Title
Fast high-resolution prediction of multi-phase flow in fractured formations

Authors
Pau, George Shu Heng
Finsterle, Stefan
Zhang, Yingqi

Publication Date
2016-02-01

DOI
10.1016/j.advwatres.2015.12.008

Peer reviewed

Fast high-resolution prediction of multi-phase flow in fractured formations

George Shu Heng Pau[a,*], Stefan Finsterle [b], Yingqi Zhang[b]

[a] Lawrence Berkeley National Laboratory, Climate and Ecosystem Sciences Division, 1 Cyclotron Road, Berkeley, CA 94720, United States

[b] Lawrence Berkeley National Laboratory, Energy Geosciences Division, 1 Cyclotron Road, Berkeley, CA 94720, United States

a r t i c l e i n f o

Keywords: Multiphase flow Fracture network Reduced order model Downscaling

a b s t r a c t

The success of a thermal water flood for enhanced oil recovery (EOR) depends on a detailed representation of the geometrical and hydraulic properties of the fracture network, which induces discrete, channelized flow behavior. The resulting high-resolution model is typically computationally very demanding. Here, we use the Proper Orthogonal Decomposition Mapping Method to reconstruct high-resolution solutions based on efficient low-resolution solutions. The method requires training a reduced order model (ROM) using high- and low-resolution solutions determined for a relatively short simulation time. For a cyclic EOR operation, the oil production rate and the heterogeneous structure of the oil saturation are accurately reproduced even after 105 cycles, reducing the computational cost by at least 85%. The method described is general and can be potentially utilized with any multiphase flow model.

1. Introduction

Efficient numerical approaches are needed to lower the com- putational barrier for performing optimization and uncertainty quantification using models that accurately represent complex multi-phase flow processes in fracture networks, with considerable impact on our ability to sustainably manage and optimize energy and water resource systems, and to effectively remediate contaminated sites. For example, a reliable evaluation of the economic viability of thermal water flood—a common enhanced oil recovery (EOR) technique—depends on whether we can predict the oil production rate accurately. Prediction of the oil production rate is typically obtained by constructing a numerical model that accurately cap- tures the geometrical and hydraulic details of the fracture network, which induces discrete, channelized flow behavior. The network also determines the effectiveness with which heat and brine penetrates the rock matrix, mobilizing and displacing the oil. Simulating an EOR operation using a discrete fracture network embedded in a low-permeability matrix is computationally very demanding, mainly because the detailed representation of the fracture network and the complex geometry of the matrix blocks bounded by randomly oriented fracture planes require high-resolution meshes.

In this paper, we apply a reduced order modeling (ROM) technique known as the Proper Orthogonal Decomposition Mapping Method (PODMM)—first proposed by Robinson et al. [1]—which allows us to reconstruct the solutions from a high-resolution model represent- ing the fracture network as a heterogeneous medium based on solu- tions obtained using low-resolution models that only have upscaled, effective properties of the fracture network and thus can be effi- ciently simulated. This technique was recently enhanced and applied to land surface models to accurately reconstruct hydrological states, heat fluxes, and carbon fluxes [2,3]. However, the suitability of the method for modeling multiphase non-isothermal subsurface prob- lems with significant nonlinear temporal and spatial dynamics has yet to be demonstrated. This work evaluates the accuracy of PODMM for a multiphase problem (an enhanced oil recovery problem).

PODMM can be viewed as a regression-based downscaling technique. Overviews of empirical downscaling techniques have been presented before (see, e.g., Wilby et al. [4], Fowler et al. [5], Gutmann et al. [6]). Previous work on regression-based down- scaling methods focuses on constructing empirical parametric models between the predictors and variables of interest [7,8]. In the context of reduced order models (ROMs), regression models can also be constructed between model parameters and the vari- ables of interest [9,10]. PODMM differs from the above regression approaches in that proper orthogonal decomposition (POD) is not just used to obtain a dimensionally reduced representation of the high-dimensional data. Instead, a least-square minimization problem that utilizes portions of the singular vectors obtained through the POD procedure [11] is solved to directly map low-resolution solutions to the high-resolution solutions. More details are provided in Section 2.2.

* Corresponding author. Tel.: +1 5104867196.
E-mail addresses: gpau@lbl.gov (G.S.H. Pau), safinsterle@lbl.gov (S. Finsterle), yqzhang@lbl.gov (Y. Zhang).

http://dx.doi.org/10.1016/j.advwatres.2015.12.008
0309-1708/

PODMM also differs from the projection-based POD methods that were previously applied to subsurface flow problems [12–15] and other engineering fields [16–18]. Specifically, the projection-based POD method is an intrusive approach that requires projecting the governing partial differential equations onto a linear space spanned by the singular vectors, and implementing the resulting discrete equations. For a multiphase non-isothermal model, the complicated nonlinear terms require additional approximations [19–21] in order to obtain a set of discrete equations

that can be solved efficiently. PODMM is considerably simpler since it only requires solving the low-resolution models once the ROMs have been trained. Thus, it does not require intrusive changes to the simulation software, mak- ing it an attractive method for complex multiphase flow problems.

