

# Enhanced History Matching Process by Incorporation of Saturation Logs as Model Selection Criteria

Jesus Manuel Aponte<sup>[1][2]\*</sup>, Robert Webber<sup>[2]</sup>, Maria Astrid Centeno<sup>[3]</sup>, Hom Nath Dhakal<sup>[1]</sup>, Mohamed Hassan Sayed<sup>[4]</sup>, Reza Malakooti<sup>[5]</sup>

1. University of Portsmouth [Faculty of Technology], Portsmouth, United Kingdom
2. CNOOC International Ltd, Uxbridge, United Kingdom
3. London South Bank University, London, United Kingdom
4. University of Southampton [Chemical Engineering], Southampton, United Kingdom
5. Computer Modelling Group Ltd, London, United Kingdom

\* Corresponding author. E-mail address: Jesus.Aponte@myport.ac.uk

## Abstract.

Nowadays, thanks to the digital transformation of the oil and gas industry, there is a more diverse range of data available from production wells to history match reservoir simulation models. Still, conventional history matching methods continue using a limited set of observed data as validation parameters or model selection criteria. Therefore, sometimes critical factors controlling the fluid dynamic, and consequently the production forecasting are overlooked.

One of the challenges found in almost any history matching process is the precise estimation of fluid saturation and waterfront movement during the simulation period, however, despite being demonstrated that using surveillance logs into reservoir characterisation improves the model representation, not much effort has been allocated to incorporate saturation well logs into the core of the history matching workflows.

Therefore, this paper proposes a methodology for an alternative history matching process enhanced by the incorporation of a simplified binary interpretation of reservoir saturation logs (RST) as objective function. Incorporating fluids saturation logs during the history matching phase unlocks the possibility to adjust or select models that better represent the near wellbore waterfront movement, which is particularly important for uncertainty mitigation during future well intervention assessments in water driven reservoirs.

For the purposes of this study, a semi-synthetic open-source reservoir model was used as base case to evaluate the proposed methodology. The reservoir model represents a water driven, highly heterogenous sandstone reservoir from Namorado field in Brazil.

To effectively compare the proposed methodology against the conventional methods, a commercial reservoir simulator was used in combination with a state-of-the-art benchmarking workflow based on the Big Loop<sup>TM</sup> approach.

A well-known group of binary metrics were evaluated to be used as the objective function, and the Matthew correlation coefficient (MCC) has been proved to offer the best results when using binary data from water saturation logs. History matching results obtained with the proposed methodology allowed the selection of a more reliable group of reservoir models especially for cases with high heterogeneity. The methodology also offers additional information and understanding of sweep behaviour behind the well casing at specific production zones, thus revealing full model potential to define new wells and reservoir development opportunities. The methodology proposed in this paper does not compromise conventional methods used to evaluate the history matching quality, instead, it should be considered as an option for multi-parameter history matching workflows. The contribution to knowledge from this research is the methodology, which has also been applied in a real case for the model build and validation of a North Sea turbidite reservoir, with similar enhancements in history matching as presented in this publication.

**Keywords:** Reservoir modelling, Objective function, Reservoir simulation, Binary classification, Data analytics, History matching, Saturation logs, Data assimilation.

# 1. Introduction

Reservoir simulation is traditionally required as a decision-making tool in reservoir management studies, as it provides the physical representation of static characteristics and dynamic behaviour of a hydrocarbon reservoir, while considering a large number of known uncertainties. The history matched model is essential for reservoir management to assess different strategies maximising hydrocarbon recovery. History matching is the model calibration exercise, which relies on the fact that if a model can accurately reproduce the production history or observed data, it will be useful to predict the future performance of the reservoir.

One of the challenges for reservoir engineers and history matching processes is to reproduce detailed information about the fluid displacement in the porous media in order to identify the time of water breakthrough for each well and each pay zone in a production well. On that topic, Benlacheheb (2014) highlighted the advantages of incorporating additional monitoring data as part of the history matching process to allow more detailed outcome related to vertical heterogeneity of reservoir properties in the validated model.

This research aims to develop a complementary methodology to improve the results obtained from the history matching process by incorporating binary saturations logs as part of the evaluation parameters. The proposed methodology leads to a more robust and reliable reservoir model.

The objectives of the study were as follow:

- To present a complementary history matching methodology which allows the incorporation of Saturation Logs as an additional evaluation parameter.
- To evaluate available misfit calculations for binary parameters to assess the history matching quality of the different models.

The novelty of the methodology proposed in this paper lies in its flexibility to include new evaluation parameters independent of the type of data. Besides, the methodology also provides a simplification of the matching responses by transforming the original format into binary output, which opens the possibility to use more complex frameworks to accelerate the matching process such as machine learning.

## 1.1. Objective Functions

The history matching quality of a model is often expressed in terms of its global objective function. The objective function is a mathematical function that allows a

measure of the misfit between simulated and observed data (Mata-Lima, 2011). Most of the existing formulations can be written as a function of a point-by-point difference between simulated and historical data. Equation 1 shows the simplest way of calculating an objective function:

$$OF(\mathbf{P}) = f\left(\left\{Y_{i,j}^d(\mathbf{P}) - Y_{i,j}^{d,H}\right\}_{i,j,d}\right)$$

Equation 1

Where  $\mathbf{P}$  is a set of parameters defined by simulation model, the function  $f$  is a real-valued function,  $Y_{i,j}^{d,H}$  corresponds to the historical data of the response  $j$ , on the object  $i$  at a date/depth  $d$ , while  $Y_{i,j}^d$  corresponds to the simulated data of the response  $j$ , on the object  $i$  at a date/depth  $d$ .

In a conventional history matching process, all reservoir data is gathered to create a 3D model through the reservoir characterisation process, then using dynamic simulation and the observed data it is possible to use the model to predict historical results, finally model simulated results are compared to the observed data using the objective function to determine the history matching quality of the model in order to determine whether the model match is acceptable.

More holistic approaches have been introduced into geomodelling processes by unlocking the potential of assessing a significant number of possible combinations of inputs used to create the 3D model, and creating a wide range of alternative solutions (Schiozer et al., 2005; Kumar et al., 2017; Luo and Bhakta, 2020). These alternative solutions are commonly described as representative group of models. The final representative group of models are selected from a bigger equally probable ensemble which is derived from the combination of all related uncertainties in an uncertainty analysis. The selection of the representative group of models is based on their history matching quality.

The rationale of these original approaches is based on the nature of the history matching as an inverse problem optimisation which means that there is no unique solution to the problem and hence different sets of inputs could lead to almost the same outcome.

The incorporation of the geologist's interpretation data as part of the whole process provides the results with more information about the most representative reservoir models and uncertainties about static reservoir data (Landa and Horne, 1997; Wang and Kovscek, 2001).

The selection of the different objective functions used for the evaluation of model's performance and efficiency was reviewed by Mata-Lima (2011). Further findings from published data show that Root Mean Squared Error (RMSE)

as well as Mean of the Deviations (AE) are the most widely used in history matching, considering the linear nature of the parameters commonly used in the process. However, many of these deviation-based statistics differ from each other in the way that differences between observed and simulated results are evaluated.

## 1.2. Limitation of conventional objective function calculations

In conventional history matching workflows, the objective function is commonly calculated using well level production data such as production rates (oil, gas, water, or total liquid rates), gauge pressures and production ratios such as well water cut.

It is well known that for any model-built process, the more the number of key performance indicators (KPIs) the model managed to represent, the better the quality of the model. Hence, matching a reservoir model using only limited data may not be enough to define a satisfactory representation of the reservoir in order to predict its future performance (Sahni and Horne, 2006).

One of the main challenges of using well production data to validate the models is to accurately capture the correct saturation changes in individual producing zones in commingled production wells. Frequent practice evokes two methods to measure near wells water saturation changes: (1) Saturation Logs and (2) 4D seismic. The use of

4D seismic technology has positively contributed to a better interpretation of fluid displacement (Kjelstadli et al., 2005; Ferreira et al., 2015). However, this technology can lead to some difficulties which require additional algorithms and statistical analysis to predict fluid saturations. Besides, 4D seismic data is not always available and has additional economic implications on the budget. On the other hand, saturation logs are commonly obtained during regular surveillance interventions.

## 1.3. Classification Metrics

Several classification techniques have been applied in different fields of sciences depending on the nature of the problem and the classification output (binary or multi-class) (Tharwat, 2020).

### 1.3.1. The Confusion Matrix (CM)

The CM is defined as a table that allows the user to analyze results and performance of a specific algorithm which classifies data. The confusion matrix is one of the most common tools used to assess binary classifiers. A CM contains information about actual and predicted classifications done by a classification system (Provost and Kohavi, 1998). The performance of such systems is commonly evaluated using the data in the matrix. Table 1 shows an example of the confusion matrix for a two-class classifier.

Table 1 The confusion matrix for two-class classification problem

		Predicted	
		Negative	Positive
Actual	Negative	$a$	$b$
	Positive	$c$	$d$

The entries in the CM, in the context of the study presented in this paper, have the following meaning:

$a$  is the number of correct predictions for a negative instance or *True Negatives (TN)*,

$b$  is the number of incorrect predictions for the negative instance or *False Positives (FP)*,

$c$  is the number of incorrect predictions for a positive instance or *False Negatives (FN)*, and

$d$  is the number of correct predictions for the positive instance or *True Positives (TP)*.

A wide portfolio of metrics used for binary classification assessments can be found in the literature and most of these metrics are derived from the confusion matrix (CM). However, many of these metrics can be only applied to specific problems due to their biases and limitations as noted by Powers (2008).

Table 2 summarise the applicability and limitations of some of the more widely used binary metrics.

### 1.3.2. The Matthews correlation coefficient (MCC)

The MCC is a confusion matrix derived metric which was introduced by Brian W. Matthews in 1975 when comparing chemical structures (Matthews, 1975). MCC represents the correlation between the observed and predicted classifications with the advantage of overcoming problems generated for cases with imbalanced data as found by Chicco and Jurman (2020) by comparison of different confusion matrix metrics. As others confusion matrix metrics, MCC can be calculated using predicted instances of the confusion matrix (TP, TN, FP and FN).

The MCC outcomes goes from -1 to 1, where a coefficient 1 indicates a perfect prediction or perfect match, -1 represents total disagreement between

prediction and true values, and zero means no better than a random prediction. MCC is the only binary classification metric that generates a high score only if the binary predictor is able to correctly predict most positive and negative data instances.

According to previous statistical studies and data science applications (Chicco and Jurman, 2020), in most of

the cases, MCC can provide more reliable statistical results than other imbalanced binary metrics such as F-measure and Accuracy.

Although the classification metrics defined in table 2 have been widely applied for data science and engineering problems, there is not public evidence of their applicability as part of history matching objective functions.

