

Production forecasting and uncertainty quantification for naturally fractured reservoirs using a new data-space inversion procedure

Wenyue Sun¹  · Mun-Hong Hui² · Louis J. Durlofsky¹

Received: 30 September 2016 / Accepted: 26 February 2017 / Published online: 22 March 2017
© Springer International Publishing Switzerland 2017

Abstract A new method for production forecasting and uncertainty quantification, applicable for realistic naturally fractured reservoirs (NFRs) represented as general discrete-fracture-matrix (DFM) models, is developed and applied. The forecasting procedure extends a recently developed data-space inversion (DSI) technique that generates production predictions using only prior-model simulation results and observed data. The method does not provide posterior (history-matched) geological models. Rather, the DSI method treats production data as random variables. The prior distribution is estimated from the flow simulations performed on prior geological models, and the posterior data-variable distribution is sampled using a data-space randomized maximum likelihood method. The DSI treatment requires the parameterization of data variables to render them approximately multivariate Gaussian. The complex production data considered here (resulting from frequent well shut-ins) is treated using a new reparameterization that involves principal component analysis combined with histogram transformation. The DSI method is first applied for two-dimensional DFM systems involving multiple fracture scenarios. In this case, comparison with a rejection sampling procedure is possible, and we show that the DSI results for P10, P50, and P90 statistics are consistent with rejection sampling results. The DSI method is then applied to a

realistic NFR that has undergone 15 years of primary production and is under consideration for waterflooding. To construct the DSI representation, around 400 prior DFM models, which correspond to different geologic concepts and properties, are simulated. Two different reference ‘true’ models, along with different data-assimilation durations, are considered to evaluate the performance of the DSI procedure. In all cases, the DSI predictions are shown to be consistent with the forecasts from the ‘true’ model and to provide reasonable quantification of forecast uncertainty.

Keywords Data-space inversion · Naturally fractured reservoir · Discrete-fracture-matrix model · Uncertainty quantification · History matching · Model inversion

1 Introduction

Naturally fractured reservoirs (NFRs), which account for a significant portion of worldwide oil and gas resources, pose significant challenges for reliable reservoir performance forecasting due to their complex geology and high degree of heterogeneity. Discrete-fracture-matrix (DFM) modeling methods have been developed and applied to simulate flow in NFRs [7, 9, 12, 15, 16, 21], and models of this type are now well established. Before these approaches can be used for practical reservoir management, however, they must first be capable of matching observed data and providing appropriate uncertainty quantification. A number of existing data assimilation, or history-matching, approaches (see [24, 25, 31, 32] for comprehensive discussions) are based on the assumption of a fixed grid that is invariant between realizations. Further, the grid is often assumed not to change between prior and posterior (i.e., history-matched) models. These assumptions may be problematic for DFM models, as

✉ Wenyue Sun
wenyue@stanford.edu

¹ Department of Energy Resources Engineering,
Stanford University, Stanford, CA, 94305, USA

² Chevron ETC, San Ramon, CA, 94583, USA

the simulation grid can change substantially when fractures are added, removed, or shifted. Thus, in order to use DFM models for practical decision making, it will be important to devise data-assimilation procedures that are not impacted by variability in grid structure.

A method recently presented by Sun and Durlofsky [30], which is referred to as a data-space inversion (DSI) procedure, has some features that may render it applicable for data assimilation with complex DFM models. DSI works solely in the data space, with the goal of providing production forecasts and associated uncertainty. The method does not involve posterior (history-matched) geological models, so limitations related to gridding and model representation do not arise. The basic idea in DSI is to view production data as a random vector with a prior distribution estimated from flow results that are generated by simulating an ensemble of prior-model realizations. The determination of the conditional distribution given observed data is then posed within a Bayesian framework, and this distribution is sampled using a data-space randomized maximum likelihood (RML) method [18, 23, 26]. Sun and Durlofsky [30] applied the DSI procedure to a bimodal channelized system and to a Gaussian permeability model. In both cases, they observed close agreement between results from DSI and those from a strict rejection sampling procedure for key quantities such as P10, P50, and P90 production rate predictions. The time-consuming step in DSI is the simulation of the prior models (in the cases considered in [30], 500 such simulations were used), though these can all be run in parallel.