2. Methods

2.1. Mapped fracture network models

We demonstrate the proposed PODMM model-reduction ap- proach for an EOR operation conducted in a fractured hydrocarbon reservoir with a single injection-extraction well. An individual cycle of the operation consists of four phases: (1) injection of hot water at 10 kg/s for 3 days, (2) an inactive soaking period of 4 days, (3) pro- duction of oil and water for 6 days at a total rate of 5 kg/s, and (4) an inactive rest period of 1 day. This two-week cycle is repeated 105 times for a total simulation time of 1470 days. The distribution of oil in the reservoir and the oil production rates are the key prediction variables of interest.

We consider a discrete fracture network within a model domain of dimensions 100 × 50 × 30 m. Fractures are generated by randomly sampled values for size, orientation, and aperture from appropriate, truncated probability distributions. Two fracture sets with an average fracture spacing of 4 m are generated using the code ThreeDFracMap [22]. In our modeling approach, the fracture network is then repre- sented by a heterogeneous continuum model. The fractures are first mapped onto a structured, uniform mesh, before upscaled, heteroge- neous, anisotropic permeabilities are calculated based on the num- ber, aperture, and orientation of the fractures intersecting the given element. Elements that do not contain any fractures are assigned a low matrix permeability of 10^{-18} m^2. The procedure is described in detail in Parashar and Reeves (2011).

mate, efficient solutions for the entire simulation period; these solu- tions are then downscaled to provide high-resolution predictions of the cyclic EOR operation based on the mapping procedure in PODMM. We examine three alternative LRMs: (1) an upscaled heteroge- neous model (LRM-HET), using the exact same conceptualization as the HRM with the exception that it uses a coarser discretiza- tion, (2) a simple homogeneous model (LRM-HOM) with a perme- ability of 10^{-13} m^2, and (3) a dual-porosity model (LRM-DPM) [24] with fracture- and matrix-continuum permeabilities of 10^{-13} m^2 and 10^{-18} m^2, respectively. Key parameters are summarized in Table 1. Note that the number of elements of LRM-DPM is twice that of LRM- HOM and LRM-HET. The permeabilities of LRM-HOM and LRM-DPM are chosen to approximately represent the fracture network perme- ability such that they produce some of the behaviors seen in the HRM. However, no rigorous upscaling technique is used to determine the permeabilities of the LRMs, e.g., by matching the oil production rate obtained from the HRM through an inverse modeling procedure. A calibrated LRM will likely improve the performance of the PODMM. However, the goal of this work is to demonstrate that the accuracy of the PODMM is not due to the calibration of LRM model parameters to fit the outputs of the HRM.

Fig. 1(b) shows the oil saturation after 3 days. The solutions ob- tained from LRM-HET is a smoother representation of the consider- ably more intricate distribution obtained with the HRM. Fig. 1(c) and (d) compares oil production rate (Q_{oil}) determined using the HRM and the 3 LRMs at the 55th, and 105th cycles, respectively. For the parameters given in Table 1, LRM-HET over-predicts the Q_{oil} while LRM-HOM and LRM-DPM under-predict the Q_{oil}.

2.2. Proper orthogonal decomposition mapping method

We summarize the Proper Orthogonal Decomposition Mapping Method (PODMM) here; details can be found in [2]. The method con- sists of a training stage and a prediction stage. During the training stage, we determine the solutions (e.g., oil saturations and fluxes at all locations) to the low- and high-resolution models (denoted as $\mathbf{g}$ and $\mathbf{f}$, respectively) at N time points. These N solutions constitute the training set. In our example, snapshots are obtained at 1-day intervals from multiple consecutive EOR cycles. We then perform a singular value decomposition (SVD) of the following matrix $\mathbf{W}$:

Based on this heterogeneous, high-resolution continuum model representing a network of discrete features embedded into a low-permeability matrix, we use the “dead-oil” (EOS8) module of TOUGH2 [23] to simulate the response of the reservoir to cyclic in-

$$\mathbf{W} = \begin{matrix} \mathbf{f}_1 - \bar{\mathbf{f}} \\ \mathbf{g}_1 - \bar{\mathbf{g}} \end{matrix} \cdots \begin{matrix} \mathbf{f}_N - \bar{\mathbf{f}} \\ \mathbf{g}_N - \bar{\mathbf{g}} \end{matrix} \quad (1)$$

jection of hot water and production of a multi-phase mixture of oil and water. Simulating a long sequence of injection-production cycles is computationally expensive, especially if a high-resolution contin-

where $\mathbf{f}_i$ and $\mathbf{g}_i$ are the high- and low-resolution solutions at the ith time point, uum representation of the discrete network is needed to capture the exchange of fluids between the reservoir rock (which contains most of the oil) and a network of discrete fractures (whose main role is

$$\bar{\mathbf{f}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{f}_i,$$

$$\bar{\mathbf{g}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{g}_i \quad (2)$$

to provide the pathways for oil extraction) embedded in that ma- trix. Moreover, the resolution also affects the system behavior and computational costs. The PODMM approach described below pre-

The POD bases, ζ_i, $i = 1,\ldots M$, are given by the resulting singular vectors and can be decomposed into dicts the high-resolution behavior using a computationally efficient low-resolution model (LRM), combined with a mapping procedure for downscali . e simulation results. Two basic grids with different

$$= \zeta^{\mathbf{g}} \qquad \zeta_i \qquad (3)$$

resolutions are thus developed: a high-resolution model (HRM) with where $\zeta^{\mathbf{f}}$ and $\zeta^{\mathbf{g}}$ are components associated with the HRM and LRMs, an element size of 2 m, and a low-resolution model (LRM) with an element size of 5 m (see Fig. 1(a)). The LRM thus has about 15 times fewer elements than the HRM, making it significantly more efficient at the expense of loss of accuracy in representing discrete flow behav- ior in the fractures and fluid exchange with the matrix. The HRM pro- respectively, and M is the chosen number of POD bases to use in an approximation.

