History matching of the PUNQ-S3 reservoir model using proxy modeling and multi-objective optimizations

Berihun Mamo Negash¹, Mariyamni Awang², Shiferaw Regassa Jufar³, and Aban John Robert⁴
Universiti Teknologi PETRONAS
Petroleum Engineering Department
32610 Bandar Seri Iskandar
Perak Darul Ridzuan, Malaysia
bmamo.negash@petronas.com.my¹, mariyamni_awang@petronas.com.my², shiferaw.jufar@petronas.com.my³, robertatomic3@gmail.com⁴

Abstract—Past studies have witnessed the wide application of assisted history matching for the calibration of dynamic reservoir models. Although the proposed algorithms have the potential to improve the history matching process in some synthetic cases, most of them have failed or have partially succeeded when applied to real, complex reservoirs. Thus far, identifying the most efficient optimization strategy for history matching has remained a challenging topic for research. In this paper a sequential approach is adopted whereby a reservoir model is replaced by a proxy model and multi-objective optimization algorithms are applied on misfit functions which was defined by combination of the proxy models and historical data. The proposed approach was tested on a case study involving a benchmark synthetic reservoir model with 14 years of production data. The data was freely provided by Imperial College London. The effectiveness of using individual optimization algorithms were quantified by using normalized root mean square error. The proposed approach is found to be efficient, robust and flexible.

Keywords: History Matching; Response Surface; Design of Experiment; Proxy modelling; multi-objective optimization

I. INTRODUCTION

Most often, simulation models do not reproduce the exact behavior of actual reservoirs in question which can greatly affect prediction results. This is due to uncertainties in measured or inferred subsurface parameters which are utilized while building reservoir simulation models. Thus, the need to quantify such uncertainties for a full field demands history matching. History matching is a powerful but time consuming technique and is an integral component of reservoir management practices. The ultimate goal of history matching is to minimize the difference between the observed field data and the simulation results so as to provide accurate forecast of future field performance. History matching is one of the most computationally demanding problem in reservoir simulation. The challenge is determining appropriate reservoir parameters such as porosity, permeability, etc. that can give a reasonable history match. This is a challenging inverse problem with non-unique solution. Conventionally, this problem is solved by tuning selected uncertain reservoir parameters one at-a- time, a process called manual history matching. The process is iteratively performed until an acceptable match between simulation and observation data is reached. Consequently, this trial and error approach is computationally expensive and highly dependent on the knowledge of the engineer involved. It is only favored for calibrating simple reservoir models but fall short in complex multi-dimensional reservoir problems. This study seeks to overcome this difficulty through application of proxy modelling and multiobjective optimization algorithms.

This paper proposes and validates the use of response surface method (RSM), design of experiment (DOE), proxy modeling and application of suitable optimization algorithms. A medium-sized benchmark synthetic reservoir model referred to as PUNQ-S3 reservoir model, has been used to evaluate performance of the proposed methodology. The model and its historical production data have been set as a benchmark case study and is freely provided by Imperial College London (UCL). It can be accessed in [1].

II. LITERATURE REVIEW

Reservoir engineers and research scientists have for long worked to develop viable options to calibrate complex reservoir models. Thus far, most researchers have taken advantage of recent advances in computer technology to speed up the history matching process [2]. This innovative technique has given rise to a new research area within reservoir engineering called assisted history matching (AHM). There is a growing interest on the niche area of AHM and a number of methods and results are being reported [3, 4]. Moreover, strengths and limitations are presented in a number of literatures for Example, [5]. The techniques approached history matching as an optimization problem which requires minimization of an objective function. The benefit is that multi-variable sensitivity analysis and multi-objective optimization can be carried out simultaneously [6]. This eliminates one at-a-time parameter variation thus, saves time.
Assisted history matching can be approached as a direct or an indirect optimization problem. Direct method is where the entire history matching job is solved by using a simulator. This involves optimizing set of uncertain parameters following subsequent simulation runs. This study investigates an indirect approach to assisted history matching. In this method the role of reservoir simulators is replaced with a non-linear proxy model which is a representative of regression functions relating input parameters and reservoir responses. Proxy model eliminate the burden of waiting for long simulation runs and therefore permits rapid convergence of objective function [7]. Figure 1 shows the sequence of indirect simulation runs and therefore measuring quality is difficult. This phenomenon and some attempts to remedy it are well explained in [9]. A related issue is that each updated realization (i.e. ensemble member) is a linear combination of the initial realizations, thus the ability of the EnKF to obtain a good estimate of the true solution is highly dependent on the quality of the initial ensemble, which needs to accurately reflect the uncertainty in the ensemble estimate; often a major challenge in real field applications. Moreover, EnKF is a Bayesian approach. However, [5] pointed out that Bayesian approaches to history matching problem assumes a simple single minimum exists, which is not always the case.