The DSI method shares some similarities with a number of ensemble-based approaches, such as ensemble Kalman filter (EnKF) procedures [1, 5, 34] and ensemble smoother (ES) techniques [2, 6]. These approaches start from an initial ensemble of prior models and associated simulation results, which are assembled into so-called state vectors. For reservoir-forecasting problems, the state vectors include both static model parameters and dynamic production data, which are jointly updated during each iteration when assimilating observed data. Because EnKF and ES techniques require the representation of multiple reservoir models on the same (fixed) grid, these approaches are not directly applicable to the DFM models considered in this paper.

Previous studies have addressed the problem of data assimilation for NFRs. These include the use of the gradual deformation method for history-matching a two-dimensional fractured model [10, 14]. With this treatment, only a single history-matched model is generated, and it is not clear how to properly construct multiple posterior models for uncertainty quantification. In other work, Cherpeau et al. [3] applied a Metropolis sampling algorithm [22] to generate multiple posterior models for a fractured system. In this case, however, only a few large-scale fractures, in contrast to the numerous fractures in NFR systems,

were considered. For the Tengiz field, a real NFR system, Dehghani et al. [4] and King et al. [17] applied experimental design procedures for uncertainty assessment. These studies used small numbers of experiments, however, so the uncertainty quantification was somewhat heuristic.

Several approaches have also been presented to generate forecasts without updating model parameters. Scheidt et al. [28] and Satija and Caers [27] proposed prediction-focused analysis (PFA) approaches that aim at building a statistical relationship between simulated well data, obtained from an ensemble of prior models, during the history-matching and forecasting periods. When data are observed, the corresponding forecasts are computed using the constructed statistical relationship. Krishnamurti et al. [19] and Mallet et al. [20] investigated multi-model ensemble-forecasting approaches, in which the forecasts are obtained as the weighted average of results generated from simulating multiple plausible physical models. The weights are determined by constraining the forecasts to match observed data. Both the PFA and multi-model ensemble-forecasting approaches have some shortcomings, however, which include limitations on the amount of data that can be handled and limited treatment of measurement error (see [29] and [30] for further discussion of these issues). This may limit the applicability of these methods for realistic production forecasting with DFM models.

In this paper, we extend the general DSI procedure and then apply it for data assimilation in complex DFM models. Because DSI works in data space, it is already well-suited for this problem. However, as we will show later, DSI requires the reparameterization of production data variables, and the approach developed for this by Sun and Durlofsky [30] cannot be directly applied to some of the cases considered here. This is because our previous treatment was based on the existence of a few production stages with discernible features or patterns (e.g., primary production at constant oil rate, followed by primary production at a specified minimum bottom-hole pressure, followed by water injection, etc). The stages that appear for the realistic system considered in this study are more complicated, in that wells experience multiple shut-ins, and this reduces the applicability of our previous treatment. We thus introduce a general histogram transform technique, in conjunction with principal component analysis (PCA), to reparameterize the production data. This new treatment is then incorporated into the DSI formulation. The generation of multiple conditional production forecasts is again accomplished using a data-space RML method. We emphasize that, even though the simulation of realistic DFM models can be time consuming, the DSI procedure enables the efficient generation of multiple forecasts once the prior models are simulated. Additional advantages and limitations of the DSI approach are discussed in [30].

This paper proceeds as follows. In Section 2, we review the general DSI framework introduced in [30]. Reparameterization of data variables using principal component analysis and histogram transformation is described in Section 3. In Section 4, the DSI procedure is applied to synthetic two-dimensional DFM systems involving realizations from a range of geological scenarios, and the DSI results are compared to those from a rejection sampling procedure. Then, in Section 5, we present DSI results for realistic NFR models. We conclude with a summary and suggestions for future work in Section 6.

2 Data-space inversion formulation

In this section, we provide a brief description of the DSI formulation introduced by Sun and Durlofsky [30]. Please refer to that paper for a complete description of the basic procedure.

