reduction of roughly 30% in computational time for an integrated RSM-based Proxy model, compared to an integration methodology with MFM, which already shows time savings compared to an explicit direct integration.

The RSM-Proxy model obtained the best values of coefficient of determination ($R^2 = 1$) and root-mean-square error (RMSE = 6.5) than the ANN-Based Proxy model (0.95 and 26.1, respectively). RSM-Proxy model demanded variable separation and variable range splitting, increasing its complexity. Further studies on ANN-Based Proxy modeling are needed since they could not represent the production system response inside the validation criteria, even when increasing dataset size or changing ANN parameters.

The production behavior related to the production of oil and water in both the field and the wells also showed similar results, such as in the injection of water in the injection wells. The same was observed with the BHP of producer and injector wells.

The application of the Proxy model integrated into the reservoir model promoted an improvement in the financial return along evaluation stages and a similar production behavior relative to the MFM in all stages of the adopted production strategy.

The promising results also indicate that applying Proxy models for integration can be evaluated for other scenarios in other fields. Ultimately, a Proxy model, which represents the production system, could be integrated into the reservoir and favor faster analysis, dynamizing production strategies, and decision-making.

Nomenclature

ANN: artificial neural network
BHP: bottom-hole pressure, [kgf/cm²]
BHPinj: bottom-hole pressures of the injection, [kgf/cm²]
BSW: basic sedimentary and water [%]
C: the cycle of production strategy
Ci_o: Injected water processing capacity, [1000.0 m³/day]
Cp_o: oil processing capacity, [1000.0 m³/day]
Cp_w: water processing capacity, [1000.0 m³/day]
E: stage of production strategy
FL: flowline length, [in] or flowline
GMDM: group method of data handling
GOR: gas-oil ratio, [m³/m³]
i: tax rate
Inj: injector well
Inv_{plat}: The platform investment, [US$ millions]
j: period
ID: internal pipe diameters, [in]
IDLHC: terative discrete latin hypercube
MFM: medium fidelity model
N: number of samples
nd: linear term
N_t: total number of periods
n_w: number of wells
NCF: net cash flow
NCFj: net cash at a specific time j
Np: accumulated production of oil, [m³]
NPV: net present value, [US$ millions]
NRMSE: normalized root-mean-square error
NW: the number of wells
PC: production column length, [m] or production column
PI: productivity index [(m³/day)/(kgf/cm²)-1]
Pj: wellhead pressure, [kgf/cm²]
Prod: producer well
Px: Proxy model
Qgi: gas lift flow rate, [m³/day]
QL: liquid flow rate, [m³/day]
RI: riser length, [in] or riser
RMSE: root-mean-square error
RS: response surface
RSM: response surface methodology
S: simulations
tj: period considered
WHP: well head pressure, [kgf/cm²]
Wi: accumulated injection water, [m³]
Wp: accumulated production of water, [m³]
x: input variable of polynomial
y: output variable of polynomial
β: term of regression

Acknowledgments

This study was conducted with the support of Energi Simulation and Petrobras within the ANP R&D tax as a “commitment to research and development investments”.

The authors are grateful for the support of the Center for Energy and Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil), and the Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP/Brazil). In addition, a special thanks to CMG, Beicip, Mathworks, and Petrobras for software licenses.