Zubarev, 2009 [10] have concluded that the use of proxy model as a substitute for mechanistic reservoir models is not justified and hence not recommended. However, recently an extensive article on the application of proxy models as the next generation of uncertainty tools to history matching problem is presented by [11]. The paper strongly argues that a major pitfall of proxy modelling is that it is a “black box” model and therefore measuring quality is difficult. This has resulted in less popularity of the technique by practitioners. In this article the application of proxy model together with an optimization algorithm is demonstrated using the PUNQ-S3 benchmark synthetic reservoir model.

III. METHODOLOGY

The methodology consists of replacing the reservoir model, which was built using physical laws, by polynomial proxy models which are functions of sensitive reservoir parameters. This is achieved through the use of DOE and RSM. A number of objective functions are then defined by combining the proxy models and historical production and pressure data. Finally minimum of the objective functions were obtained using optimization algorithms. Thus, the parameter values obtained are plugged in to the original reservoir model to update. The procedure can be repeated to improve the history matching process. The detail steps are as outlined in Figure 1.

A. Response Surface Method (RSM) and Design of Experiment (DOE).

RSM is a statistical modeling approach that is used to explore the relationship between multiple input variables called factors and their associated response variables. The process is demonstrated in Figure 2. The concept of RSM is widespread in many industrial designs such as optimizing chemical processes [12], maximizing strength of metals [13], etc. DOE is an important aspect of RSM that has been introduced to the petroleum industry for modeling subsurface uncertainties that affect performance of hydrocarbon reservoirs [14]. It is a sampling tool used to generate random number of experiments for simulation inputs. A properly executed DOE makes it possible to gather sufficient information on multiple reservoir parameters from only few simulation runs in order to fit a proxy model. The developed proxy can be enhanced by optimizing it with suitable algorithms to determine optimal reservoir parameters that minimize the difference between simulated and
observed responses [15]. This method is faster than using a simulator because the proxy model structures are simple but powerful polynomial functions. In fact, the speed of simulating a full field can be enhanced by replacing a reservoir simulator with proxy model that is representative of a given number of uncertain parameters. Suitable optimization algorithms can then be used to optimize responses based on proxy models, also called surrogate or meta-models.

![Diagram](image)

Figure 2: Process for the design of experiment (DOE)

The process of DOE is accomplished by utilizing different response surface modeling techniques as discussed hereafter.

Central Composite design (CCD): This is a RSM design technique for building a first order (2N) model for the output variable often called response variables without going through a third factorial design level. Three principle components of CCD are prominent and include factorial points, center points and axial or star runs.

Box-Behnken design (BBD): An equally useful RSM for the design of experiment which has the same level of display as CCD. The most important feature of Box-Behnken design is taking into consideration the location of sampling points. Unlike CCD, BBD does not considers star and corner points. The number of experiments resulting from both CCD and BBD are determined from a simple relationship shown in eqn. (1)

\[ NE = 2^N + 2N + C \]  

where:  
- \( NE \) = number of experiments  
- \( N \) = number of factors.  
- \( C \) = number of center points.

For instance, when \( N = 4 \), BBD will have \( C = 3 \) and total of 27 experiments to run while CCD will have 31 experiments + specified number of center points. Our experience has shown that CCD generates more experiments for a specified number of factors (input variables) than BBD. Hence, allowing it to capture more details in the feasible regions of the parameter search space. Other commonly used Response Surface Methods (RSM) for experimental design not discussed include Plackett-Burma, Latin Hypercube, as can obtain from the literature [14].