Let \mathbf{m} denote the vector containing reservoir model parameters such as grid-block permeabilities and porosities. The forward model relating model parameters to the dynamic production data is expressed as

$$\mathbf{d}_{\text{full}} = \mathbf{g}(\mathbf{m}), \tag{1}$$

where \mathbf{d}_{full} is a column vector of dimension N_d that includes dynamic data of different types (e.g., oil and water production rates) from multiple wells at different time steps. Data contained in \mathbf{d}_{full} are normally uncertain as the model parameters in \mathbf{m} are initially uncertain. We refer to the elements in \mathbf{d}_{full} as data variables. We consider \mathbf{d}_{full} to include data variables during both the history-matching and forecast periods, i.e., $\mathbf{d}_{\text{full}} = [\mathbf{d}_{\text{hm}}^T, \mathbf{d}_{\text{pred}}^T]^T$, where \mathbf{d}_{hm} and \mathbf{d}_{pred} are column vectors containing data variables during the history-matching and forecast/prediction periods, respectively.

The observed (measured) data variables during the history-matching period are assembled into a column vector, \mathbf{d}_{obs} , of dimension N_{obs} . Our goal is to predict the future quantities of interest, \mathbf{d}_{pred} , conditioned to observations \mathbf{d}_{obs} . We emphasize that the goal of the DSI procedure is to generate estimates of \mathbf{d}_{pred} without calibrating model parameters.

We begin by writing the conditional probability density function (PDF) of \mathbf{d}_{full} given \mathbf{d}_{obs} as

$$p(\mathbf{d}_{\text{full}}|\mathbf{d}_{\text{obs}}) = \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{d}_{\text{full}})p(\mathbf{d}_{\text{full}})}{p(\mathbf{d}_{\text{obs}})} \propto p(\mathbf{d}_{\text{obs}}|\mathbf{d}_{\text{full}})p(\mathbf{d}_{\text{full}}), \tag{2}$$

where $p(\mathbf{d}_{\text{full}})$ is the prior PDF of \mathbf{d}_{full} and $p(\mathbf{d}_{\text{obs}}|\mathbf{d}_{\text{full}})$, referred to as the likelihood function of \mathbf{d}_{full} , represents the

conditional PDF of observing \mathbf{d}_{obs} given \mathbf{d}_{full} . The production forecasts, which comprise part of \mathbf{d}_{full} , can be generated by sampling the posterior distribution of \mathbf{d}_{full} using Eq. 2 once the likelihood function and prior PDF are defined.

Because \mathbf{d}_{obs} contains direct measurements of \mathbf{d}_{hm} , the relationship between \mathbf{d}_{obs} and \mathbf{d}_{full} can be written as

$$\mathbf{d}_{\text{obs}} = H\mathbf{d}_{\text{full}} + \boldsymbol{\epsilon}, \tag{3}$$

where H is an $N_{\text{obs}} \times N_d$ matrix, consisting of zeros and ones, that simply extracts the elements in \mathbf{d}_{full} corresponding to \mathbf{d}_{hm} , and $\boldsymbol{\epsilon}$ denotes the measurement error vector. The measurement errors are assumed to be Gaussian distributed with zero mean and covariance matrix C_D . Therefore, the likelihood function of \mathbf{d}_{full} is Gaussian and can be written as follows:

$$p(\mathbf{d}_{\text{obs}}|\mathbf{d}_{\text{full}}) = p(\boldsymbol{\epsilon} = \mathbf{d}_{\text{obs}} - H\mathbf{d}_{\text{full}}) \propto \exp\left(-\frac{1}{2}(\mathbf{d}_{\text{obs}} - H\mathbf{d}_{\text{full}})^T C_D^{-1}(\mathbf{d}_{\text{obs}} - H\mathbf{d}_{\text{full}})\right). \tag{4}$$