During the prediction stage, we first determine a coarse- resolution solution $\mathbf{g}$, and solve vides simulation data for a relatively short training phase; it is also used in this study as the reference solution needed to demonstrate the accuracy of the proposed approach. The LRM provides approxi-

$$\boldsymbol{\alpha}(\mathbf{g}) = \arg\min \left\| \mathbf{g} - \mathbf{g}^{-} - \sum_{i=1}^{M} \gamma_i \zeta^{\mathbf{g}} \right\|_2 \qquad (4)$$

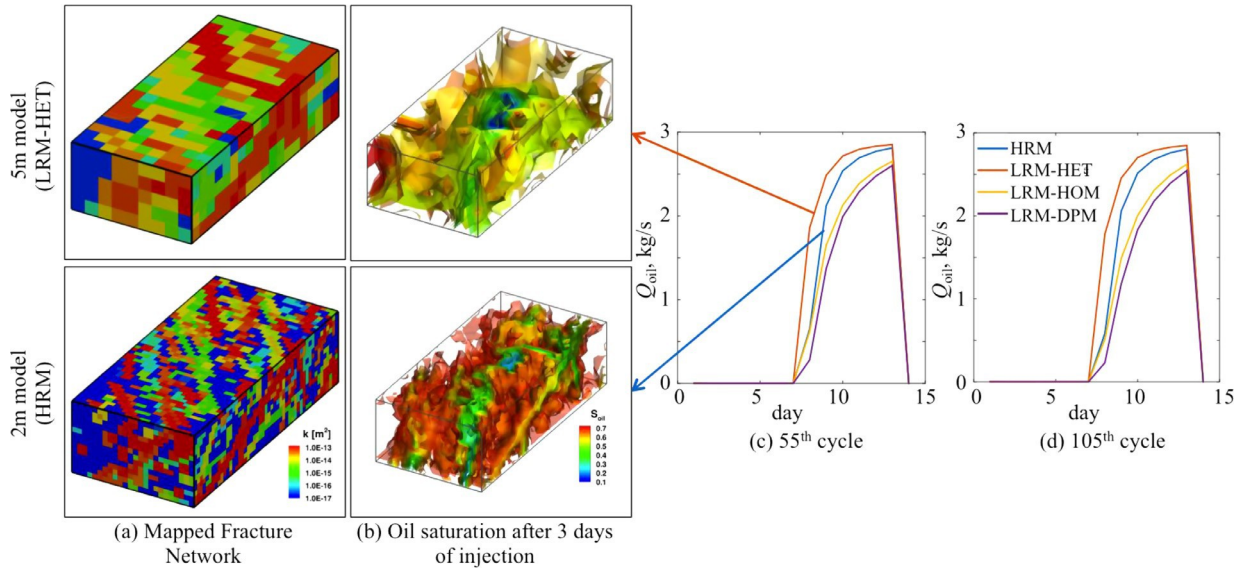

Fig. 1. Discrete fracture network mapped onto TOUGH2 continuum grid using (a) a low resolution of 5 m (LRM-HET) (top) , and a high resolution of 2 m (HRM) (bottom). Oil saturation after 3 days of hot water injection, calculated with (b) LRM-HET (top), and HRM (bottom). The comparison of oil production rate (Q_{oil}) curves at the (c) 55th cycle, and (d) 105th cycle for the HRM and the LRMs.

Table 1
Summary of key geometric, hydraulic, and thermal model parameters.

Parameter	HRM	LRM		
Domain size, m	100×50×30			
	LRM-HOM	LRM-DPM	LRM-HET	
Element size, m	2	5	5	5
Number of elements	18,750	1,200	2,400	1,200
Matrix permeability, m^2	10^{-18}	N/A	10^{-18}	10^{-18}
Mapped fracture network permeability, m^2	Heterogeneous, anisotropic	10^{-13} isotropic	10^{-13} isotropic	Heterogeneous, anisotropic
porosity, %	4	4	4	4
heat conductivity, W m^{-1} K^{-1}	2.5	2.5	2.5	2.5
heat capacity, J kg^{-1} K^{-1}	1000	1000	1000	1000
pore compressibility, Pa^{-1}	10^{-9}	10^{-9}	10^{-9}	10^{-9}

where $\|\cdot\|$ the root mean square of a vector, and γ_i is the mixing coefficient of POD basis ζ^g. The downscaled high-resolution solution $\mathbf{f}^{ROM}$ is then gi$\|\cdot\|$ by $\mathbf{f}^- + {}^M\alpha_i(g)\zeta^f$.