Table 2: Metrics used for binary classification, adapted from Tharwat (2020)

Binary Metric	Key features	Application	Formulae
Confusion Matrix (CM)	CM measures the correlation between the observed and predicted data as quality of a binary response (true/false), (positive/negative)	The CM allows the application of the different metrics to correlate the data.	$M = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix}$
False Positive Rate (FPR)	FPR represents the proportion of positive cases that are incorrectly classified as positive from the total number of negative outcomes.	Also recognized as fallout and false alarm rate. This metric is not affected by imbalanced data.	$FPr = \frac{FP}{(FP + TN)}$
True Negative Rate (TNR)	TNR represents the proportion of negative cases that are properly identified as negative from the total number of negative outcomes.	It is also called specificity or inverse recall. This metric is less affected by imbalanced data.	$TNr = \frac{TN}{(FP + TN)}$
False Negative Rate (FNR)	FNR represents the proportion of negative cases that are incorrectly identified as negative from the total number of negative outcomes.	It is also called miss rate or inverse recall. This metric is less affected by imbalanced data.	$FNr = \frac{FN}{(TP + FN)}$
Precision (P)	Represents the ratio of correct predictions that are relevant. When the prediction is yes, how often is it correct?	It is also called "confidence" metric. It does not consider the number of true negatives.	$P = \frac{TP}{(TP + FP)}$
Recall (R)	Measure the accuracy on the positive class. Thus, when the correct prediction is yes, how often does it predict yes?	The metric is valuable to measure the real positive cases that are predicted. The metric is represented as a rate of discovery of positive classifiers.	$R = \frac{TP}{(TP + FN)}$
F-Measure (FM)	It is the ratio of metrics Precision/Recall. It is the harmonic mean of precision and recall metrics.	It considers the ratio of True Positives to the arithmetic mean of predicted positives and real positives. This metric is sensitive to changes in the class distribution.	$Fm = \frac{2TP}{(2TP + FP + FN)}$
Accuracy (A)	Represents the ratio between correct predictions to all predictions. The best value is 1 and the worst value is 0.	The metric is not reliable for imbalanced data. It can provide an overoptimistic estimation of the classifier.	$A = \frac{(TP + TN)}{(TP + TN + FP + FN)}$

Matthew's correlation coefficient (MCC)

Represents the relation between the observed and predicted classes.

The outcome ranges from +1 to -1, +1 represents a perfect prediction and -1 total disagreement. The metric is sensitive to imbalance data.

$$MCC = \frac{A - B}{\sqrt{C * D * E * F}}$$

Where:  
 A=TP\*TN      B=FP\*FN  
 C=TP+FP      D=TP+FN  
 E=TN+FP      F=TN+FN

## 2. Methodology

In order to evaluate the advantages and quality enhancement of the proposed methodology compared with a conventional history matching objective function which only use production rates, a semi-synthetic geological model developed by UNISIM-M (Gaspar et al., 2016) was used in combination with a benchmarking workflow.

### 2.1. Benchmarking the proposed methodology

To benchmark the proposed methodology, a modified "Big Loop" workflow was used. This benchmarking

workflow differs from the original presented by Kumar (2017) as it only uses the first iteration of the loop to generate the ensemble of models and it does not iterate to improve the history matching quality of the ensembles.

The modified workflow is divided in two stages, Stage 1, or Pre-Loop, which is used to generate a base or control case to represent the real reservoir outcomes or "observed" data. In the Stage 2, a final ensemble of models is created from the base case. The final ensemble of model is later used to benchmark the proposed methodology by comparing the performance of both, conventional and enhanced proposed approaches in a model selection assessment. The modified "Big Loop" workflow is illustrated in Fig. 1

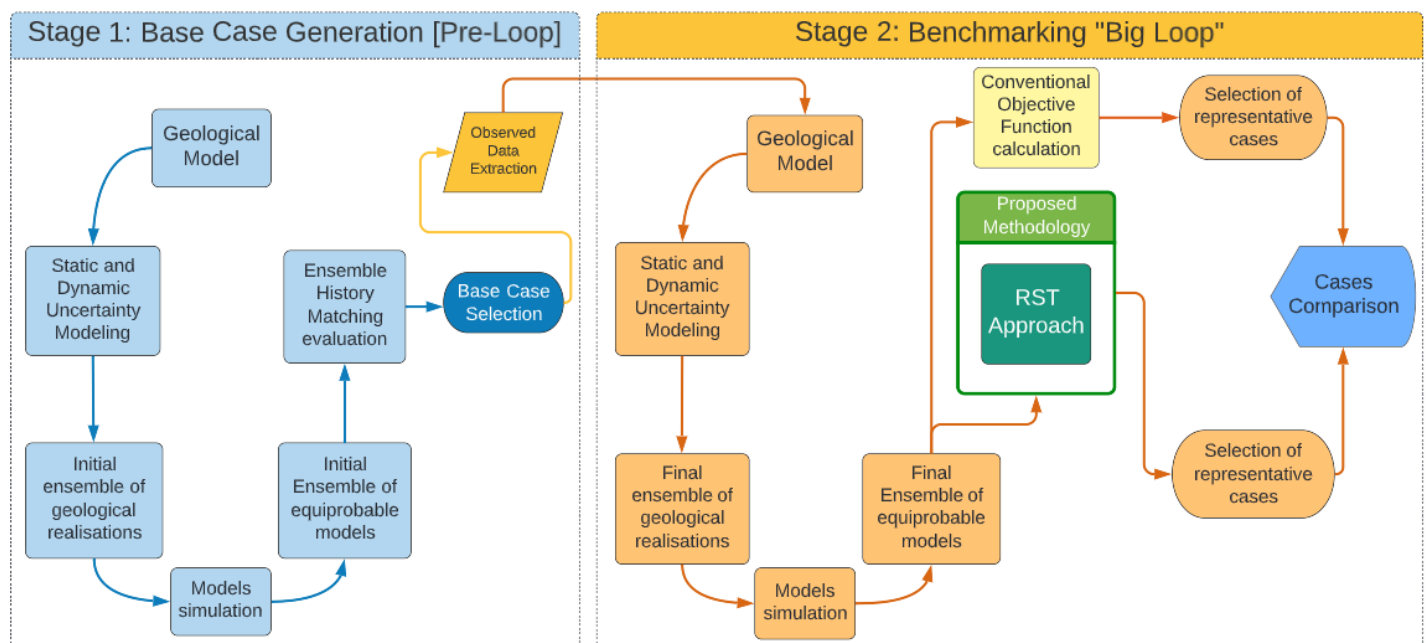


Fig. 1: Modified "Big Loop" workflow including proposed methodology

### Stage 1: Base case generation [Pre-Loop]

This stage includes the base reservoir model building (or observed data case) using all available data. Geological and dynamic inputs are analysed and included in the workflow. Variables and ranges of uncertainty are defined and analysed. The initial ensemble of models is generated as result of an uncertainty analysis loop using MonteCarlo permutation in combination with a Latin-hypercube sampling method.

After the base case is generated, key producers are identified to evaluate the methodology. In real applications, this step will depend on the data available in terms of number of saturation logs per well. For the purpose of the methodology benchmarking, all producer wells existing in the model were considered as well as one RST log per year per well.

## Stage 2: Benchmarking workflow

To generate the final ensemble of models, a second uncertainty analysis loop is performed using the same inputs and variables defined in the previous stage. The difference between stage 1 and 2 uncertainty analyses is that the observed data used in stage 2 correspond to the simulation data of the base case selected in stage 1. After the final ensemble of models is generated, both, traditional and proposed methodologies are used to select the best history matching models from the ensemble. To select the models, the traditional approach uses an objective function calculation based on producer's water cut. The proposed RST approach methodology uses binary RST logs and a confusion matrix metric to select the group of best matching cases.

To assess the history matching quality of individual cases in each group, different key performance indicators

(KPIs) such as the well and layer production rates are used to compare the selected cases against the base case.

## 2.2. Proposed RST Approach Methodology

The proposed methodology is an **add-in module** that can be used as an additional validation step in any history matching process as indicated previously in Fig. 1.

By incorporating reservoir water saturation changes, derived from cased hole saturation logs, into the history matching process; the proposed methodology adds more granularity in terms of matching zonal level water saturation changes around producer wells. The new proposed methodology for enhanced history matching process using RST logs is represented in more details in Fig. 2 and each step is explained on this section.

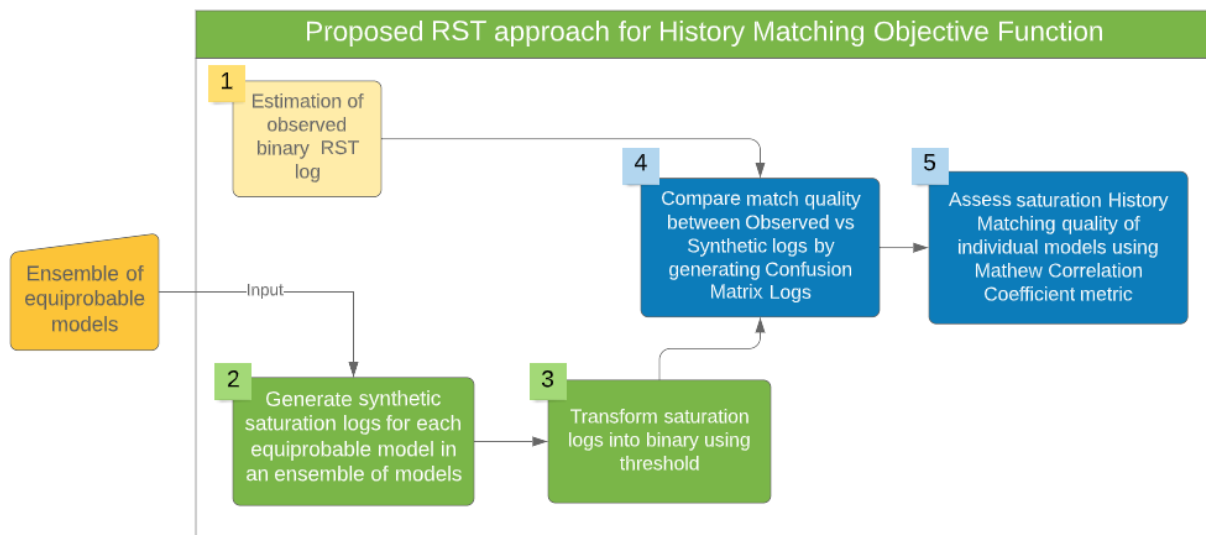


Fig. 2: Methodology proposed for enhanced history matching process using RST logs

### Step 1: Estimation of observed binary RST logs

The location and movement of the waterfront or sweep for each producer well is a key uncertainty in understanding and modelling the behaviour of an oil reservoir through production. Water saturation changes in the reservoir can be monitored by acquiring cased hole saturation logs.

The standard interpretation approach for estimating cased hole saturation changes is based on log analysis of the cased hole log data. The process requires the use of a formation evaluation model that includes rock and fluid properties to interpret the log responses, and additional parameters to model the borehole configuration. The product of this interpretation is an estimate of water saturation that varies on a continuous basis from 0 to 100% at each depth increment in the zone of interest. The results obtained from this interpretation approach contain high uncertainty, as there are many unknown parameters in the

formation evaluation model, and the data can be noisy, which represents some challenges in the use of this data as a history matching parameter.