**B. Polynomial proxy models:**

Model of first or second order regression functions that is used to relate input variables \( x \) with model responses \( y \). The objective is to find \( \beta \) so that a relationship between \( x \) and \( y \) is established.

\[ y(x) = \beta_0 + \sum_i \beta_i x_i + \sum_i \sum_j \beta_{ij} x_i x_j + \sum_i \beta_{ii} x_i^2 \]  

where:
- \( x \) = vector of \( N \) input variables  
- \( \beta_i \) = coefficient of linear model,  
- \( \beta_{ij} \) = coefficient of bilinear model and  
- \( \beta_{ii} \) = coefficient of second order.

The resulting values of \( y \) from the proxy model are then compared with \( y \) from the measured reservoir data. Any misfit in the two results is further minimized by updating the proxy with new parameter search values or optimization with an algorithm.
C. Objective function

The objective function is an expression that is used to minimize the mismatch between predicted and observed responses. Different types of regression equations are used to quantify objective function and include: Linear Least Square, Generalized Least Square and Weighted Least Square [16]. In this study, linear least square, shown in Eqn. (3) is used to define the misfit between the proxy model and historical production data.

$$\sum_{i=1}^{N}(Y_{obs} - Y_{sim})^2$$  \hspace{1cm} (3)

Moreover, Normalized Root Mean Square Error (NRMSE) approach, presented in Eqn. 3 has been used to validate the efficacy of the proposed approach.

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{N}(Y_{obs} - Y_{sim})^2}{\sum_{i=1}^{N}(Y_{obs} - \bar{Y}_{obs})^2}}$$  \hspace{1cm} (4)

Where, N is the number of experiments, $Y_{obs}$ is the observed response, $Y_{sim}$ is simulation response and $\bar{Y}_{obs}$ is the average observed data.

D. Optimization Algorithms

Optimization is the process of fine-tuning the inputs of a system to obtain maximum or minimum of a function. The objective is to obtain optimum range of uncertain reservoir parameters that provide minimum NRMSE of the objective function. The simplicity or complexity of the problem is determined by the nature of fitness function and number of uncertain variables involved. Simple processes or functions are easily optimized through single objective-function optimization algorithms. However most real world applications are faced with multiple conflicting objective functions to be optimized. In such cases, the solution becomes complex and therefore calls for multi-objective optimization strategies [17].

Several optimization strategies have been studied in the last decades and are categorized as either gradient based (deterministic) algorithms or gradient free (stochastic) or global optimization strategies. Gradient based methods have been used for optimization problems for decades. Several studies including [18] have pointed the weakness of gradient based algorithms in high dimensional non-linear problems. The main concern is their intensity to get stuck in local minima and inability to reach multiple global minima. More so, the process requires expensive computation of the gradients in an attempt to minimize the objective function. Consequently, Gradient based algorithms provide a single output that is proximate to the initial guess of the function. Such an output might not be an optimal solution to the problem as potential responses can be found from multiple search spaces in the global optima. These algorithms include; adjoint methods, conjugate gradient, Gauss-Newton methods, Steepest Descent methods, Quadratic approximation and Marquardt’s techniques, etc as can be obtained from literatures [19, 20].

Conversely, many engineering and scientific problems deals with multiple global optima points with conflicting objective functions. This means that there is usually not a single optimal solution but set of pareto ultimate ones with tradeoffs in the objective functions. The search for appropriate pareto-optima or set of design parameters in feasible regions of the response surface pose relatively expensive task. Recent researchers [21, 22] have hinted the concept of global optimization as a powerful tool to address misfits in multi-objective functions. Such strategies are viable means to explore the entire global optima in search for the most likely matching parameters. Recent advances in global optimization include: Ensemble Kalman Filter (ENKF), simulated Annealing, Ant Colony, differential evolution, Genetic Algorithm, Global search, Goal Attainment, etc. The problem of multi-objective optimization is generally expressed as

$$\text{Minimize } (f_1(x), f_2(x), ..., f_m(x))$$
$$\text{s.t. } X \in S$$

Where x is the vector of decision variables, $f_1(x) ... f_m(x)$ are the m objective functions and S is the feasibility region. The methodology proposed in this paper applies Genetic Algorithm and Goal Attainment algorithms.