We can jointly consid$\}$, multiple variables in a single ROM to take advantage of correlations between the variables by modifying matrix $\mathbf{W}$ in Eq. (1) such that each $i=1$ mn consists of concatenated solutions of these variables (from both low- and high-resolution models). See [3], for a more complete descrip on of the method. For the current study, we jointly predict all primary variables of the model, i.e., the oil saturation ($\boldsymbol{S}_{oil}$), pressure ($\boldsymbol{P}$) and temperature ($\boldsymbol{T}$); $\mathbf{g}$, and $\mathbf{f}$ in Eq. (1) are then given by $[\boldsymbol{S}_{oil}; \boldsymbol{P}; \boldsymbol{T}]$ of the low-, and high-resolution models, respectively. Since these variables have disparate magnitudes, they are first rescaled by their means. Similarly, we create a ROM for the fluxes by jointly predicting the flux distributions of oil ($\boldsymbol{F}_{oil}$), water ($\boldsymbol{F}_{water}$) and heat ($\boldsymbol{F}_{heat}$) over the entire domain.

HRM and the LRMs and determine M_R by finding M that minimizes the absolute error between $\mathbf{f}$ and $\mathbf{f}^{ROM}$ within that cycle.

2.3. *Quantifying and estimating error*

To measure the accuracy of $\mathbf{f}^{ROM}$ relative to $\mathbf{f}$, we define the aver- aged root mean square error (RMSE) of the ROM approximation over a cycle as E_{RMSE}:

$$E_{RMSE} = \frac{1}{14}\sum_{i=1}^{14} e_{RMSE}(\mathbf{f}^{ROM}_i, \mathbf{f}_i) \tag{5}$$

The main model parameter for the ROM is the parameter M, the number of bases to use in the approximation. It controls the accuracy and stability of the approximation. A common heuristic approach to determining an appropriate M, denoted by M_R, is by specifying a desired percentage of variance explained, obtained by summing the normalized singular values associated with the M_R selected POD bases [10,15]. However, this approach favors the use of large M_R, which can be unsuitable in a PODMM approximation since it can lead to overfitting when solving the minimization problem given by Eq. (4). To ensure our approximation is stable, we have instead used an empirical approach: we simulate an additional cycle using the

where $\mathbf{f}^{\mathrm{ROM}}$ is the ROM approximation of $\mathbf{f}_i$, $e_{\mathrm{RMSE}}(\mathbf{f}^{\mathrm{ROM}}, \mathbf{f})$ is the RMSE between $\mathbf{f}^{\mathrm{ROM}}$ and $\mathbf{f}$, and E_{RMSE} is averaged over the two-week cycle. Similarly, we define the PODMM approximation on the coarse grid, given by $\mathbf{g}^{\mathrm{ROM}} = \bar{\mathbf{g}} + \sum_{i=1}^{M} \alpha_i(\mathbf{g})\zeta^{\mathbf{g}}_i$, and determine its RMSE error, denoted as $e_{\mathrm{RMSE}}(\mathbf{g}^{\mathrm{ROM}}, \mathbf{g})$, where $\alpha_i(\mathbf{g})$ is determined from Eq. (4). We can then define an error estimator based on $e_{\mathrm{RMSE}}(\mathbf{g}^{\mathrm{ROM}}, \mathbf{g})$:

$$e_{\mathrm{EST}}(\mathbf{g}) = C_0 e_{\mathrm{RMSE}}(\mathbf{g}^{\mathrm{ROM}}, \mathbf{g}), \tag{6}$$

where C_0 is a positive constant; C_0 is typically greater than 1 because Eq. (4) minimizes the e_{RMSE} of $\mathbf{g}^{\mathrm{ROM}}$, and thus the e_{RMSE} of $\mathbf{g}^{\mathrm{ROM}}$ will be smaller than the e_{RMSE} of $\mathbf{f}^{\mathrm{ROM}}$. We determine C_0 by computing the maximum ratio of the e_{RMSE} of $\mathbf{f}^{\mathrm{ROM}}$ to the e_{RMSE} of $\mathbf{g}^{\mathrm{ROM}}$ for the days in the additional cycle used to determine M_R.

The error estimator e_{EST} serves two purposes. First, it can be used to determine time scales for which the ROM will be accurate when