The proposed approach to overcome some of the challenges associated with the use of saturations logs for history matching is to create an interpreted binary (yes/no) 'sweep' flag that represents the break-through of the waterfront. Although this binary sweep interpretation approach is a simplification of the normal process; it can be a valid characterisation of the saturation in several ways. The simplification reduces the grade of uncertainty created to determine a specific saturation value which is usually well defined when the waterfront has arrived.

Analysis of time-lapse saturation logs shows that for any specific depth interval, there is a single and appreciable change of water saturation from a low initial value to a high one, corresponding to the arrival of the waterfront.

This large, one-time change in water saturation is followed by little subsequent changes. Given this observed behaviour of the waterfront, it becomes reasonable to simplify the existing interpretation of the saturation log response by defining the waterfront arrival at the time and interval when the sigma log response changes. One limitation for this approach resides on the need of a previous comparable log, to define a baseline.

As the timing of the arrival of the waterfront varies at different intervals depending on reservoir properties and vertical heterogeneities, these are the key adding observations for the history matching objective function.

This method of characterisation of the ‘style of sweep’ in a reservoir is consistent with the observed behaviour of many special core analysis tests in which water breakthrough occurs in a ‘piston-like’ type behaviour.

Gradual changes in water saturation are not usually observed in core flood experiments or in actual production surveillance observation data.

In the example shown in Fig. 3, the water saturation changes in the reference case simulation model are captured at discrete times (January 2004, 2011, 2012 and 2013). These water saturation scenarios are then forward modelled to create synthetic ‘Sigma’ cased hole saturation responses. This process simulates or reproduces the observation data that would be available in an actual field development with reservoir surveillance monitoring.

Besides, in Fig. 3 the Sigma data is analysed to create a series of binary interpretation products that show the sweep response of the reservoir at the different time steps as shown in the penultimate 4 tracks.

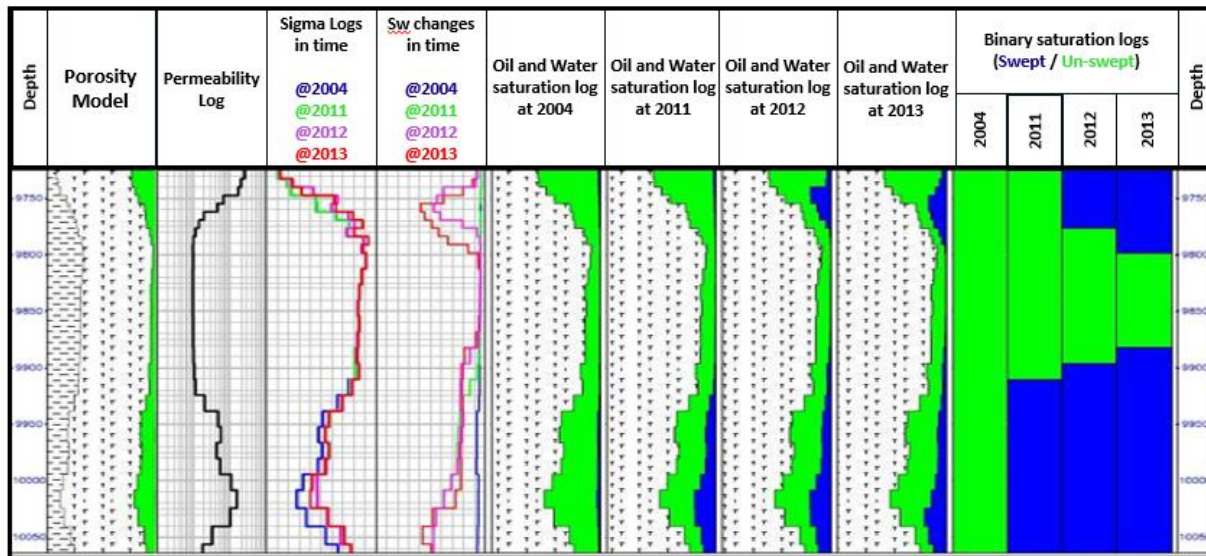


Fig. 3: Illustration of water saturation changes in the reservoir, as identified by cased hole saturation logs and interpreted in ‘sweep’ binary log

Through this process, the binary yes/no ‘sweep’ /‘un-swept’ interpretation is used to represent the most significant changes in saturation in the reservoir. Simplification of the changes in water saturation from a continuous variable to a binary variable has the additional benefit of mitigating the inherent uncertainty in the precise change in saturation, which may be unknowable from cased hole saturation logs.

For many decades both, Sigma and Carbon/Oxygen logging technology have been established to provide saturation estimates through casing. Recent improvements to this technology have reduced the uncertainty in saturation estimates. The innovative interpretation approach proposed on this research opens the possibility to use diverse sources of well logging interpretation data (including open-hole interpretations from new infill wells),

to define waterfront arrival times over decades of field development, irrespective of the data logging technology.

### Step 2: Generating synthetic saturation logs for each equiprobable model

As spatial fluid property changes are recorded for each individual grid cell, changes of water and oil saturation near the wellbore are also recorded, thus allowing the estimation of synthetic saturation well logs at any time during the simulation period. Synthetic saturation logs can capture the fluid saturation for each grid block located along the path of a specific well.

Information from saturation logs is critical in producer wells because an increase in water saturation, above a specific threshold, can be related to water breakthrough. Fig. 4 shows an example of a 2D vertical section of water saturation profile along the well trajectory of a pair of

producer and injector wells at a specific point in time during the simulation. Fig. 4 also shows a well log view of the corresponding synthetic water saturation log of the producer well at the same simulation time step. Moreover Fig. 4 also shows how injected water in the injector well

INJ017V has created preferential paths towards the producer RJS19 at the top and the bottom of the reservoir. These preferential waterfront paths are also captured as high-water saturation intervals in the producer synthetic water saturation log

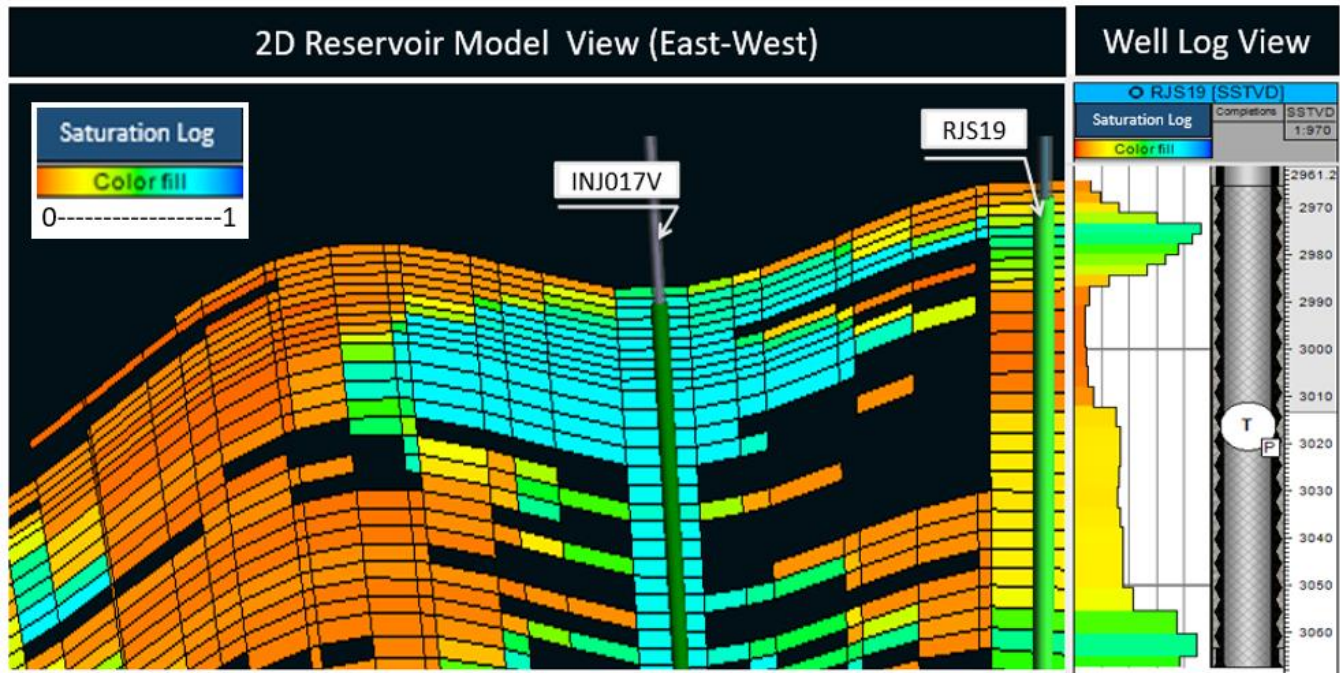


Fig. 4: 2D vertical section of water saturation profile containing the well trajectory of a producer and an injector, along with the producer synthetic water saturation log

For the proposed workflow, synthetic saturation logs are generated for all producer wells with available saturation logs, providing an additional quality check parameter used during the validation of the simulation model.

### Step 3: Transforming saturation logs into binary logs using a threshold

To transform the previously generated synthetic saturations logs into binary (“Swept”, “un-Swept”) simulation logs, a water saturation threshold is used. This threshold represents the minimum water saturation of the near wellbore cells, at the moment the waterfront arrives at the producer well.

Classifications in the binary saturation log are obtained as follow:

- 1) **Swept** class (after waterfront has arrived). A log segment is classified as swept when the water saturation in specific zones is above the threshold.
- 2) **Un-swept** class (dry production). A log segment is classified as un-swept when the water saturation has not yet reached the threshold.

**Threshold estimation.** Threshold or the value of saturation at water breakthrough can be estimated either empirically or analytically, depending on the field data available. In order to capture results impact and uncertainty ranges, both methods have been included in this research.

**a. Analytical Method.** When special core analysis (SCAL) data is available, the fractional flow curve can be used in combination with the Welge method (Wheaton, 2016) to estimate the average water saturation at the waterfront, as suggested by Welge (1975). The common procedure to estimate the average water saturation at the waterfront is by drawing a tangent line to the fractional flow curve from the initial water saturation. Fig. 5 shows an example of calculation of the water saturation at the front ( $S_{wf}$ ) using the Welge method, where  $S_w$  represents the water saturation percentage and  $f_w$  correspond to the fractional flow or water cut.



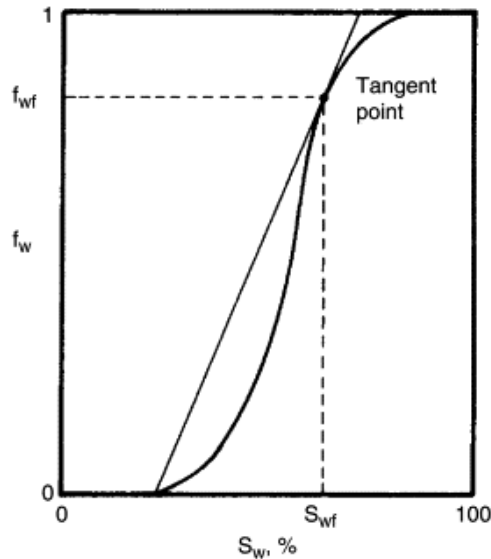


Fig. 5: Welge plot of fractional flow curve.