References

1. Kosmala, A., Aanonsen, S. I., Gajraj, A., Biran, V., Brusdal, K., Stokkenes, A. & Torrens, R. (2003). Coupling of a surface network with reservoir simulation, SPE 84220. Richardson: Society of Petroleum Engineers, doi: 10.2118/84220-MS.
2. Rotondi, M., Cominelli, A., Di Giorgio, C., Rossi, R., Vignati, E. & Caraati, B. (2008). The benefits of integrated asset modeling: lesson learned from field cases, SPE 113831, SPE Europec/EAGE Annual Conference and Exhibition held in Rome, Italy, 9-12, doi: 10.2118/113831-MS.
3. Hohendorff Filho, J. C. V. & Schiozer, D. J. (2018a). Effect of Reservoir and Production System Integration on Field Production Strategy Selection, Oil & Gas Science and Technology, 73(1-15), doi: 10.2516/ogst/2018042.
4. Victorino, I. R. S., Hohendorff Filho, J. C. V., Castro, M. S., Mello, S. F. & Schiozer, D. J. (2018) Influence of Well and Gathering Systems Parameters on Integrated Petroleum Reservoir and Production System Simulations, Journal of the Brazilian Society of Mechanical Sciences and Engineering, v. 40, 9(1-21), doi: 10.1007/s40430-018-1341-z.
5. Lopes, M., von Hohendorff Filho, J. & Schiozer, D. J. (2021). The effect of dynamic data adjustments in production system simulation models on oil production forecasting applied to reservoir simulation models, Journal of Petroleum Science and Technology, 11(1), 17-28. doi: 10.22078/jpst.2021.4339.1701.
6. Gaspar, A. T., Barreto, C. E., Muñoz Mazo, E. O., and Schiozer, D. J. (2014). Application of assisted optimization to aid oil exploitation strategy selection for offshore fields, SPE Latin America and Caribbean Petroleum Engineering Conference, doi: 10.2118/169464-MS.
7. Silva, L. M., Avansi, G. D. and Schiozer, D. J. (2020). Development of proxy models for petroleum reservoir simulation: a systematic literature review and state-of-the-art, International Journal of Advanced Engineering Research and Science, 7(10), 36–62. doi:10.22161/ijaers.710.5.
8. Maschio, C., Avansi, G. D., Silva, F. B. M. and Schiozer, D. J. (2022). Data assimilation for uncertainty reduction using different fidelity numerical models, Journal of Petroleum Science and Engineering, 209, 2022, 109851, ISSN 0920-4105, doi: 10.1016/j.petrol.2021.109851.
9. Shields, A., Tihonova, S., Stott, R., Saputelli, L. A., Haris, Z. & Verde, A. (2015). Integrated Production Modelling for CSG Production Forecasting, Society of Petroleum Engineers, doi:10.2118/176881-MS.
10. Mohaghegh, S. D. & Abdulla, F. A. S. (2014). Production management decision analysis using ai-based proxy modeling of reservoir simulations – a look-back case study, Society of Petroleum Engineers, 27, doi: 10.2118/170664-MS.
11. Peng, C. H. & Gupta, R. (2003). Experimental design in deterministic modelling: assessing significant uncertainties, SPE 80537. SPE Asia Pacific Oil and Gas Conference, Jakarta, Indonesia, September 9-11. doi: 10.2118/80537-MS.
12. Risso, F. V. A., Risso, V. F. & Schiozer, D. J. (2007). Risk assessment of oil fields using proxy models: a case study, In: 8th Canadian International Petroleum Conference, Calgary, Alberta, Canada, 12-14. doi: 10.2118/2007-138.
13. Madeira, M. G. (2005). Comparison of techniques for risk analysis applied to petroleum field development, Campinas: Petroleum Engineering Department, Mechanical Engineering Faculty, State University of Campinas, Dissertation, 132.
14. Schiozer, D. J., Ligero, E. L., Maschio, C., Risso, F. V. A. (2008). Risk assessment of petroleum fields – use of numerical simulation and proxy models, Petroleum Science and Technology, 26, 10-11(1247-1266), doi: 10.1080/10916460701833913.
15. Risso, F. V. A., Risso, V. F. & Schiozer, D. J. (2006). Application of statistical design and proxy models in risk analysis of petroleum fields, In: Rio Oil & Gas. Rio de Janeiro, Brazil, 11-14.
16. Denney, D. (2010). Pros and cons of applying a proxy model as a substitute for full reservoir simulations, Society of Petroleum Engineers, doi:10.2118/0710-0041-JPT.
17. Hohendorff Filho, J. C. V. & Schiozer, D. J. (2019). Methodology to accelerate explicit integration between reservoir and production system simulators, 11/2019, XL CILAMCE, 1, 1-3, Natal, RN, Brazil.
18. Venkataraman, R. (2000). Application of the method of experimental design to quantify uncertainty in production profiles, SPE 59422. SPE Asia Pacific Conference on Integrated Modelling for Asset Management, Yokohama, Japan, 25-26, doi: 10.2118/59422-MS.
19. Oliveira, F. S., Gomes, D. M., Cavalcante, J. R. & Leitão, H.C. (2015). Optimizing steam injection scheduling using analytical models in a probabilistic approach, In: SPE Canada Heavy Oil Technical Conference, Calgary, Alberta, Canada. doi: doi.org/10.2118/174427-MS.
20. Camponogara, E. & Nakashima, P. H. R. (2006) Optimal allocation of lift-gas rates under multiple facility constraints: a mixed integer linear programming approach, Journal of Energy Resources Technology, 128(4), 280-289, doi: 10.1115/1.2358143.
21. Silva, T. L. & Camponogara, E. (2014). A computational analysis of multidimensional piecewise-linear models with applications to oil production optimization, European Journal of Operational Research, 232, 3(630-642), doi:10.1016/j.ejor.2013.07.040.
22. Luguesi, C., Camponogara, E., Seman, L. O., González, J. T. & Leithardt, V. R. Q. (2023). Derivative-free optimization with proxy models for oil production platforms sharing a subsea gas network, IEEE Access, 11(8950-8967), doi: 10.1109/ACCESS.2023.3239421.