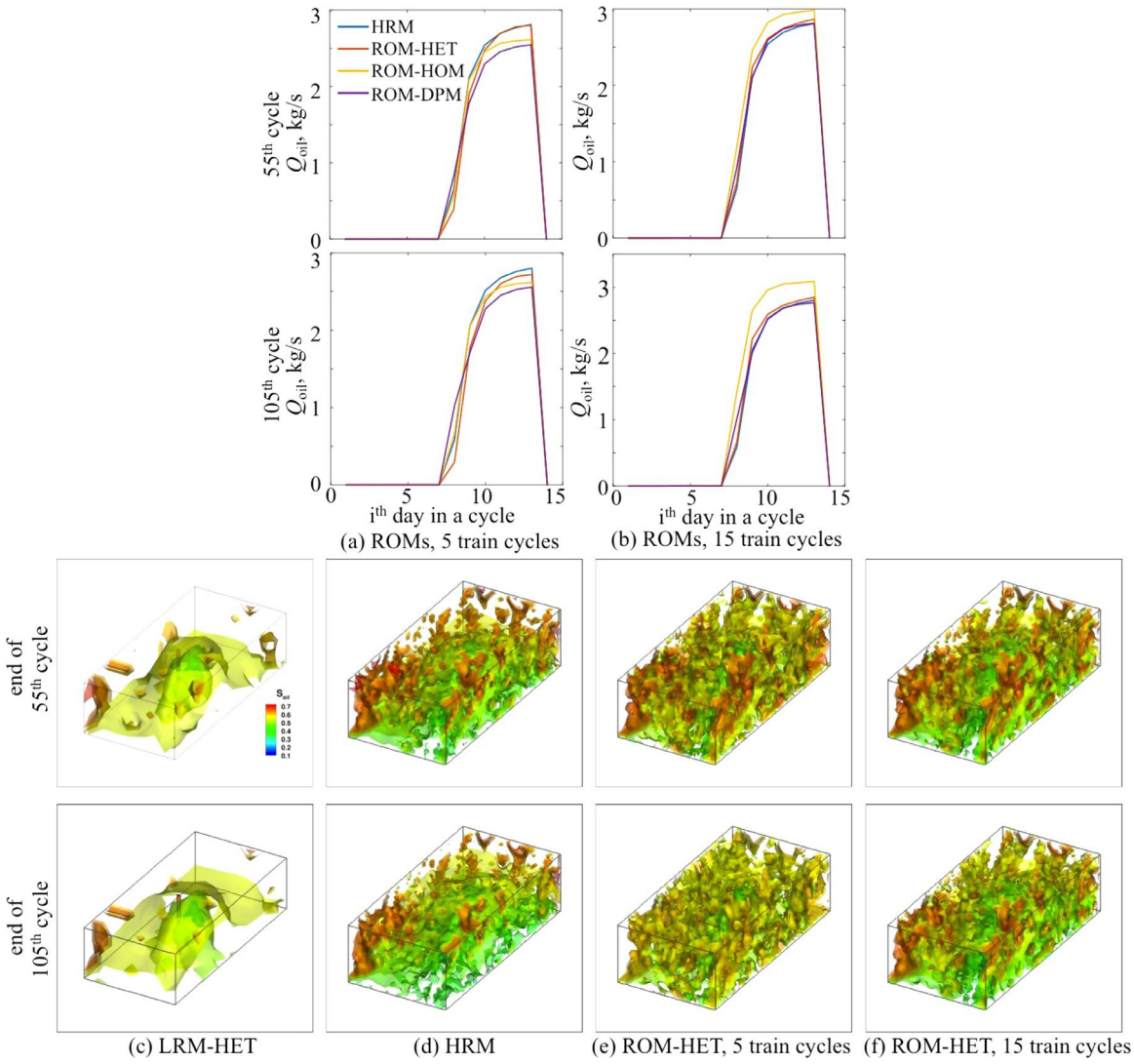

Fig. 2. Comparison of the oil production rate (Q_{oil}) obtained for the 55th (top), and 105th (bottom) cycles from the 3 different ROMs trained using (a) 5, and (b) 15 training cycles; and oil saturation S_{oil} in the entire system from (c) LRM-HET, (d) HRM, (e) ROM-HET with 5 training cycles, and (f) ROM-HET with 15 training cycles for the last day of the 55th (top), and 105th (bottom) cycles.

predicting into the future. Since the e_{EST} is only a function of $\mathbf{g}$, it can be efficiently evaluated when determining $\mathbf{f}^{ROM}$. For a desired level of accuracy, τ_{tol}, we can evaluate time scales in which $e_{EST} \leq \tau_{tol}$; we demonstrate this in Section 3.2. Second, e_{EST} can also be used to determine the length of the training period in a similar manner. By specifying a time scale in which we require $e_{EST} \leq \tau_{tol}$, we can adaptively increase the length of the training period until the requirement is fulfilled.

2.4. Training and validation datasets

The HRM described in Section 2.1 serves as the reference model in this work. Based on this computationally demanding model, we simulate a relatively short period (i.e., the initial few injection- production cycles) to train the ROM. To distinguish between the var- ious ROMs, we use ROM-HET, ROM-HOM, and ROM-DPM to denote ROMs constructed using the LRM-HET, LRM-HOM, and LRM-DPM, respectively. We focus on the prediction of oil saturation $\boldsymbol{S}_{oil}$ in the entire model domain and oil production rate (Q_{oil}), which is com- puted by integrating predicted $\boldsymbol{F}_{oil}$ over the boundaries of the well. We examined two training periods: 5 and 15 EOR cycles (i.e., 70 and 210 days). Daily solutions from the low- and high-resolution models

within these periods are used to train the ROMs. Since the injection- production cycle consists of three distinct stages (injection, soak- ing/rest and production), separate ROMs are constructed for the three stages because the solutions in each stage can have unique charac- teristics. For example, $\boldsymbol{F}_{oil}$ is close to zero during the soaking stage, and of different characteristics during the injection and production stages. We note that M_R will be different for the three stages.

In the results section, we study the accuracy of the ROMs using validation sample sets consisting of the daily HRM solutions deter- mined up to the 105th cycle (Day 1470). Depending on the training periods, the validation sample set starts from the 6th cycle (Day 71) for ROMs trained using 5 cycles, and the 16th cycle (Day 211) for ROMs trained using 15 cycles.