Sun and Durlofsky [30] provided analytical solutions for the posterior distribution $p(\mathbf{d}_{\text{full}}|\mathbf{d}_{\text{obs}})$ under the condition that the prior PDF of \mathbf{d}_{full} is Gaussian. For cases when this prior PDF is far from Gaussian, they showed that improved DSI performance can be obtained by reparameterizing the data variables in \mathbf{d}_{full} into new variables that are nearly multivariate Gaussian. If we let $\boldsymbol{\eta}$ denote the new (approximately) multivariate Gaussian vector with prior mean $\boldsymbol{\eta}_{\text{prior}}$ and covariance matrix C_η , the relationship between \mathbf{d}_{full} and $\boldsymbol{\eta}$ can be written as

$$\mathbf{d}_{\text{full}} = \mathbf{f}(\boldsymbol{\eta}). \tag{5}$$

Thus, instead of directly sampling $p(\mathbf{d}_{\text{full}}|\mathbf{d}_{\text{obs}})$ from Eq. 2, we first formulate and sample the conditional distribution of $\boldsymbol{\eta}$ given \mathbf{d}_{obs} . The ensemble of forecasts for \mathbf{d}_{full} can then be generated by applying Eq. 5 to the corresponding conditional samples of $\boldsymbol{\eta}$. Combining Eqs. 4 and 5, the conditional distribution of $\boldsymbol{\eta}$ can be expressed as

$$p(\boldsymbol{\eta}|\mathbf{d}_{\text{obs}}) \propto p(\mathbf{d}_{\text{obs}}|\boldsymbol{\eta})p(\boldsymbol{\eta}) \propto \exp\left(-\frac{1}{2}(\mathbf{H}\mathbf{f}(\boldsymbol{\eta}) - \mathbf{d}_{\text{obs}})^T C_D^{-1}(\mathbf{H}\mathbf{f}(\boldsymbol{\eta}) - \mathbf{d}_{\text{obs}}) - \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_{\text{prior}})^T C_\eta^{-1}(\boldsymbol{\eta} - \boldsymbol{\eta}_{\text{prior}})\right). \tag{6}$$

In the following section, we describe the reparameterization of data variables and then present the data-space randomized maximum likelihood (RML) method to sample the resulting posterior distribution.

3 Reparameterization of data variables

As discussed in the Introduction, Sun and Durlofsky [30] introduced a pattern (or production-stage) based mapping technique suitable for handling time-series production data variables that exhibit a few discernible patterns or features. This approach requires pre-specifying the production stages (and data patterns), and it may encounter difficulties when the production data display complex behavior with no obvious patterns, as occurs with the data considered in this paper. In this section, we describe a new reparameterization procedure that does not require any pre-specification of data patterns and is thus applicable to any type of production data.

3.1 PCA representation of data variables

DSI entails first performing flow simulation for multiple (N_r) independently generated prior reservoir models. We designate $(\mathbf{d}_{full})_1$ to $(\mathbf{d}_{full})_{N_r}$ to be the data vectors obtained by solving Eq. 1 for prior models $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{N_r}$. We then apply principal component analysis (PCA) to the simulated data. As discussed in [30], this treatment eliminates the need to compute the inverse of C_η , and also reduces the

number of variables to be determined. To apply PCA, we first construct the centered matrix $X \in \mathbb{R}^{N_d \times N_r}$ as

$$X = [(\mathbf{d}_{full})_1 - \mathbf{d}_{prior} \quad (\mathbf{d}_{full})_2 - \mathbf{d}_{prior} \quad \dots \quad \dots \quad (\mathbf{d}_{full})_{N_r} - \mathbf{d}_{prior}]. \tag{7}$$