This method relies on the availability of core samples for the different type of rocks and SCAL results from laboratory experiments. The method is affected by reservoir characteristics, fluid and reservoir properties and pressure draw-down. The saturation determined from the Welge method is an average saturation and from the mathematical point of view, there are some limitations to determine the exactly tangent point when the fractional curve does not show appreciable changes with water saturation. Results reported by Iscan (2021) showed a good match between water saturation values estimated by fractional flow and production logging tools PLT for different rock types.

**b. Empirical Method.** As previously captured, water saturation threshold is the near wellbore water saturation at the producers when water arrives to the well, hence if a saturation log has been taken at the point of time when a specific well starts producing water, the maximum water saturation registered in the log can be used as the threshold. This method relies on the availability of water saturation logs at the time the well started producing water.

Fig. 6 shows swept and un-swept areas of a vertical section of a model, highlighting the producer well name "RJS19" and the closest injector well "INJ017V" at a specific time step, Fig. 6 also shows the corresponding synthetic binary RST log (right side), the equivalent saturation log and the applied threshold to transform saturation log into binary RST, for this example 0.2 was used.

#### Step 4: Comparing match quality between Observed vs Synthetic logs by generating a confusion matrix Log

Each binary synthetic saturation log, derived from individual models, can be directly compared with the observed binary RST log. The direct comparison process is performed using the confusion matrix classification metrics, after segmenting both observed and synthetic logs in depth units. As a result, a confusion matrix log is created for the entire log depth. Example of one model confusion matrix log and its corresponding confusion matrix table is shown in Fig. 7. Moreover, Fig. 8 shows confusion matrix logs generated for the first twenty different models of an ensemble.

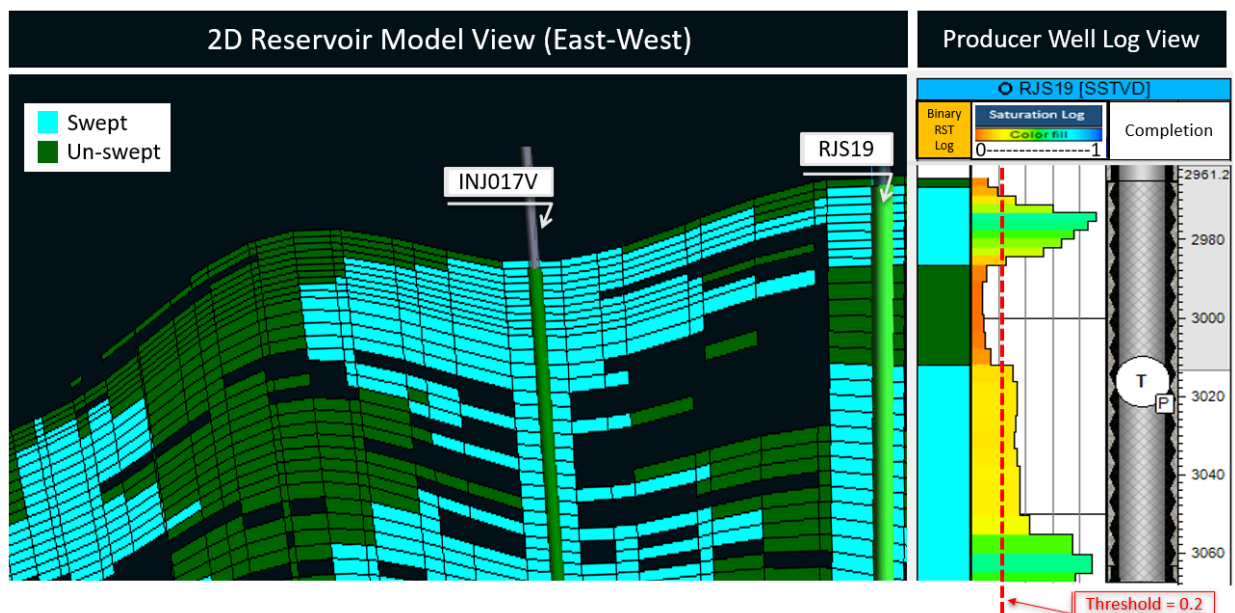


Fig. 6: Swept and un-swept areas of a 2D reservoir model slice highlighting the producer RJS19 and the closest injector at a specific time step

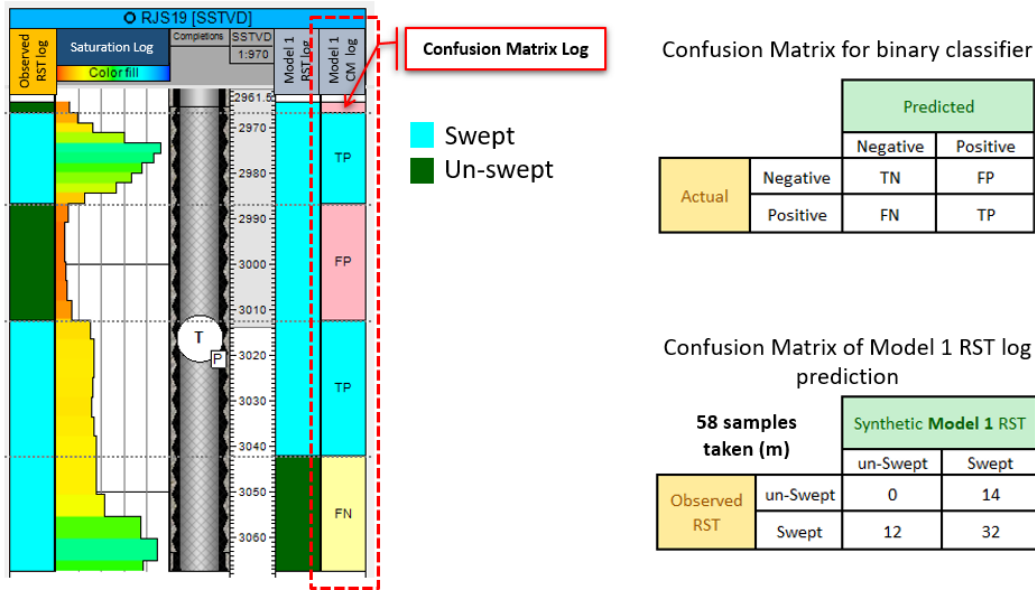


Fig. 7: Example of one model confusion matrix log and its corresponding confusion matrix table

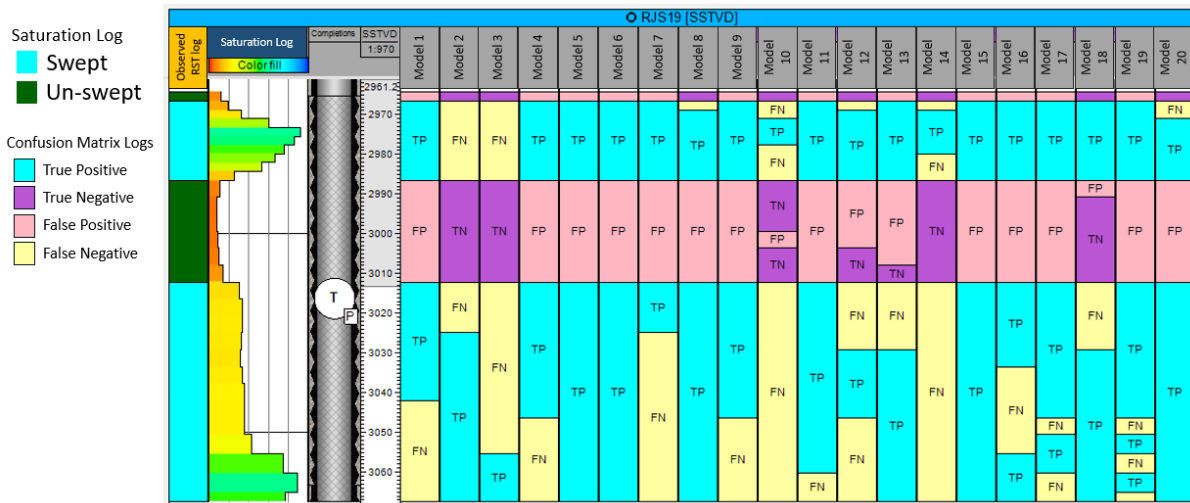


Fig. 8: Confusion matrix logs of the first 20 different models for an ensemble

### Step 5: Assessing history matching quality of individual models using binary Confusion Matrix derived metrics

During the development of the proposed methodology, the top five widely recommended confusion matrix metrics were assessed to identify the most suitable to address the binary RST mismatch. As result of this assessment, the Matthew Correlation Coefficient (MCC) was selected as the only metric that could numerically represent similarities between observed and synthetic saturation logs. Results from this analysis are included further down in the methods application section (section 3.1).

The Matthew Correlation Coefficient (MCC) metric is calculated for all the models in the ensemble. Afterwards, all models are ranked by the MCC metric results, and the best ranked model is selected.

### 2.3. Geological model used

The UNISIM semisynthetic 3D reservoir model from UNICAMP was used to evaluate the applicability of the new methodology.

The chosen geological model was a case study project developed by Cepetro educational centre (UNICAMP university, Brazil). UNISIM is a black oil semi-synthetic sector model built in a high-resolution 3D grid using public data from the Namorado Field, located in Campos Basin (offshore Brazil).

One of the main objectives of the development of UNISIM was ensuring that all relevant reservoir geological details were captured to makes UNISIM one of the best benchmarking models to evaluate new methodologies.

The reservoir is contained in an anticlinal tramp with a bottom driven active aquifer. The field is divided in three main flooding units defined by three depositional sequences separated by discontinuous sequences of marls

and poor vertical connection. The reservoir is divided in two main compartments, separated by a sealing fault with two Oil-Water contacts. The developing strategy of the field considers water injection as pressure support mechanism, with 14 producers, 11 injectors and 7 years of historical production.

A modified high resolution waterflooding UNISIM model obtained by additional modifications of the original 3D model was used to assess method feasibility and reservoir heterogeneity impact.

Modifications introduced to the original UNISIM-I-H model were aimed to increase heterogeneity and waterflooding applicability. Relevant modifications are explained as follow:

**A. Model vertical resolution increased.**

In order to incorporate more heterogeneity to the model, which allows better representation of the reservoir fluid movement, the model grid was subdivided in three separated zones and the vertical cell resolution was increased 1:2 times in all the zones. Fig. 9 shows the model before and after the modification of the vertical resolution.