3. Results

3.1. Oil production rate

The ROMs are able to reproduce the oil production rates (Q_{oil}) very accurately (Fig. 2 (a) and (b)). The relative error between Q_{oil} of HRM and ROM-HET, constructed using 15 training cycles, is less than 6% averaged over the production stage during the 55th cycle. The

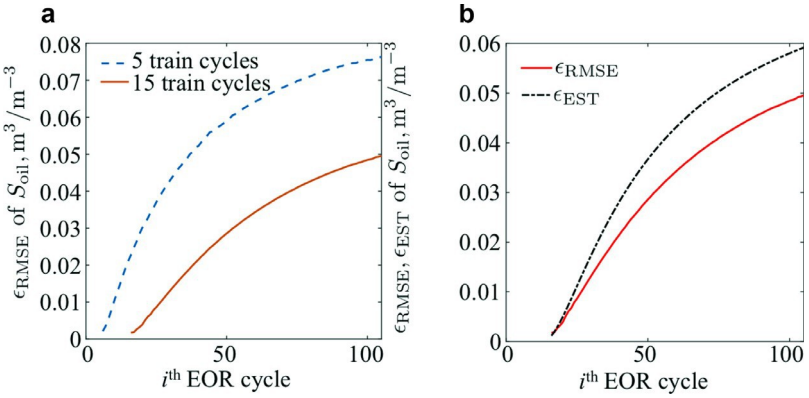

Fig. 3. (a) The E_{RMSE} of ROM-HET increases with number of cycles after the training period. Increasing the number of training cycles from 5 to 15 reduces E_{RMSE}. (b) For ROM-HET trained using 15 cycles, $E_{EST} = \frac{1}{14}\sum_{i=1}^{14} e_{EST}(\mathbf{g}_i)$ is close to E_{RMSE}, indicating that e_{EST} is a good error estimator.

error stays approximately the same even when predicting up to the 105th cycle. Compared to the Q_{oil} obtained using LRMs, ROM-HET reduces the LRM-HET's errors by 83%, and 85% for the 55th, and 105th cycles, respectively. If a lower accuracy is acceptable, for example due to underlying uncertainty in the HRM, 5 training cycles may be sufficient since it leads to only slight increases in the error. By per- forming only 6 cycles of HRM simulation instead of 105 cycles based on our proposed PODMM, we can reduce the computational cost by 94%; running the LRM, and setting up and evaluating the ROM are computationally very efficient. We elaborate further on the condition under which a shorter training period is possible in Section 3.3.

Comparing the three different ROMs (Fig. 2 (a) and (b)), ROM-HET is consistently more accurate than ROM-HOM and ROM-DPM. This is expected because (1) LRM-HET is more accurate than LRM-HOM and LRM-DPM, and (2) the variation of the approximation error with in- creasing M has a more well-defined minimum, resulting in a more accurate determination of M_R. Both reasons can be attributed to the fact that LRM-HET is a better approximation of the HRM since it re- tains some of the heterogeneous structure of the HRM.

When the ROMs are trained using 15 training cycles, ROM-DPM can be as accurate as ROM-HET. ROM-HOM, however, consistently over-predicts Q_{oil}. This over-prediction is likely due to overfitting. Since the approximation error increases with time, M_R should de- crease over time to avoid overfitting. We will study how M_R should be reduced over time in the future. We note that if we have determined M_R by minimizing the absolute error of Q_{oil} between the ROMs and the HRM over the entire validation period, we are able to get good approximations for all ROMs. However, determining the actual error of Q_{oil} over the entire validation period is not practical since we want to avoid simulating the HRM over an extended period of time.

3.2. Reproduction of fine-scale saturation distribution

We first determine how the accuracy of the ROMs in reproducing fine-scale saturation solutions changes over time. Initial tests suggest that ROM-HOM and ROM-DPM cannot reproduce the heterogeneity in the solution obtained using the HRM; therefore, we will only dis- cuss results obtained using ROM-HET in this section. Fig. 3(a) shows that the error, E_{RMSE}, increases with the number of cycles after the initial training period. The use of 15 instead of 5 training cycles re- duces E_{RMSE} by 35%. In addition, E_{RMSE} of ROM-HET trained with 5 cycles grows much faster than ROM-HET trained with 15 cycles. The poorer performance of ROM-HET trained with 5 cycles indicates that the initial transient dynamics of the first 5 cycles are poor represen- tations of the long-time behavior of $\boldsymbol{S}_{oil}$. Subsequent slower growth of E_{RMSE} indicates that $\boldsymbol{S}_{oil}$ changes more slowly at later times.

Since E_{RMSE} increases with time, the ROM-HET we constructed is only accurate for a finite number of cycles. To determine the range of cycles for which the ROM-HET is valid without performing HRM simulation over an extended per $\}_{i=1}^{4}$, we will use the error estimator e_{EST} defined in Eq. (6). For the ROM-HET trained with 15 cycles, C_0, determined based on the 16th cycle, is found to be 10.334. Fig. 3(b) shows that $E_{EST} = \frac{1}{14}\sum_{i=1}^{14} e_{EST}(\mathbf{g}_i)$, the mean of e_{EST} over a cycle, is a good predictor of E_{RMSE}. We can thus determine a time scale for which the ROM-HET has desired accuracy by specifying an appropriate tolerance for E_{EST}. For example, if we specify a tolerance of 0.05, the ROM-HET that we have constructed has the desired accuracy up to the 74th cycle for predicting $\boldsymbol{S}_{oil}$.