E. Multi-objective Genetic Algorithm (MOGA).

This is an evolutionary or a population-based optimization algorithm which is developed based on Darwin’s theory of evolution and natural selection. It was first developed by John Holland in the 1970s in University of Michigan. It employs three basic approaches namely; selection, recombination and mutation to complete an entire optimization process. It is among the most applied evolutionary search strategy in design optimization tasks [17]. Starting with a parent population of a defined size N, the order of genetic algorithm evaluation is illustrated in Figure 3.
Due to its derivation from biological processes and a searching behavior that mimics living species, MOGA is highly flexible and can be adopted in any simulation work that involves optimization. In essence, each individual in a parent population represents an initial set of the problem and are usually marked by genetic codes. Through the process of recombination, new offspring with binary codes (genomes) are reproduced. This new population undergoes mutation where they are subjected to a selection constraint (survival of the fittest test). Only individuals that pass the test will enter into the next generation and are considered to be highly fit and having potential to withstand further tests. The process is iteratively repeated until the most efficient and dominant parameters are selected.

F. Multi-objective Goal Attainment Algorithm (MOGAA)

MOGAA works to achieve a set of goals by constraining the design function to a set of objective functions. Goal attainment optimization problems are framed such that the objective functions can be under or over achieved [23]. This makes it possible to keep track of the cost function relative to the initial design goals. Unique part of goal attainment is that the relative degree of under or over achievement of the goals is controlled by a vector of weighting coefficients. These coefficients are obtained from the design parameters through the design of experiment. A standard expression of MOGAA is precisely formulated as in Eqn.6.

\[ y = \min_{y, x \in S} \text{such that } F_i(x_i) - \omega_i y \leq F^*_i \]

where, \( i = 1, 2, 3 \ldots m \), \( y \) is the slackness factor, \( F^* \) is the goal to be attained, \( F \) are the \( m \) objective functions and \( x \) is set of design variables.

The coefficient \( \omega \), is the weighting vector that enables the designer to measure the relative tradeoffs between the objective functions. Such design specifications provide an intuitive approach to rapidly optimize several design objective functions simultaneously. This study captures to implement and weigh the effectiveness of goal attainment method in comparison to genetic algorithm.

IV. CASE STUDY

PUNQ-S3 is a synthetic reservoir model owned by Elf exploration and production Company. The reservoir model was constructed by a group of European Research Universities/Centers and Companies under the courtesy of the European Union [24]. It was developed purposely for Production Forecasting with Uncertainty Quantification (named, PUNQ) and has become a benchmark for researchers to test new methods for history matching. The model consists of 5 layers at a top depth of 2430 m dipping at 1.5 degree and consists of 19x28x5 grid blocks. 1761 blocks are active and covers a uniform area of 180 x 180 m2. The top structure on the model is shown in Figure 4. A fault strikes through the East and South of the model while the north and west are strongly connected to a large (Carter Tracy) aquifer and bounded by a gas cap. Due to the presence of an aquifer
and a gas cap in Layer 1, no injection wells were considered. There are six production wells in this reservoir (PRO-1, PRO-4, PRO-5, PRO-11, PRO-12 and PRO-15) and production schedule was put under well flow constrains and constant monitoring. Wells are scheduled to flow for the first year to allow for well testing followed by build-up test for three years prior to resuming production.

Figure 4: Top structure of the PUNQ-S3 reservoir model

When production commenced, a period of two weeks was needed to shut-in the well for every year of production to enable pressure data gathering for surface analysis. The uncertain parameters in this model are the porosity, permeability and transmissibility which must be adjusted. The available historical production data includes WBHP, WGOR, WWCT, FOPT, FGPT and FWPT. These data are compared against simulation data as shown in Figure 5 for field cumulative productions. The interest of this study is to minimize the mismatch between the observed and simulated production rates.