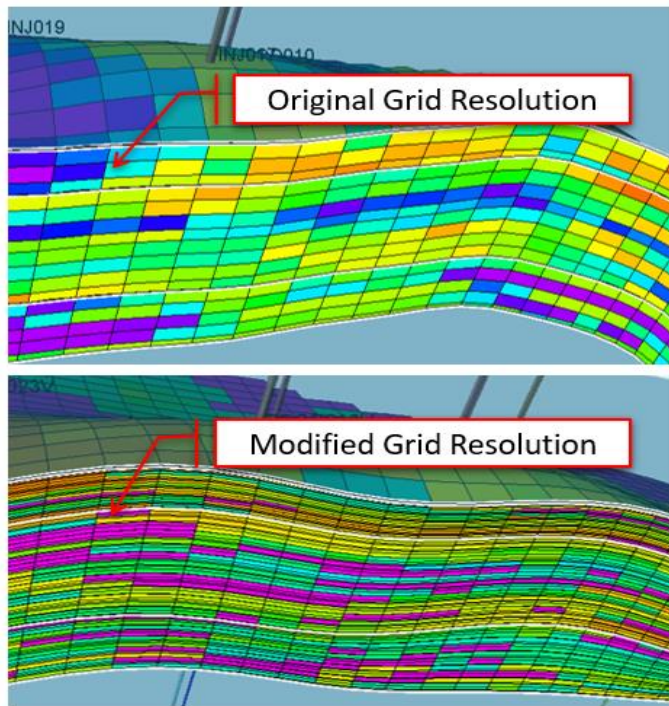


Fig. 9: 3D reservoir model before and after including vertical resolution changes

**B. Well trajectories.**

Existing well trajectories were changed from horizontal to vertical to simplify the model for the waterfront monitoring study. In order to expand the model usability and incorporate a more general view of the reservoir development, well trajectories and injection patterns were changed. The image presented in Fig. 10

shows a 2D view of the changes performed to well injectors and producers.

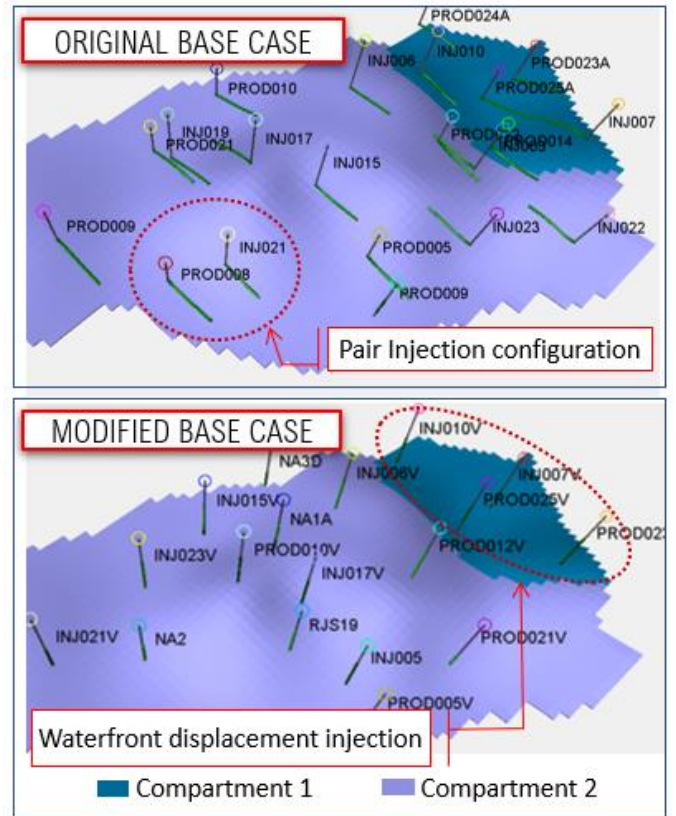


Fig. 10: Injectors and producers well trajectory modification

**C. Porosity and Net-to-Gros log correction**

After incorporating more vertical grid resolution and changing the well path to vertical, original porosity and Net-to-Gross (NTG) property logs became obsolete. As porosity and NTG logs are inputs required in the stochastic model-built workflow, missing logs were regenerated using data from vertical wells in conjunction with a neural network algorithm.

This last modification to the model also incorporated more stratigraphic resolution and better vertical sand-shale definition.

After all model enhancements were introduced, an uncertainty matrix was created considering the most relevant UNISIM model uncertainties captured by Gaspar (2016). This procedure is a widespread practice applied in reservoir modelling for data quality check, relevant variables and workflows definition, and estimation of ranges of uncertainty and sampling distributions in preparation to any uncertainty analysis.

Model modifications and uncertainty ranges defined in the uncertainty analysis were sufficient to generate a diverse initial ensemble of 160 3D models. Fig. 11 shows porosity and permeability diversity of four randomly selected models from one of the ensembles.

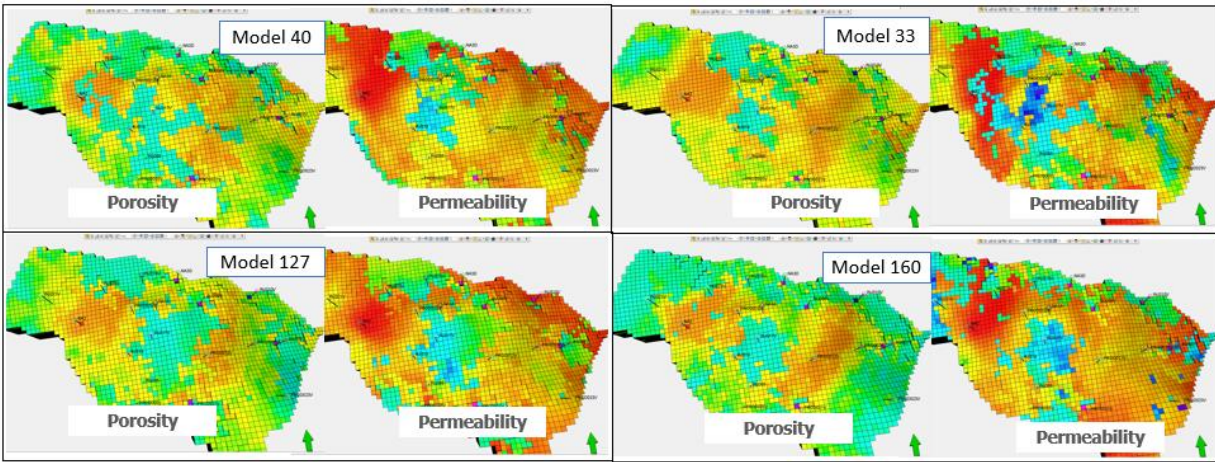


Fig. 11: Porosity and permeability diversity of four randomly selected models from the modified case ensemble

After initial ensemble was generated and the base case (case 127) was selected, a final ensemble of 200 models was generated in the second stage of the benchmarking workflow.

To mitigate the uncertainty associated to aggregation methods, one producer observation well and one RST date were identified to be used in the assessment and evaluation of classification metrics (section 3.1). The well RJS19 was selected for this assessment due to its high vertical permeability contrast and the date 08/08/2013 was selected as RST date to ensure that the waterfront has already arrived to the well at least in one zone.

Additionally, both analytical and empirical threshold calculation were used in this assessment in order to mitigate the uncertainty associated to threshold estimation. For the empirical method, the water saturation threshold of the observation well RJS19 was estimated using water saturation log and water production data from the base case. For the analytical method, synthetic UNISIM SCAL data was extracted from the base case saturation model and used in combination with Welge method to estimate the threshold.

Fig. 12 shows empirical threshold estimation using the base case water saturation log at the time of water breakthrough obtained from the water production data.

Fig. 13 shows analytical average water saturation at breakthrough using saturation model extracted from base case. After estimating the thresholds, synthetic water saturation logs were created and converted to binary logs for RJS19 well in all ensemble models as well as the base cases. Subsequently, confusion matrix logs were generated for all model cases.

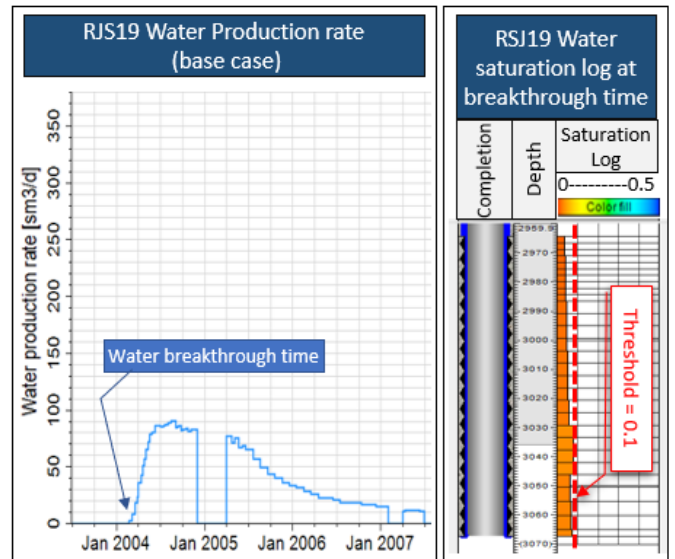


Fig. 12: Empirical threshold estimation using base case water saturation log at the time of water breakthrough

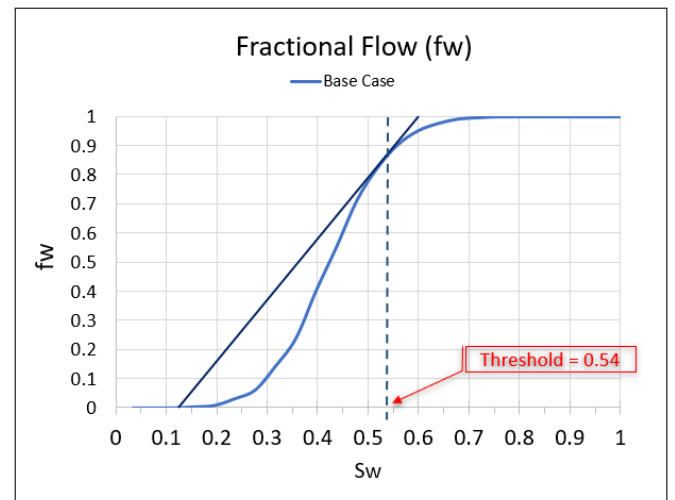


Fig. 13: Water saturation threshold using Welge method

For the comparison of proposed vs conventional history matching methodologies (section 3.2) all producers and one RST per year during the entire historical data were considered. Due to the data limitations of the Welge method highlighted in the methodology section (section 2.2), empirical threshold calculation was only used in the methodology's comparison.

### 3. Method application and discussion

#### 3.1. Assessment and evaluation of classification metrics

This preliminary assessment incorporated the testing and evaluation of the top binary metrics derived from the confusion matrix to quantify the matching quality of saturation logs. As mentioned before, for the binary metric assessment, the observation well RSJ19 was selected to be used as representative well. As a result of this analysis the best binary metric was selected to be used with the proposed methodology.

In order to capture the threshold uncertainty impact into the metric selection, empirical and analytical thresholds estimation methods were used to transform water saturation logs into binary sweep logs for the 200 cases ensemble (defined as Group A and B, respectively). To evaluate the matching quality, individual metrics were used to calculate the RST mismatch and all cases were compared and ranked in each threshold group. After the ranking, the top ten best matching cases were selected from the ranked list to visually compare with the observed RST log for final qualitative analysis of the selected cases.