We now determine how well PODMM reproduces the $\boldsymbol{S}_{oil}$ distribu- tion. Fig. 2(d) shows the heterogeneous structure in $\boldsymbol{S}_{oil}$ at the end of the 55th cycle (Day 770) and 105th cycle (Day 1470). Fig. 2 (f) shows that ROM-HET trained with 15 cycles is able to reproduce most of the intricate structure of $\boldsymbol{S}_{oil}$ on the 2 m scale. However, ROM-HET is not able to reproduce the very high $\boldsymbol{S}_{oil}$ values; the approximated solu- tions are smoother since only 9 POD bases are included in the approx- imation. Although the ignored POD bases amount to only 0.04% of the variance in the solutions within the training period, they can have a significant impact on the approximation of the solutions at later times, especially if these solutions are significantly different from the solutions in the training period. Consistent with the results in Fig. 3, ROM-HET trained with 5 cycles has poorer predictive capability (Fig. 2 (e)). Nonetheless, in both cases, ROM-HET added a significant amount of small-scale structure to the LRM results shown in Fig. 2(c). In this respect, the PODMM can be viewed as a downscaling procedure. With 15 training cycles (plus one additional cycle for determining M_R), the computational cost of simulating 105 cycles is reduced by 85%.

3.3. Dependence of training period on fine-scale reproduction

Compared to Q_{oil}, the estimation of $\boldsymbol{S}_{oil}$ requires a longer train- ing period. Approximating $\boldsymbol{S}_{oil}$ over the entire domain requires repro- ducing a significant amount of fine-scale information and is thus a challenging problem. On the other hand, Q_{oil} is a scalar, integrated quantity over a smaller region of the domain near the well and does not need to accurately capture fine-scale information in the $\boldsymbol{F}_{oil}$ over the entire domain. We can thus obtain sufficiently accurate predic- tion even with a shorter training period (Fig. 2(a)). For the current problem where the goal is to

predict future solutions, we can con- clude that a longer training period is needed to accurately capture fine-scale features in a solution. If we are interested in capturing the effects of operational parameters (e.g., injection rate, injection dura- tion, production rate and production duration) and model parameters (e.g., rock properties), the training period will additionally depend on how the $\boldsymbol{S}_{\text{oil}}$ and Q_{oil} vary with these parameters within the modeled parameter space.

4. Conclusions and future work

We have demonstrated that PODMM is capable of predicting long- time behavior of the oil production rate (Q_{oil}) and oil saturation ($\boldsymbol{S}_{\text{oil}}$). The main conclusions are

(1) The number of training cycles depends on the properties of the variables we are approximating. We showed that Q_{oil} is ac- curately approximated using only 5 training cycles while 15 training cycles are needed to reproduce the intricate structure of $\boldsymbol{S}_{\text{oil}}$ using the PODMM method.

[4] Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, et al. Statisti- cal downscaling of general circulation model output: a comparison of methods. Water Resour Res. 1998;34:2995–3008. http://dx.doi.org/10.1029/98WR02577.

[5] Fowler HJ, Blenkinsop S, Tebaldi C. Linking climate change modelling to im- pacts studies: recent advances in downscaling techniques for hydrological modelling. Int J Climatology. 2007;27:1547–78. http://dx.doi.org/10.1002/joc. 1556.

[6] Gutmann E, Pruitt T, Clark MP, Brekke L, Arnold JR, Raff DA, et al. An intercompar- ison of statistical downscaling methods used for water resource assessments in the United States. Water Resour Res. 2014;50:7167–86. http://dx.doi.org/10.1002/ 2014WR015559.

[7] von Storch H, Zorita E, Cubasch U. Downscaling of global climate change esti- mates to regional scales: an application to Iberian rainfall in wintertime. J Cli- mate. 1993;6:1161–71.

(2) The approximation errors grow with time, indicating that the resulting ROMs have a limit to how far they can predict into the future without significant loss of accuracy. We show that it is possible to determine the limit through the use of an error estimator.
(3) Different LRMs can be used with PODMM. For some variables (e.g., Q_{oil}), simpler LRMs (e.g., a dual-porosity model) can pro- duce a ROM that is sufficiently accurate for predictive purposes. However, ROMs constructed using more complex LRMs that are better representations of the HRM will typically have more consistent predictive capabilities.

For future work, we plan to study how PODMM can be used in cases where operational, and model parameters are varied. In addi- tion, we will apply PODMM to higher-resolution models and more complex models (e.g., subsurface flow coupled to geomechanics) to study and improve the robustness of PODMM.

Acknowledgment

We would like to thank A. Guadagnini and the two anonymous reviewers for their constructive comments. This research was sup- ported, in part, by the U.S. Department of Energy under Contract #DE- AC02-05CH11231. We thank Rishi Parashar of the Desert Research In- stitute for making ThreeDFracMap available to this project.