Figure 5: Observed Vs Simulation model FOPT

The proposed workflow used to evaluate history matching of the synthetic PUNQ-S3 case study is shown in Figure 1. The flow chart shows a generalized procedure for assisted history matching in eight (8) subsequent steps.
A. Design of Experiment (DOE) and reservoir simulation.

DOE involves generating random values of factors (input variables) by using their initial minimum and maximum ranges. The uncertain parameters in this study are the multipliers for porosity, permeability and Transmissibility as seen in Table 1. For the 7 parameters, 84 experimental values were randomly generated from DOE as inputs for reservoir simulation.

<table>
<thead>
<tr>
<th>S/No.</th>
<th>Factors</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A:PORO (Porosity)</td>
<td>0.68</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>B:MULTPERMX (Permeability Multiplier in the X direction)</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>C:MULTPERMY (Permeability Multiplier in the Y direction)</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>D:MULTPERMZ (Permeability multiplier in the Z direction)</td>
<td>0.35</td>
<td>1.0</td>
</tr>
<tr>
<td>5</td>
<td>E:MULTX (Transmissibility multiplier in the X direction)</td>
<td>0.15</td>
<td>1.0</td>
</tr>
<tr>
<td>6</td>
<td>F: MULTY (Transmissibility multiplier in the Y direction)</td>
<td>0.15</td>
<td>1.0</td>
</tr>
<tr>
<td>7</td>
<td>G: MULTZ (Transmissibility multiplier in the Z direction)</td>
<td>0.05</td>
<td>1.0</td>
</tr>
</tbody>
</table>

The experiments were run at different time steps for each of the six (6) wells. The 15 responses for WBHP, 15 responses for WWCT and 15 responses for WGOR in each of the six wells were selected at specified time steps of the simulation. The simulation result was used to complete the DOE in order to fit a response surface (proxy model). In total, each well produced $84 \times 15 \times 3 = 3780$ responses and $3780 \times 6$ responses for all the six wells.

B. Building and validation of Proxy model.

A multidimensional non-linear regression proxy model that represents feasible regions of the search space was built through RSM and DOE. During DOE, the 7 input parameters interacted to produce 37 combination of uncertain parameters over the feasible region.

Figure 6 above shows an estimated effect of these parameters on Well bottom hole pressure at time step one (WBHP_1). As observed, there is a unique influence of each parameter on WBHP_1. Parameters in blue have a negative impact on the model as compared to those in grey. Parameters which pose less influence on the model were removed from the analysis while retaining most influential parameters as illustrated in Figure 7. The proxy was validated with the retained parameters in order to improve its accuracy for further analysis.
The proxy model is said to be accurate if it reproduces the same behavior trend as the simulation model. This is judged by quantifying its R-Squared (R2) error which should be close to unity (1) as revealed by Figure 8. Eqn. 6 is used to quantify this error.

\[
R^2 = 1 - \frac{\sum(Y_{\text{obs}} - Y_{\text{sim}})^2}{\sum(Y_{\text{obs}} - Y_{\text{obs}})^2}
\]  

(6)

C. Objective function.

With the help of the constructed proxy model, the objective function is defined prior to performing optimization process. This enables calculation of the misfit between observed and simulated data. History matching is successful when the difference between predicted and the observed responses have been reduced to a specified selection criterion. NRMSE (Equation 3) as described previously is used to quantify this difference.

D. Optimization process.

The ultimate goal of optimization is determining the values of uncertain reservoir parameters that minimize the objective function to an acceptable tolerance. In this workflow, multi-objective global optimization algorithms have been applied. The idea was to compare the effectiveness of MOGA and MOGAA as tools to find the best uncertain parameters for history matching. The best algorithm is selected based on its convergence rate and minimal objective function.