##### 3.1.1. Precision.

For the group A (empirical threshold) a complete subset of ten cases was selected using precision metrics after the 200 cases were ranked as it is showed in Fig. 14. However, the expected selection of the top ten best cases was not possible for the group B (analytical threshold) as the first 14 cases shared the maximum precision score as it is showed in Fig. 15. As a result, precision metric could not be used to select smaller subsets of cases or a best representative case as the top ten cases in group A, and top 14 cases in group B shared the maximum precision score of 1.

As has been highlighted by Powers (2008) and then corroborated in this analysis, precision metric or True Positive Rate tends to be biased by the positive category, in the RST binary context, the un-swept category.

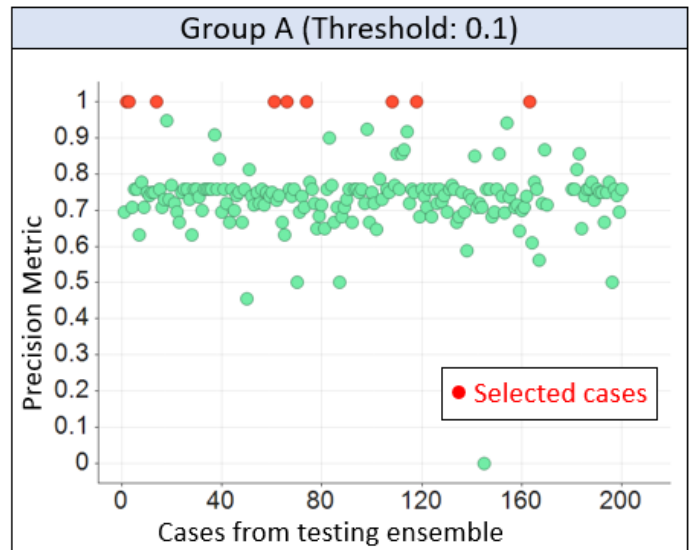


Fig. 14 Precision scores of all 200 models in group A

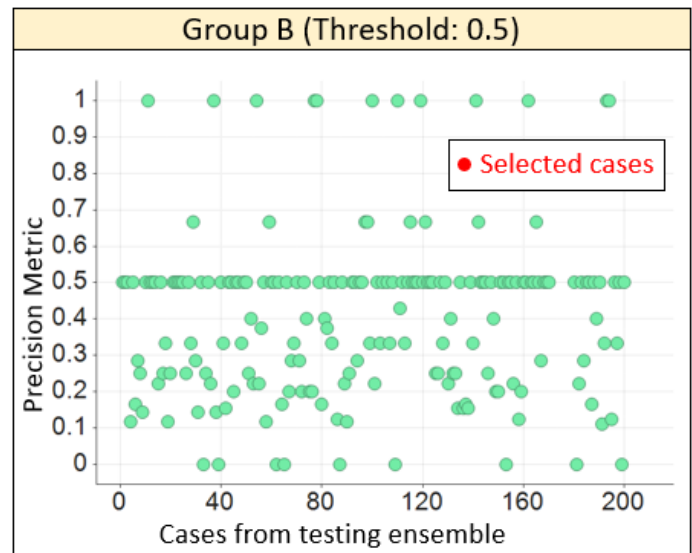


Fig. 15 Precision scores of all 200 models in group B

##### 3.1.2. Accuracy, F-Measure and Recall

Like when using the Precision metric, it was not always possible to select a group of ten cases when using Accuracy, F-Measure or Recall metrics. For some of them a maximum group of six cases was able to be selected, however it was a norm along these metrics that most of the cases in the ensemble tend to share the same calculated value, which makes impossible to rank the models and select a representative case.

##### 3.1.3. MCC.

Opposite to the other assessed metrics, a full subset of ten cases were selected using the MCC metric in the group A, and the difference in case scores in the MCC subset allowed the selection of a representative case with the best matching quality (case 37). This metric also allowed the selection of a reduced group of cases if needed. In the group B only six cases were able to be selected, however

MCC showed consistency on selecting the same top ranked case in both groups.

MCC metric score for both groups of cases are shown in Fig. 16 and Fig. 17.

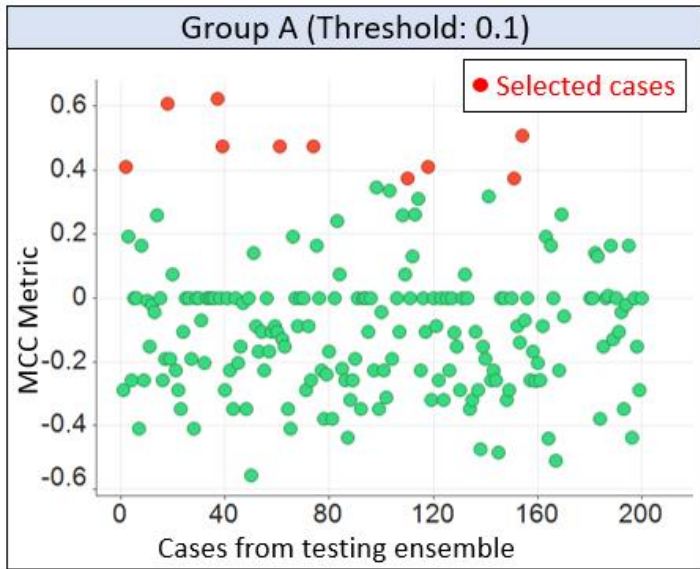


Fig. 16 MCC scores of all 200 models in group A

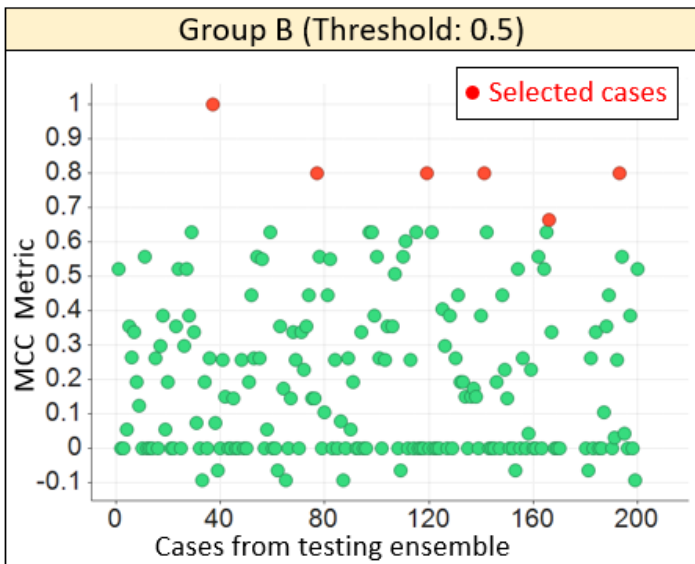


Fig. 17 MCC scores of all 200 models in group B

### 3.1.4. Binary metrics assessment summary.

A visual comparison of the observed and models generated binary logs for the top selected cases using both threshold groups is presented in Fig. 18. On this figure, dashed yellow

lines indicate boundaries between the two categories “swept” and “un-swept”. For the group A, all the selected cases using the precision metric perfectly match the un-swept category; however, the matching quality of the swept category is inaccurate. Comparable results were observed for the group B for selected cases using the recall metric. However, for this metric, all cases perfectly matched the swept category but show inconsistent results for the un-swept category matching.

Unpredicted results were also found for the accuracy and f-measure metrics when comparing both threshold groups. For group A both metrics present inaccuracies matching for both swept and un-swept categories.

For both, empirical and analytical threshold groups, Accuracy, f-measure and MCC metrics selected the same model 37 as top case with best matching quality among the entire ensemble of cases. However, MCC metric is the only metric which included cases 154 (in group A), and 166 (in group B) within the top selected models; as can be seen in Fig. 18, both 154 and 166 cases seem to be visually similar to the observed RST logs, hence they have better history matching quality than other selected cases.

MCC metric presented the most balanced results between swept and un-swept classes when compared to observed binary RST, independently of the threshold used.

This assessment demonstrated that MCC metric is the best to be used as history matching quality metric when two equally important categories are assessed. Besides, results indicate that the MCC metric can mitigate any category bias by considering all categories equally important to be matched. MCC score distinction adds extra flexibility to the model selection when incorporating additional KPIs into the selection process.

Table 3 shows a summary of selected cases by metric, colour coded to highlight the commonly selected cases between metrics, for instance, cases highlighted in green were cases commonly selected by 4 metrics, red colour cases were those commonly selected by 3 metrics, blue highlighted cases are those selected by 2 metrics, and black coloured cases were unique selected by individual metrics.



### 3.2. Comparing proposed RST methodology versus conventional history matching approach

For traditional approach, the selection of the subset of ten cases was performed using a water-cut well based objective function. The objective function was created using standardised RMSE calculation method. For the RST methodology approach, the best ten cases were selected using the MCC metric.

For this comparison, data from all producer wells were included, and for both, conventional and proposed methodologies, global misfits were calculated as cumulative error of individual well's mismatches, as captured in Equation 2.

Where,  $f$  represents the partial objective function of each producer and  $GOF$  is the global misfit of a specific case; other parameters were previously described in Equation 1. To facilitate the comparison, global misfits were normalised (0-1) in a way that cases with better history matching quality from both methodologies were those with values closer to zero.

$$\begin{aligned}
 GOF = & f_1 \left( \left\{ Y_{i,j}^d(\mathbf{P}_1) - Y_{i,j}^{dH} \right\}_{i,j,d} \right) \\
 & + f_2 \left( \left\{ Y_{i,j}^d(\mathbf{P}_2) - Y_{i,j}^{dH} \right\}_{i,j,d} \right) \\
 & + f_n \left( \left\{ Y_{i,j}^d(\mathbf{P}_n) - Y_{i,j}^{dH} \right\}_{i,j,d} \right)
 \end{aligned}$$

Equation 2

Calculation of the global misfits of the 200 cases and selection of the 10 cases with the best history matching quality was performed for both methods and results are showed in Fig. 19 and Fig. 20. Selected subsets of cases are also showed in Fig. 19 and Fig. 20.

The selection criteria for both methodologies was based on GOF values, and the objective was to select the top ten ranking cases with smaller GOF.

Cases selected by individual methods and the values of each GOF are showed in Table 4. There are two cases which were commonly selected by both methods, the top ranked case (case 37) and case 151 which was ranked differently in both subsets.

Top ranked cases highlighted in yellow in Table 4 were analysed and qualitatively compared to the base case to assess the history matching quality, hence, to judge the ability of both methodologies to select a case with the best history matching quality. Comparison was performed using well and zone water cut ratio.