References

[1] Robinson T, Eldred M, Willcox K, Haimes R. Strategies for multifidelity optimiza- tion with variable dimensional hierarchical models. In Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Con- ference. American Institute of Aeronautics and Astronautics, Newport, Rhode Is- land, 2006.
[2] Pau GSH, Bisht G, Riley WJ. A reduced-order modeling approach to represent subgrid-scale hydrological dynamics for land-surface simulations: application in a polygonal tundra landscape. Geosci Model Dev. 2014;7:2091–105. http://dx.doi. org/10.5194/gmd-7-2091-2014.
[3] Pau GSH, Shen C, Riley WJ. Accurate and efficient prediction of fine-resolution hy- drologic and carbon dynamic simulations from coarse-resolution models. (2015). Submitted.
[8] Hanssen-Bauer I, Forland EJ, Haugen JE, Tveito OE. Temperature and precipita- tion scenarios for Norway: comparison of results from dynamical and empirical downscaling. Climate Res 2003;25:15–27. http://dx.doi.org/10.3354/cr025015.
[9] Higdon D, Gattiker J, Williams B, Rightley M. Computer model calibration using high-dimensional output. J American Stat Assoc 2008;103:570–83.
[10] Wilkinson RD. Bayesian calibration of expensive multivariate computer experi- ments. Large-Scale Inverse Problems Quantif Uncert 2011;707:195–215.
[11] Everson R, Sirovich L. Karhunen–Loeve procedure for gappy data. J Opt Soc Amer A. 1995;12:1657–64.
[12] Winton C, Pettway J, Kelley CT, Howington S, Eslinger OJ. Application of proper orthogonal decomposition (POD) to inverse problems in saturated groundwater flow. Adv Water Resour. 2011:1–24. http://dx.doi.org/10.1016/j.advwatres.2011. 09.007.
[13] Boyce SE, Nishikawa T, Yeh WWG. Reduced order modeling of the Newton formu- lation of MODFLOW to solve unconfined groundwater flow. Adv Water Resour. 2015;83:250–62. http://dx.doi.org/10.1016/j.advwatres.2015.06.005.
[14] Li H, Luo Z, Chen J. Numerical simulation based on POD for two-dimensional so- lute transport problems. Appl Math Model 2011;35:2489–98. http://dx.doi.org/10.1016/j.apm.2010.11.064.
[15] Siade AJ, Putti M, Yeh WWG. Snapshot selection for groundwater model reduction using proper orthogonal decomposition. Water Resour Res. 2010;46:W08539. http://dx.doi.org/10.1029/2009WR008792.
[16] Kunisch K, Volkwein S. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. SIAM J Numer Anal 2002;40:492–515. http://dx.doi.org/10.1137/S0036142900382612.
[17] Luo Z, Du J, Xie Z, Guo Y. A reduced stabilized mixed finite element formulation based on proper orthogonal decomposition for the non-stationary Navier–Stokes equations. Int J Numer Meth Eng 2011;88:31–46. http://dx.doi.org/10.1002/nme. 3161.
[18] Willcox K, Peraire J. Balanced model reduction via the proper orthogonal decom- position. AIAA J 2002;40:2323–30.
[19] Barrault M, Maday Y, Nguyen NC, Patera AT. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equa- tions. Comptes Rendus Mathematique. 2004;339:667–72. http://dx.doi.org/10. 1016/j.crma.2004.08.006.
[20] Constantine PG, Wang Q. Residual minimizing model interpolation for param- eterized nonlinear dynamical systems. SIAM J Scientif Comput 2012;34 A2118– A2A44. http://dx.doi.org/10.1137/100816717.
[21] Chaturantabut S, Sorensen D, Steven J. Nonlinear model reduction via discrete empirical interpolation. SIAM J Scientif Comput 2010;32:2737–64.
[22] Parashar R, Reeves DM. Computation of flow and transport in fracture networks on a continuum grid. In: Proceedings of MODFLOW and More: Integrated Hydro- logic Modeling, Golden, CO; 2011.
[23] Pruess K, Moridis G, Oldenburg C. TOUGH2 User's Guide, Version 2.1, Report LBNL- 43134. In: Lawrence Berkeley National Laboratory, Berkeley, California; 1999.
[24] Pruess K, Narasimhan TN. A practical method for modeling fluid and heat-flow in fractured porous-media. Soc Petroleum Eng J 1985;25:14–26. http://dx.doi.org/ 10.2118/10509-PA.

value decomposition (SVD) of the following matrix:

$$\mathbf{W} = \begin{bmatrix} \mathbf{f}_1 - \bar{\mathbf{f}} & & \mathbf{f}_N - \bar{\mathbf{f}} \\ & \cdots & \\ \mathbf{g}_1 - \bar{\mathbf{g}} & & \mathbf{g}_N - \bar{\mathbf{g}} \end{bmatrix} \tag{1}$$

where $\mathbf{f}_i$ and $\mathbf{g}_i$ are the high- and low-resolution solutions at the ith time point,

$$\bar{\mathbf{f}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}_i, \quad \bar{\mathbf{g}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{g}_i \tag{2}$$

The POD bases, $\zeta_i, i = 1, \ldots M$, are given by the resulting singular vectors and can be decomposed into

$$\zeta_i = \begin{bmatrix} \zeta_i^{\mathbf{f}} \\ \zeta_i^{\mathbf{g}} \end{bmatrix} \tag{3}$$

where $\zeta_i^{\mathbf{f}}$ and $\zeta_i^{\mathbf{g}}$ are components associated with the HRM and LRMs, respectively, and M is the chosen number of POD bases to use in an approximation.

During the prediction stage, we first determine a coarse-resolution solution $\mathbf{g}$ and solve

$$\boldsymbol{\alpha}(\mathbf{g}) = \arg\min_{\gamma} \left\| \mathbf{g} - \bar{\mathbf{g}} - \sum_{i=1}^{M} \gamma_i \zeta_i^{\mathbf{g}} \right\|_2 \tag{4}$$