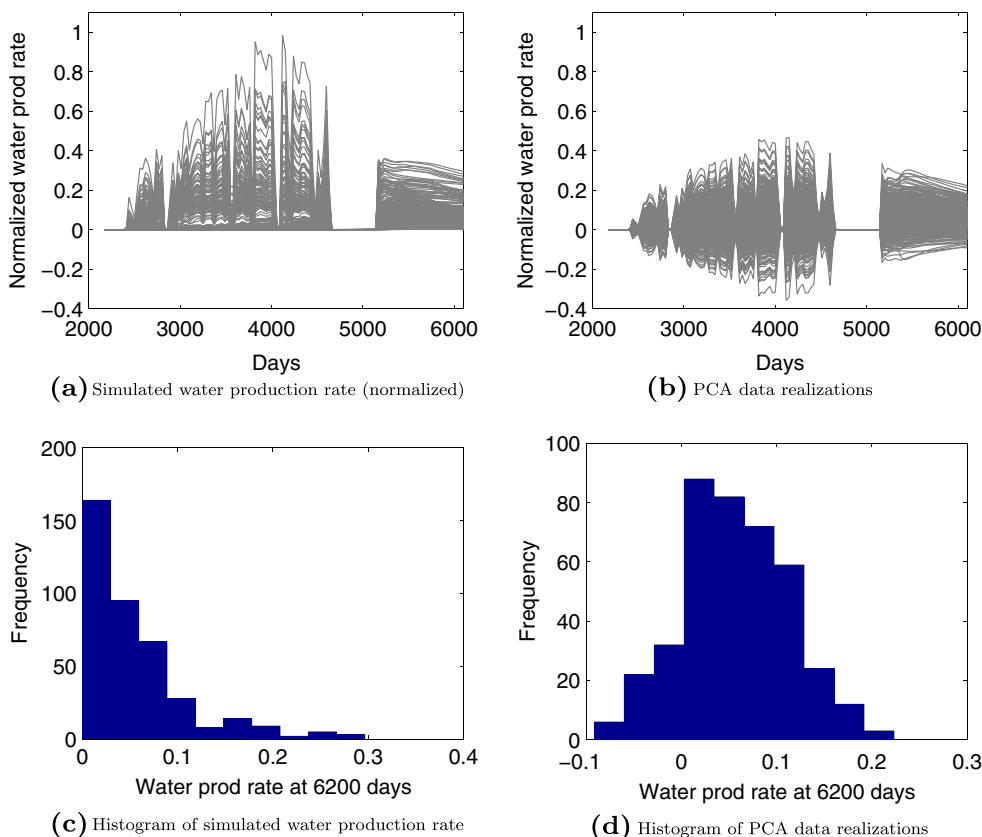
We assume there is no modeling error associated with the forward model (Eq. 1). Therefore, $(\mathbf{d}_{full})_i, i = 1, 2, \dots, N_r$ can be viewed as prior realizations of the random vector \mathbf{d}_{full} , with the prior mean of \mathbf{d}_{full} estimated as $\mathbf{d}_{prior} = (1/N_r) \sum_{i=1}^{N_r} (\mathbf{d}_{full})_i$.

We then perform singular value decomposition of $X/\sqrt{N_r - 1}$, which provides the eigenvalues and eigenvectors of the covariance matrix $XX^T/(N_r - 1)$. The square root of the eigenvalues (ordered by descending value) form the diagonal elements of the matrix Σ , and the corresponding eigenvectors comprise the columns of the matrix U . Thus, new data realizations, which we refer to as \mathbf{d}_{PCA} , can be generated through application of

$$\mathbf{d}_{PCA} = U \Sigma \boldsymbol{\xi} + \mathbf{d}_{prior} = \Phi \boldsymbol{\xi} + \mathbf{d}_{prior}, \tag{8}$$

where $\boldsymbol{\xi}$ is a vector containing uncorrelated variables with standard normal distribution, and Φ is the basis matrix. The

Fig. 1 Example illustrating the impact of the PCA parameterization of production data



leading columns in Φ are associated with the largest eigenvalues of the covariance matrix $XX^T/(N_r - 1)$. In general, only the first $N_l < N_r$ columns are retained, which means that $\Phi \in \mathbb{R}^{N_d \times N_l}$ and $\xi \in \mathbb{R}^{N_l \times 1}$. See [30] for further details on the construction of Φ .

We now apply the PCA procedure described above to simulated water production rate (WPR) data from a set of 400 prior models. The data, shown in Fig. 1a, are normalized by the maximum WPR value among all wells at all time steps. The prior geological models, which are intended to represent a real naturally fractured reservoir, will be described below. The data shown in Fig. 1a are reported monthly for 11 years, from around 2200 to 6200 days. Assuming for now that data vector \mathbf{d}_{full} contains only these WPR data, with $N_d = 132$, we have $(\mathbf{d}_{full})_i \in \mathbb{R}^{132 \times 1}$ ($i = 1, 2, \dots, N_r$). Using the PCA parameterization (Eq. 8), we generate 400 new data realizations. These are shown as the gray curves in Fig 1b. It is clear that the PCA realizations display unphysical behavior, particularly negative WPR values. Figure 1c and d shows the histograms of the simulated data and PCA data realizations at 6200 days. The histogram from the PCA realizations (Fig. 1d) is close to Gaussian and clearly quite different than the histogram from the simulated data.