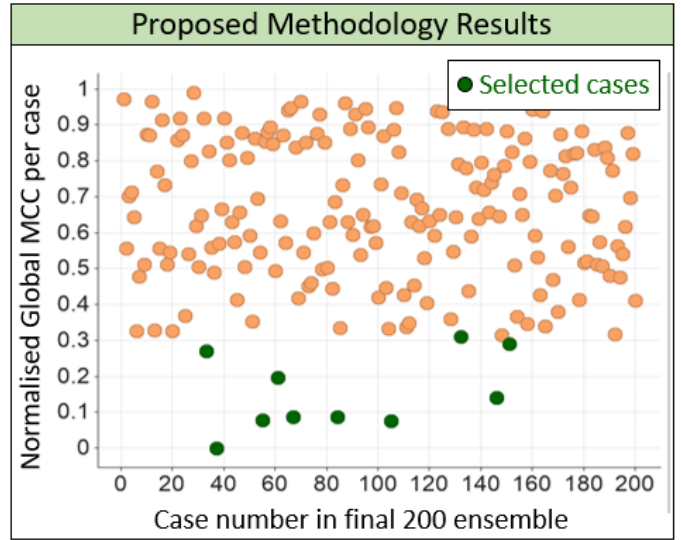


Fig. 19: GOF of 200 cases using proposed method and highlighting selected subset of cases

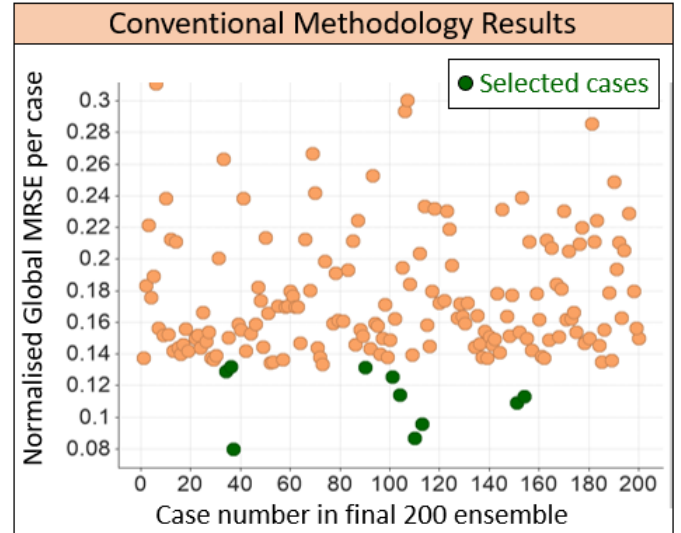


Fig. 20: GOF of 200 cases using conventional method and highlighting selected subset of cases

Table 4: Results obtained by using standard RMSE and proposed MCC metrics

Case Ranking	Proposed Approach		Standard Approach	
	Case Number	GOF Value	Case Number	GOF Value
1	37	0.0000	37	0.0800
2	105	0.0767	110	0.0871
3	55	0.0789	113	0.0958
4	84	0.0865	151	0.1093
5	67	0.0882	154	0.1131
6	146	0.1419	104	0.1142
7	61	0.1962	101	0.1257
8	33	0.2703	34	0.1292
9	151	0.2900	90	0.1315
10	132	0.3099	36	0.1323



For illustration purposes, three representative wells were selected to show history matching quality results of the top ranked cases.

For the well and zonal level history matching comparison. NA2, NA3D and PROD021V wells were selected based on their field location and stratigraphic column in order to capture different heterogeneity degrees of the reservoir properties (NA2 and NA3D have thick and good quality sands divided by a poorest intra layer sand, and PROD021V has poor vertically connected thin sands separated by small shale layers). Full

stratigraphic column of selected wells can be seen in Fig. 21. As well as the comparison of the base case RST log against RST logs of the two top ranked cases selected from the different methodologies.

As expected, near wellbore saturation of the case selected by using RST methodology (case 105) seems to have a better history matching quality.

Other wells were also analysed, and same responses were observed for all producers.

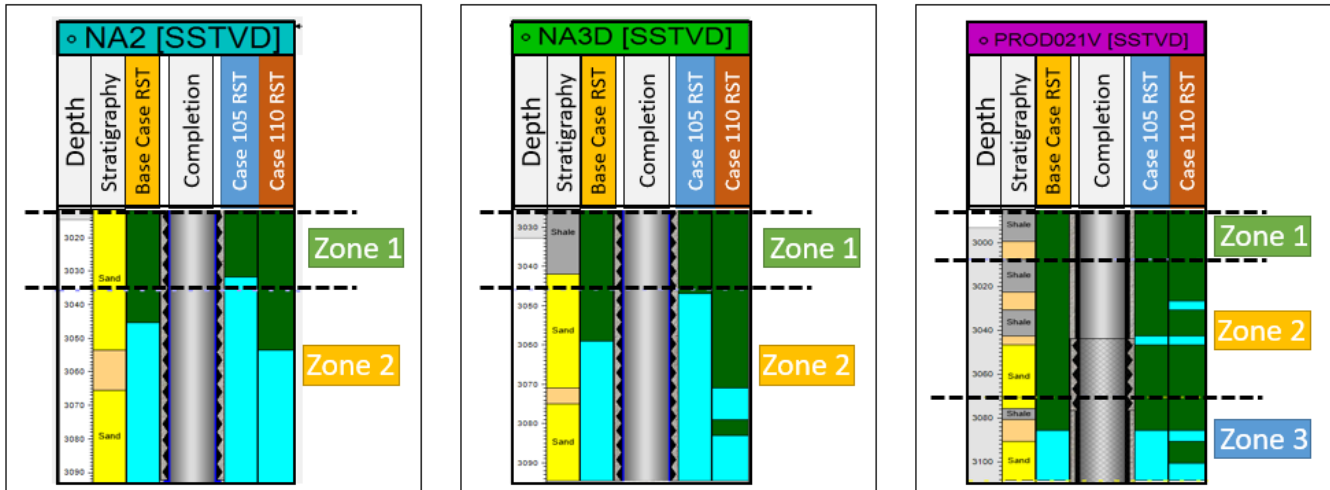


Fig. 21: Binary RST logs of top ranked cases for the representative wells

Well level water cut results from representative wells are shown in the following figures: Fig. 22, Fig. 23 and Fig. 24. From well results, both selected cases seem to have reasonable high-quality match, however proposed methodology case seems to outperform the conventional method in wells NA3D and PROD021V (Fig. 23 and Fig. 24).

As it is showed in Fig. 23, the history matching quality of well PROD021V for the case selected by the RST methodology is remarkable. As explained before, PROD021V is the well with the highest degree of heterogeneity in property distribution from the selected group of wells.

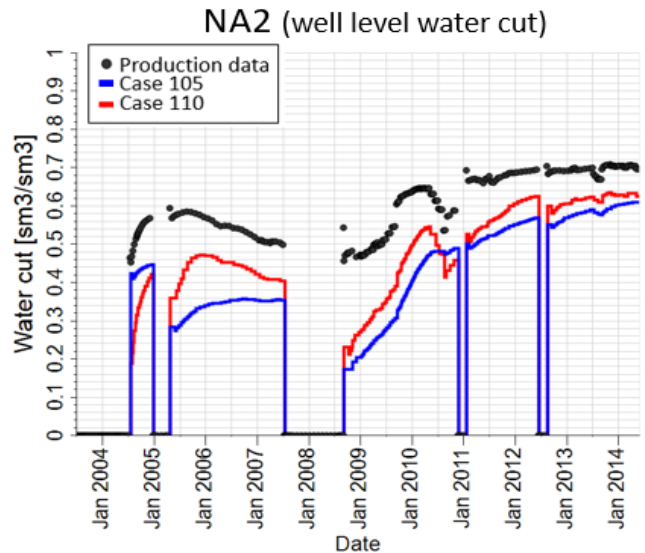


Fig. 22: Well NA2 water cut

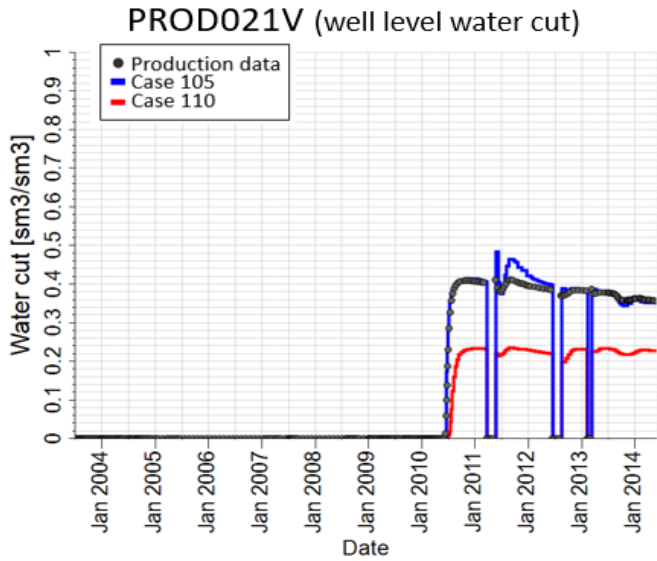


Fig. 23: Well PROD021V water cut

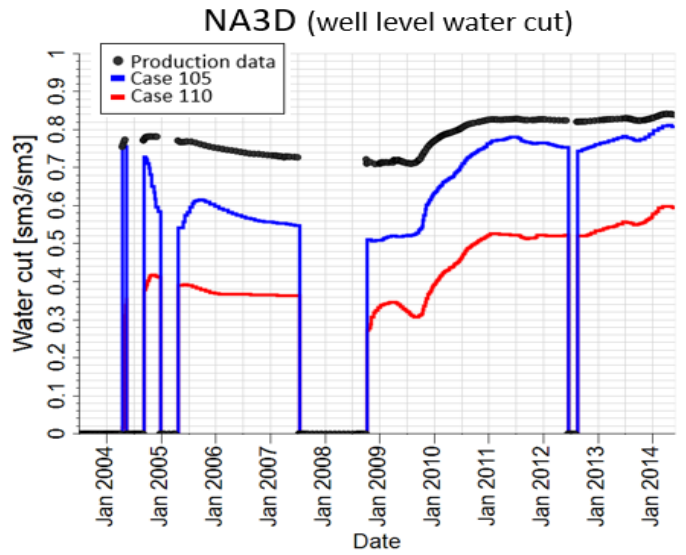


Fig. 24: Well NA3D water cut

Fig. 25 shows zonal level results of selected wells for both cases compared to the observed data. As expected, and corroborated by Figures 27 b, c and d, the proposed methodology performs better over the conventional

method when selecting cases with better zonal history matching quality. The well drilled across high heterogeneity formations (PROD021V) seems to be the one with better matching results in case 105 (proposed methodology case).

