Journal of Petroleum Science and Engineering 146 (2016) 392–399

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: www.elsevier.com/locate/petrol

An alternative proxy for history matching using proxy-for-data approach and reduced order modeling

Jincong He, Jiang Xie*, Xian-Huan Wen, Wen Chen

Chevron Energy Technology Company, Houston, Texas, USA

ARTICLE INFO

Article history:
Received 27 January 2016
Received in revised form
25 April 2016
Accepted 22 May 2016
Available online 28 May 2016

Keywords:
History matching
Mismatch function
Response surface methodology
Reduced-order modeling

ABSTRACT

Response surface methods are commonly used in history matching process to approximate the functional relationship between the input parameters and the aggregated mismatch. The quality of the proxy (accuracy in prediction) degrades as the nonlinearity of the response surface increases. However, commonly-used definitions of aggregated mismatch, such as root mean squared error (RMSE) or mean absolute error (MAE), are highly nonlinear. As a result, the quality of the proxy for aggregated mismatch can be unsatisfying in many cases.

In this work, we propose the proxy-for-data (PFD) approach, in which one proxy is built for each observation data point and then the data values predicted by those proxies are used to calculate the aggregated mismatch. Because proxies are constructed for the data themselves rather than for the aggregated mismatch, the nonlinearity of the aggregated mismatch definition will not affect the quality of the proxy. It is shown in multiple test cases that the new approach could potentially improve proxy quality for different types of proxies and different aggregated mismatch definitions. For cases with a large amount of observation data points, we also show that the use of reduced-order modeling can efficiently reduce the number of proxies needed and achieve similar improvement. The new approach is successfully applied to both synthetic and field examples and both examples show improved proxy quality.

1. Introduction

History matching is the process to update the reservoir models using measurement data, such as historical production, pressure, and seismic data. It plays an essential role in reservoir modeling and management. Existing methods for history matching can generally be divided into two categories: deterministic methods and stochastic methods (Oliver and Chen, 2011). Deterministic methods, such as the various gradient-based method (Sarma et al., 2006), try to find a single model that best matches with historical data. Stochastic methods, such as rejection sampling (Caers, 2011) and Markov chain Monte Carlo (MCMC) (Ma et al., 2008), try to quantify the posterior distribution of the reservoir model parameters after incorporating the data.

One key component in most history matching methods is the mismatch function. A mismatch function, typically defined as the root mean square error (RMSE) or the mean absolute error (MAE) between the simulated data and the observed data (Tarantola, 2005), quantifies the degree of consistency of a reservoir model with the historical data. Each evaluation of the mismatch function requires one reservoir simulation, which, for practical field cases, can already take hours if not days. Furthermore, most history matching methods typically require large numbers of mismatch function evaluations. The computational cost for history matching can be very expensive if all evaluations are done through reservoir simulation.

One way to reduce the computational cost is to construct a response surface proxy for the mismatch function. A response surface proxy is a parameterized mathematical expression that can be calibrated on a set of training data to approximate the input/output relations of the mismatch function. For typical problems, the response surface proxies take less than a second to construct and evaluate. Once the response surface proxy is constructed, it can be used in place of the simulator to evaluate the mismatch function and to significantly speed up the history matching process. Response surface proxies have been widely used for history matching (Landa and Güyagüler, 2003; Castellini et al., 2006; Friedmann et al., 2003). Many types of proxies have been introduced, including Kriging, splines, and polynomial functions. Yeten et al. (2005) provides a review and a comparison of different types of response surface proxies and Bhark and Dehghani (2014) provides a benchmarking of different experimental design techniques for assisted history matching.

One problem of the response surface methodology often faced

* Corresponding author.
E-mail address: jiang.xie@chevron.com (J. Xie).

http://dx.doi.org/10.1016/j.petrol.2016.05.026
0920-4105/

by engineers is the inaccuracy of the mismatch proxy. For example, (Slotte and Smorgrav, 2008) applied Markov chain Monte Carlo and response surface methodology to history match production data and they reported a correlation coefficient of only 0.52 (as a rule of thumb this number needs to be at least 0.85 for the proxy to be considered accurate). Zubarev (2009) investigated the pros and cons of proxy models and concluded that "with increasing complexity of the solution space", "the application of the proxy-modeling methodology is not recommended for history matching". One of the reasons why the mismatch proxy is not accurate is that, for most types of mismatch definitions, be it RMSE or MAE, the mismatch function is highly nonlinear, while most response surface proxies perform best when the underlying objective function is linear. It is true that increasing the number of training simulations may improve the proxy quality (Castellini et al., 2010). However, that translates to a higher computational cost.

In this paper, we are interested in improving the mismatch proxy quality with the same number of simulations by changing the way mismatch function is evaluated. We propose a proxy-for-data (PFD) approach, in which, instead of building proxy directly on the (nonlinear) mismatch, we build proxy for each simulated data point that is potentially more linear and calculate the mismatch using the data values predicted from these data proxies. Because the proxies are constructed for simulated data points rather than for the mismatch, the quality of the proxies would not be affected by the nonlinearity introduced by the definition of mismatch function (MAE or RMSE). We will show that the proposed PFD approach can substantially improve the mismatch proxy quality in two cases for different types of data, different types of proxy and different types of mismatch definition.