3.2 Histogram transformation

To improve the ability of the PCA parameterization to preserve the physical character of the production data variables, we introduce an additional histogram transformation step. The basic idea is to transform a PCA realization \mathbf{d}_{PCA} (obtained using Eq. 8) to another realization $\tilde{\mathbf{d}} = h_T(\mathbf{d}_{PCA})$ that has a specified target histogram for each variable in $\tilde{\mathbf{d}}$. In this paper, the target histograms/distributions are obtained empirically from the ensemble of simulated data, as these data vectors represent the prior realizations of random vector \mathbf{d}_{full} (Fig. 1c, for example, shows the target histogram of WPR at 6200 days).

Because \mathbf{d}_{PCA} is a linear transformation of the multivariate (standard) normally-distributed vector ξ (Eq. 8), the distribution of \mathbf{d}_{PCA} is multivariate Gaussian with mean \mathbf{d}_{prior} and covariance matrix $\Phi\Phi^T$. Thus, the initial histogram/distribution of each variable in \mathbf{d}_{PCA} is Gaussian, with mean defined by the corresponding row of \mathbf{d}_{prior} and variance defined by the diagonal term in $\Phi\Phi^T$. Let $f_I(d_{PCA}^j)$ and $f_T(d_{PCA}^j)$ denote the initial and target cumulative distribution functions (CDFs) for row j in \mathbf{d}_{PCA} . Then, each element in the target realization $\tilde{\mathbf{d}}$ is obtained as $\tilde{d}_{PCA}^j = f_T^{-1} f_I(d_{PCA}^j)$, $j = 1, \dots, N_d$.

Figure 2 shows the PCA data realizations before and after this histogram transformation. It is clear that this treatment significantly improves the ability of PCA to reproduce the

physical, non-negative character of the simulated WPR data (Fig. 1a). The red, blue, and green curves in Fig. 2a indicate three particular PCA data realizations. The corresponding realizations after histogram transformation are shown in Fig. 2b. Although the red curve in Fig. 2a displays negative WPR values, these are all non-negative after the histogram transformation.

3.3 DSI formulation with reparameterization

Incorporating PCA and the histogram transform, we express the data vector \mathbf{d}_{full} as a function of ξ as

$$\mathbf{d}_{full} \approx \mathbf{f}(\xi) = h_T(\mathbf{d}_{PCA}) = h_T(\Phi\xi + \mathbf{d}_{prior}). \tag{9}$$

We take ξ to be the reparameterized vector representing \mathbf{d}_{full} in Eq. 5, that is $\eta = \xi$. Because ξ is multivariate standard normal, Eq. 6 becomes

$$p(\xi|\mathbf{d}_{obs}) \propto \exp\left(-\frac{1}{2}(\mathbf{H}\mathbf{f}(\xi) - \mathbf{d}_{obs})^T C_D^{-1}(\mathbf{H}\mathbf{f}(\xi) - \mathbf{d}_{obs}) - \frac{1}{2}\xi^T \xi\right), \tag{10}$$

where $\mathbf{f}(\xi)$ is defined in Eq. 9. Equation 10 characterizes the posterior distribution of new parameters ξ given \mathbf{d}_{obs} . Note that the (possibly ill-conditioned) $(\eta - \eta_{prior})^T C_\eta^{-1}(\eta - \eta_{prior})$ term in Eq. 6 no longer appears. Rather, after application of PCA, this term now appears as $\xi^T \xi$.

To sample this distribution, we apply the randomized maximum likelihood (RML) method in the data space. The objective function is thus defined as

$$\xi_{rml} = \arg \min_{\xi} \left[(\mathbf{H}\mathbf{f}(\xi) - \mathbf{d}_{obs}^*)^T C_D^{-1}(\mathbf{H}\mathbf{f}(\xi) - \mathbf{d}_{obs}^*) + (\xi - \xi^*)^T (\xi - \xi^*) \right]. \tag{11}$$

Here, \mathbf{d}_{obs}^* are perturbed observations from the Gaussian distribution $N[\mathbf{d}_{obs}, C_D]$, and ξ^* are sampled from the standard normal distribution $N[\mathbf{0}, I]$. In this work, the quasi-Newton line search algorithm within Matlab is used to find the minimum of the objective function in Eq. 11. This minimization is generally accomplished in just a few seconds since the objective function evaluations are very fast.