V. RESULTS AND DISCUSSION.

The proposed workflow for assisted history matching has been successfully applied with appreciable outcome. Response surface and design of experiment are the most critical attributes in the success of this work. As observed in Figure 6, the influence of the 7 input parameters and their interactions on well responses was scrutinized. Through a step-wise validation process, the accuracy of the proxy model was improved to fit simulation model. This makes it a potential driver replacing the use of reservoir simulator for global optimization of the uncertain reservoir parameters. The established proxy is cost effective, fast and can handle large number of uncertain parameters simultaneously. It was used to estimate the 7 uncertain reservoir parameters over a feasible region (response surface).
A. Proxy predicted results

Figure 9 shows a 3D desirability plot of the overall desirability function for porosity and horizontal permeability plane where a value of 1 indicate the perfect response and 0 indicate a completely undesired value. Several values of porosity, permeability and transmissibility multipliers were tested each time during the simulation to identify the best matching combinations.

An improved match was realized following successive simulation runs by employing the estimated proxy parameters (Table 2) as illustrated in Figure 10 and Figure 11. However, NRMSE % calculation revealed a significant mismatch in WBHP (Figure 12). A different approach was used to minimize this error by using reliable optimization strategies as discussed in case 2.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Simulation</th>
<th>Proxy model</th>
<th>MOGA</th>
<th>MOGAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PORO</td>
<td>0.68</td>
<td>0.84</td>
<td>0.89</td>
<td>0.916</td>
</tr>
<tr>
<td>MULTPERMX</td>
<td>0.70</td>
<td>0.85</td>
<td>0.86</td>
<td>1.107</td>
</tr>
<tr>
<td>MULTPERMY</td>
<td>0.70</td>
<td>0.85</td>
<td>0.91</td>
<td>0.953</td>
</tr>
<tr>
<td>MULTPERMZ</td>
<td>0.35</td>
<td>0.675</td>
<td>0.72</td>
<td>0.890</td>
</tr>
<tr>
<td>MULTX</td>
<td>0.15</td>
<td>0.575</td>
<td>0.56</td>
<td>0.750</td>
</tr>
<tr>
<td>MULTY</td>
<td>0.15</td>
<td>0.575</td>
<td>0.56</td>
<td>1.005</td>
</tr>
<tr>
<td>MULTZ</td>
<td>0.05</td>
<td>0.525</td>
<td>0.92</td>
<td>0.774</td>
</tr>
</tbody>
</table>

Figure 9: Estimated response surface for multiple responses
Figure 10: PRO-15 Gas-Oil ratio profile (without optimization)

Figure 11: PRO-4 Water cut production profiles (without optimization)

Figure 12: PRO-1 Bottom hole pressure profile (without optimization)
B. Optimization of the proxy models using goal attainment and genetic multiobjective algorithms

In the second case, the estimated proxy parameters in case 1 were optimized using multi-objective genetic algorithm (MOGA) and multi-objective genetic goal attainment algorithm (MOGAA). Interestingly, application of the two algorithms provided a satisfactory history matching for all the well responses (WBHP, WWCT, and WGOR) as seen in Figures 13-15. Specifically, MOGAA resulted in a 0.0019% NRMSE which is much better when compared with MOGAA which resulted in a NRMSE of 0.016%. Errors resulting from the application of either proxy model, MOGA and MOGAA were quantified using the defined NRMSE and compared to the error before history matching as shown in Table 3. It is observed that the error diminishes drastically with the introduction of optimization algorithms.

![Figure 13: Optimized PRO-15 Gas-oil ratio profile](image)

![Figure 14: Optimized PRO-4 water cut profile](image)
An integrated workflow for history matching using computer aided techniques has been demonstrated. The importance of proxy modeling for sensitivity analysis and its potential to improve history matching has been illustrated. It is proven that use of proxy model for predicting uncertain reservoir parameters is faster and more convenient than the traditional vary one at-a-time approach. Introduction of MOGA and MOGAA in the workflow provide significant improvement in the history matching results. It is hoped that the developed approach can be applicable for real field case studies and related optimization problems.

ACKNOWLEDGMENT

The authors would like to thank Elf Exploration Company and Imperial College of earth Sciences and Engineering for making PUNQ-S3 model dataset available online. The authors also thank Universiti Teknologi PETRONAS for providing access to the commercial softwares required to complete the work.

REFERENCES

5. Tavassoli, Z., J.N. Carter, and P.R. King, Errors in History Matching.

© IEOM Society International