23. Kohler, M., Vellasco, M., Silva, E. & Figueiredo, K. (2020). simproxy decision support system: a neural network proxy applied to reservoir and surface integrated optimization, IEEE Systems Journal, 14, 4(5111-5120), doi: 10.1109/JSYST.2020.2968239.
24. Hohendorff Filho, J. C. V. & Schiozer, D. J. (2018b). Integrated production strategy optimization based on iterative discrete latin hypercube, ECMOR XVI, 3-6, Barcelona, Spain, doi: 10.3997/2214-4609.201802213.
25. Avansi, G. D., Schiozer, D. J., Suslick, S. B., Risso, F. V. A. (2009). Assisted procedures for definition of production strategy and economic evaluation using proxy models, SPE 122298, Annual Conference and Exhibition held in Amsterdam, The Netherlands, 8–11 doi: 10.2118/122298-MS.
26. Heris, M. K. (2015). Group method of data handling (GMDH) in MATLAB (URL: yarpiz.com/263/ypml113-gmdh), Yarpiz.
27. Alexey, I. (1971). Polynomial theory of complex systems (PDF), IEEE Transactions on Systems, Man and Cybernetics, SMC-1 (4), 364–378, doi:10.1109/TSMC.1971.4308320.
28. Negash, B. M., Tufa, L. D., Ramasamy, M., & Awang, M. B. (2017). System identification based proxy model of a reservoir under water injection. Modelling and Simulation in Engineering, 1–10, doi: 10.1155/2017/7645470.
29. Victorino, I. R. S., Hohendorff Filho, J. C. V., Castro, M. S. & Schiozer, D. J. (2019). Analysis of the production of a pre-salt based carbonate reservoir through integrated simulation of reservoir and production system, International Journal of Petroleum Engineering, 3, doi: 10.1504/IJPE.2019.105645.
30. Correia, M. G., Hohendorff Filho, J. C. V., Gaspar, A. T. F. S. & Schiozer,. D. J. (2015). UNISIM-II-D: benchmark case proposal based on a carbonate reservoir, SPE LACPEC, 18-20 November, Quito, Equator, doi: 10.2118/177140-MS.
31. Santos, A. A. S., Gaspar, A. T., Hohendorff Filho, J. C. V., Correia, M. G., Santos, S. M. G. & Schiozer, D. J. (2018). Case study for field development and management - selection of production strategy based on UNISIM-II, Available in: https://www.unisim.cepetro.unicamp.br/benchmarks/files/UNISIM-II-D.pdf.
32. Beggs HD, Brill JP (1991) Two-Phase Flow in Pipes. 6ed.
33. Standing, M. B. (1947). A pressure-volume-temperature correlation for mixtures of california oils and gases, Drill, & Prod. Prac. 275.
34. Hohendorff Filho, J. C. V., Maschio, C. & Schiozer, D. J. (2016). Production strategy optimization based on iterative discrete Latin Hypercube. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 38(2473–2480),doi: 10.1007/s40430-016-0511-0.