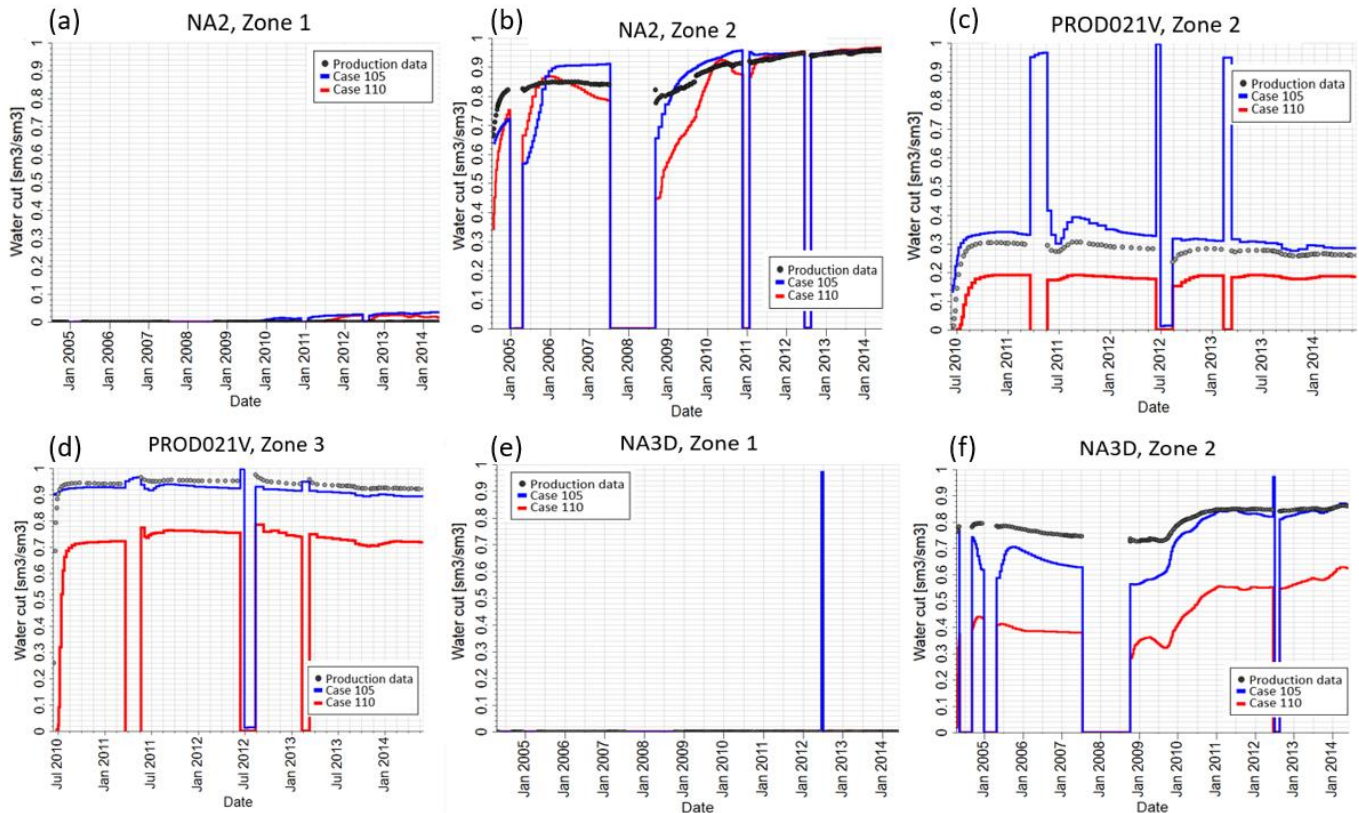


Fig. 25: Water Cut per zone for NA2, NA3D and PROD021V wells using conventional and proposed methodologies.

A global analysis of the obtained results suggests that although the conventional method exceeds on selecting cases where the total water cut of a well is closer to the

observed data, it fails to identify the waterfront arrival at different zones as the zonal match is poor.

Depending upon model objectives, having a better representation of the actual zonal and intra-zonal water

displacement could potentially provide critical information for the decision-making process. For instance, if the purpose of the model is to assess the value of adding or closing well perforations to reduce water production, having a model with detailed representation of the water displacement from different zones is crucial.

## 4. Conclusions and future work

The proposed methodology has been successfully evaluated using a high-resolution 3D gridding, from public data from the Brazilian Namorado Field model, and saturation logs data to assess the history matching quality in order to understand and select models with better sweep pattern match. The main conclusions are summarized below:

- The zonal history matching quality of a reservoir model can be measured using statistical analysis metrics derived from the confusion matrix, such as the Matthew Correlation Coefficient MCC.
- MCC metric can mitigate any category bias as it considers all categories as equally important, allowing the selection of more category balanced groups of cases.
- The use of binary RST logs as matching parameter improves the selection of models with better history matching quality at zonal/sands level.
- Reservoir heterogeneity seems to play a significant role when selecting different history matching methodologies. Thus, the use of RST logs as matching parameter contributed the most in highly heterogeneous waterflooded reservoirs.
- As expected, the quality of the match when using RST as a matching criteria is highly correlated to the number of RSTs available per well. Approach is less effective when scarce data is available.
- The history matching quality by well using RST logs for the selected cases relies on the number of RST samples available and the dates when they were taken. Having a better distributed set of RST logs along the history will positively impact selecting a model with better matching quality throughout the entire historical period.
- Further studies should be performed when merging both conventional and proposed methodology for history matching processes.

- The proposed methodology using binary classifiers as history matching parameters opens the application of other types of high uncertainty interpreted data such as 4D seismic, well tracer concentration or production logs between others.

## Credit authorship contribution statement

Jesus Manuel Aponte: Conceptualisation, Methodology, Validation, Formal analysis, Investigation, Data processing, Data curation, Writing – Original Draft, Visualisation.

Robert Webber: Conceptualisation, Methodology, Formal analysis, Investigation, Data processing, Writing – Original Draft, Visualisation.

Dr Maria Astrid Centeno: Methodology, Investigation, Data processing, Writing – Original Draft.

Professor Hom Nath Dhakal: Supervision, Writing – review & editing.

Professor Mohamed Hassan Sayed: Supervision, Writing – review & editing.

Dr Reza Malakooti: Supervision, Writing – review & editing.

## Acknowledgements

This paper and the research behind would not have been possible without the remarkable support of many individuals and organizations, among those we would wish to particularly thank Portsmouth and London South Bank Universities for providing technical support and resources. We would also like to extend our sincere gratitude to CNOOC International limited as well as all CNOOC International's peers and colleagues which one way or another have contributed with knowledge and experience towards the completion of this research, amongst them we would like to especially thank Azra Kovac, Jeremy Rhodes, Maria Trujillo, Murdo Paterson, Clive Sirju and Ben Fletcher for their outstanding expertise, guidance and experience.

## Nomenclature

K: absolute permeability

Kr: relative permeability of the fluids in the reservoir rock

$\mu$  : fluid viscosity

B: fluids formation volume factor

P: reservoir pressure,

$q$  : fluid flow rate,

$\phi$  : rock porosity,

S: fluids saturation,

x: is the coordinate along the x-axis and

$x_0$  : the initial point,

t: time coordinate.

n: number of cells

$p_a$ : Constant initial pressure

RST: Reservoir Saturation Tool

OF(P): Objective function for a set of the parameters (P)

AE: Mean of the Deviations

RMSE: Root Mean Squared Error  
 RMS: Root Mean Square  
 SSR: Sum of square residual  
 SMS: Sum of mean square  
 SAR: Sum of absolute residuals  
 WRMS: Weighted root mean square  
 MARE: Mean average relative error  
 RE: Relative error  
 CRM: Coefficient of residual mass  
 EF: Efficiency function  
 CM: Confusion Matrix  
 MCC: Matthew's correlation coefficient  
 P: Precision  
 R: Recall  
 FM: F-Measure  
 A: Accuracy

## References

- Benlacheheb, M., Al Meer, H. A., Kandil, A., Ross, F. and Rimach, M. (2014) Integration of the PLT and PBU Data in Permeability Modeling Workflow, in: *International Petroleum Technology Conference*, International Petroleum Technology Conference.
- Bertolini, A. C. and Schiozer, D. J. (2011) Influence of the objective function in the history matching process, *Journal of Petroleum Science and Engineering*, 78 (1), pp. 32-41.
- Chicco, D. and Jurman, G. (2020) The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, *BMC Genomics*, 21 (1), pp. 6.
- Ferreira, C. J., Davolio, A., Schiozer, D. J., Vernon, I. and Goldstein, M. (2015) Use of emulator and canonical correlation to incorporate 4D seismic data in the reduction of uncertainty process, in: *EUROPEC 2015*, Society of Petroleum Engineers.
- Gaspar, A. T., Avansi, G. D., Maschio, C., Santos, A. A. and Schiozer, D. J. (2016) UNISIM-IM: Benchmark case proposal for oil reservoir management decision-making, in: *SPE Trinidad and Tobago Section Energy Resources Conference*, Society of Petroleum Engineers.
- Iscan, A. G. (2021) Water saturation calculation using fractional flow and production logging data in a caspian region sandstone petroleum reservoir, *Journal of Petroleum Science and Engineering*, 200 , pp. 108355.
- Kjelstadli, R. M., Lane, H. S., Johnson, D. T., Barkved, O. I., Buer, K. and Kristiansen, T. G. (2005) Quantitative history match of 4D seismic response and production data in the Valhall field, in: *Offshore Europe*, Society of Petroleum Engineers.
- Kumar, S., Wen, X., He, J., Lin, W., Yardumian, H., Fahruri, I., Zhang, Y., Orribo, J. M., Ghomian, Y. and Marchiano, I. P. (2017) Integrated Static and Dynamic Uncertainties Modeling Big-Loop Workflow Enhances Performance Prediction and Optimization, in: *SPE Reservoir Simulation Conference*, Society of Petroleum Engineers.
- Landa, J. L. and Horne, R. N. (1997) A procedure to integrate well test data, reservoir performance history and 4-D seismic information into a reservoir description, in: *SPE Annual Technical Conference and Exhibition*, Society of Petroleum Engineers.
- Luo, X. and Bhakta, T. (2020) Automatic and adaptive localization for ensemble-based history matching, *Journal of Petroleum Science and Engineering*, 184 , pp. 106559.
- Mata-Lima, H. (2011) Evaluation of the objective functions to improve production history matching performance based on fluid flow behaviour in reservoirs, *Journal of Petroleum Science and Engineering*, 78 (1), pp. 42-53.
- Matthews, B. W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme, *Biochimica Et Biophysica Acta (BBA)-Protein Structure*, 405 (2), pp. 442-451.
- Oliver, D. S., Reynolds, A. C. and Liu, N. (2008) *Inverse theory for petroleum reservoir characterization and history matching*. .
- Powers, D. M. (2008) Evaluation Evaluation. in: *ECAI* , pp. 843-844.
- Provost, F. and Kohavi, R. (1998) Glossary of terms, *Journal of Machine Learning*, 30 (2-3), pp. 271-274.
- Sahni, I. and Horne, R. N. (2006) Stochastic history matching and data integration for complex reservoirs using a wavelet-based algorithm, in: *SPE annual technical conference and exhibition*, Society of Petroleum Engineers.
- Tharwat, A. (2020) Classification assessment methods, *Applied Computing and Informatics*, .
- Wang, Y. and Kovscek, A. R. (2001) No title, *A Streamline Approach for History-Matching Production Data*,