We have now completed the general description of the DSI procedure with our new reparameterization. Additional algorithmic details are provided in [30]. The overall approach applied here can be summarized as follows:

1. Perform flow simulation on an ensemble of prior model realizations, conditioned to prior geological information, to obtain simulated data $(\mathbf{d}_{full})_i$, $i = 1, 2, \dots, N_r$.
2. Using PCA and histogram transformation, represent data variables \mathbf{d}_{full} as a function of ξ (Eq. 9).

- Apply the RML method (in the data space) to sample the conditional distribution of ξ given \mathbf{d}_{obs} (Eq. 10). Then, generate the corresponding forecasts through application of Eq. 9.

It is important to reiterate that the only time-consuming step in the entire procedure is performing the flow simulations for the prior models. These simulations can all be run in parallel, so the elapsed (wall-clock) time might correspond to only the time required for a few simulations if a sufficient number of processors are available. If necessary, these runs can be accelerated by using upscaled DFM models. The recent procedure presented in [15] could be applied for this purpose. The total computation time needed to generate multiple forecasts with the data-space RML procedure is on the order of minutes.

4 Numerical results for synthetic models

In this section, the DSI procedure described above is applied to synthetic two-dimensional DFM examples. The prior models in these cases involve DFM realizations generated from a range of fracture scenarios. The DSI results for uncertainty quantification will be compared to those obtained from rejection sampling (RS).

4.1 Problem setup and prior-model predictions

The model covers an area of 2100 m \times 2100 m and is of thickness 100 m. An oil-water system is considered. Viscosities of oil and water at standard conditions are 1 cp and 0.3 cp, respectively. The initial oil and water saturations are 0.9 and 0.1. The oil and water matrix relative permeability functions (k_{ro}^m and k_{rw}^m) are specified as

$$k_{ro}^m = \left(\frac{1 - S_{or} - S_w}{0.9 - S_{or}} \right)^2, \quad k_{rw}^m = 0.8 \left(\frac{S_w - 0.1}{0.9 - S_{or}} \right)^2, \quad (12)$$

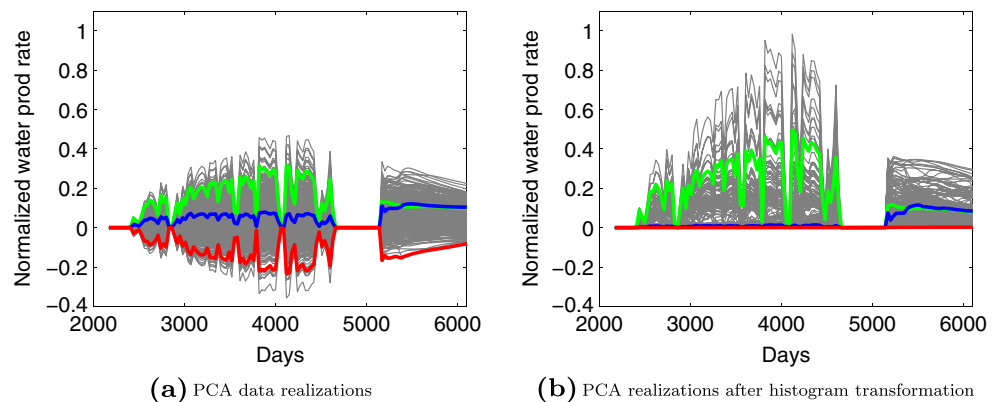
where S_w denotes water saturation and S_{or} is the residual oil saturation (in this example S_{or} will be treated as an uncertain parameter). In the fractures, straight-line relative permeability functions are specified for both phases, i.e., $k_{ro}^f = 1 - S_w$ and $k_{rw}^f = S_w$. Capillary pressure effects are ignored. The matrix and fracture porosities are assumed to be constant at 0.2 and 0.3, respectively. One water injection well is placed at the center of the model, and four production wells are located near the corners, as shown in Fig. 3. The injector operates at a fixed bottom-hole pressure (BHP) of 3600 psi, and all producers operate at a fixed oil production rate (OPR) of 700 STB/day, subject to a minimum BHP of 2500 psi.