The proposed PFD approach does require constructing one proxy for each data point. For cases with a large number of observation data, the additional computational cost may become a concern. For such cases we propose the use of reduced-order modeling (He et al., 2011, 2013; He and Durlofsky, 2014) to project the high-dimensional data vector into a much lower-dimensional subspace. We will show that the use of reduced-order modeling can significantly reduce the number of proxies needed while preserving the benefits of the PFD approach.

The paper proceeds as follows. We will first present the traditional proxy-for-mismatch (PFM) approach of constructing the mismatching and demonstrate its limitation. We then present the new PFD approach and demonstrate its improvement over the PFM approach with numerical result. Then, we introduce the reduced-order modeling treatment for the PFD approach and apply it to the Brugge case where we have a large number of data. Finally, a summary of our findings and suggestions for future work will be provided at the end of the paper.

We note that an earlier version of this work was presented in an SPE conference proceeding paper (He et al., 2015).

2. Problem formulation

2.1. Mismatch function

In history matching, the degree a model $\mathbf{m}$ is consistent with the measurement data $\mathbf{d}$ is usually quantified by a so-called mismatch function $J(\mathbf{m})$. Here $\mathbf{m}$ is the vector of uncertain parameters in the system, such as permeability multipliers, relative permeability end point, and compressibility. $\mathbf{d} = [d_1, d_2, \ldots, d_{nd}]$ is a vector of measurement data, such as water cut, BHP or PLT data (n_d is the number of data points).

We denote as $\mathbf{g}(\mathbf{m}) = [g_1(\mathbf{m}), g_2(\mathbf{m}), \ldots, g_{nd}(\mathbf{m})]$ the vector of data values predicted by the simulator. The mismatch function J is usually defined as the norm of the difference between the actual measurement $\mathbf{d}$ and $\mathbf{g}(\mathbf{m})$. Depending on the kind of norm used, there are different definitions, such as the root mean square error (RMSE)

$$J_{\mathrm{RMSE}}(\mathbf{m}) = \sqrt{\frac{1}{n_d}\sum_{i=1}^{n_d}(g_i(\mathbf{m}) - d_i)^2} \tag{1}$$

and mean absolute error (MAE)

$$J_{\mathrm{MAE}}(\mathbf{m}) = \frac{1}{n_d}\sum_{i=1}^{n_d}\left|g_i(\mathbf{m}) - d_i\right|. \tag{2}$$

It is clear that the mismatch function $J(\mathbf{m})$ is a highly nonlinear function in terms of $g_i(\mathbf{m})$ whichever norm definition is used.

2.2. Response surface methodology

Each evaluation of the mismatch function $J(\mathbf{m})$ requires a reservoir simulation to be performed to calculate the predicted data value $\mathbf{g}(\mathbf{m})$. While one reservoir simulation can be expensive already, existing history matching methods usually require hundreds or thousands of evaluations of the mismatch function $J(\mathbf{m})$. Therefore, the computational cost can be significant if all the evaluations are performed using reservoir simulation.

To reduce the number of simulations needed, response surface methods are often used. A response surface proxy is a parameterized mathematical function that approximates the input/output relationship of the objective function, in the case, the mismatch function $J(\mathbf{m})$. Once a proxy is constructed, it can be used in place of reservoir simulation to evaluate $J(\mathbf{m})$ for history matching to significantly reduce the computational cost. Proxy methods are widely used for history matching and optimization applications (Landa and Güyagüler, 2003; Castellini et al., 2006; Friedmann et al., 2003). Popular proxy methods include linear Kriging, splines, and polynomial functions. See Yeten et al. (2005) for a more thorough review.

A proxy is constructed by fitting a parameterized mathematical function to a set of input/output pair $(\mathbf{m}_i, J_i)$, $i = 1, \ldots, n_s$ called the training data (n_s is the number of training data points). Each training data point is obtained from one reservoir simulation called a training simulation. The input parameters $\mathbf{m}_i$, $i = 1, \ldots, n_s$ on which training simulations are performed are usually determined through a process called experimental design, which determined a set of points $\mathbf{m}_i$ in the parameter space, such that the point set spans the parameter space as evenly as possible and the simulation result at those points can best represent the input-output relationship of the mismatch function. Frequently used methods include D-Optimal design and space-filling designs (Yeten et al., 2005). In this paper, space-filling designs will be employed in all cases.

Here and throughout the rest of the paper, we will use Blackboard Bold font to denote the proxy function in order to distinguish it from the true objective function. For example, the proxy function for the mismatch function $J(\mathbf{m})$ will be denoted as $\mathbb{J}(\mathbf{m})$. With this notation, the procedure to construct a proxy for $\mathbb{J}(\mathbf{m})$ is summarized as below.

- Step 1. Perform experimental design to generate n_s design points in the uncertainty parameter space.
- Step 2. Perform a reservoir simulation on each of the design points and calculate the values of the mismatch function J
- Step 3. Fit a parameterized mathematical function $\mathbb{J}(\mathbf{m})$ to the point set (m_i, J_i), $i = 1 \cdots l_s$.
- Step 4. Use the mathematical function $\mathbb{J}(\mathbf{m})$ in place of the reservoir simulation to evaluate the mismatch function $J(\mathbf{m})$ for history matching