Figure 3 shows two model realizations corresponding to different fracture densities and orientations. Fractures are displayed as black or red lines and the background gray region corresponds to matrix. The red fractures are the fractures intersected by wells. These are taken to be hard data and are the same in all realizations. In each model, there are two sets of black fractures, which are assumed to be orthogonal to each other.

Table 1 presents the six model parameters that are considered to be uncertain. Fracture density denotes the average fracture surface area per rock volume. The fracture densities for the models shown in Fig. 3a and b are 0.006 and 0.01/m, respectively. Fracture orientation, in the x - y plane, represents the orientation for one set of fractures (the other set of fractures is oriented orthogonally). Matrix permeability is uncertain but always nonzero, which enables some flow through the matrix. The fracture permeability k^f depends on the fracture aperture d via the relationship $k^f = 1000d^2$ (the units of k^f and d are md and mm, so for $d = 10$ mm, $k^f = 10^5$ md). Finally, residual oil saturation, which impacts the relative permeability curves (see Eq. 12), and rock compressibility, are also considered to be uncertain.

All parameters are assumed to be independently and uniformly distributed a priori within the specified ranges shown

Fig. 2 Example illustrating improvement in WPR realizations when PCA is used in conjunction with histogram transformation



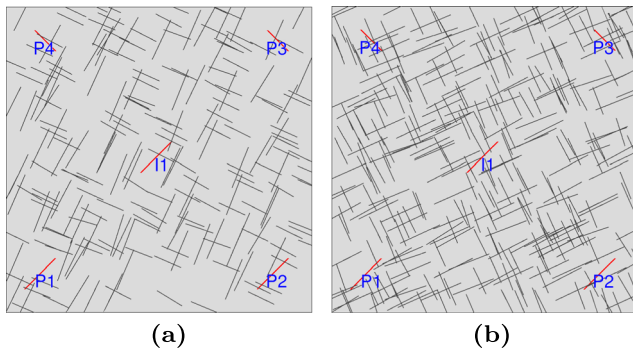


Fig. 3 Fracture networks for two DFM realizations. Fractures are shown in *black* or *red*, and matrix is in *gray*. Injection (I1) and production (P1 to P4) wells are also displayed

in Table 1. Any set of these six model parameters defines a particular geological scenario. Within a scenario, an infinite number of realizations can be generated. In this work, we randomly select 100 geological scenarios (consistent with the ranges in Table 1), and for each of these scenarios we generate 100 fracture realizations. This results in a total of 10,000 prior-model realizations. It is important to emphasize that the fracture networks are different for all realizations, though realizations in the same scenario display the same average fracture density and orientation.

Figure 4 presents simulation results, for injector I1 and producers P1 and P2, for 200 prior models (selected randomly from the set of 10,000). The simulation data at all wells are reported every 60 days over a total simulation time frame of 7680 days. The variability in these prior simulation results is very large. For example, the predicted OPR plateau periods, shown in Fig. 4c and f, range from the start to the end of simulation time frame for both P1 and P2. In addition, the impact of the (uncertain) fracture network on flow can be seen from the water production rate (WPR) results. When the injector and producer are connected by fractures, the injected water reaches the producer quickly. This phenomenon is evident in Fig. 4d, where we see very early water breakthrough times (less than 100 days) for some realizations.

Table 1 Uncertain model parameters, ranges, and values for ‘true’ models. All parameters are assumed to be uniformly distributed a priori

Parameter	Lower bound	Upper bound	‘True’ model 1	‘True’ model 2
Fracture density ($\times 10^{-3}/\text{m}$)	6	10	8.8	6.6
Fracture orientation	25	65	65	62
Fracture aperture (mm)	8	12	8	9.4
Matrix permeability (mD)	0.3	0.5	0.43	0.34
Rock compressibility ($\times 10^{-5}/\text{psi}$)	1	5	2.1	4.3
Residual oil saturation	0.1	0.3	0.19	0.16

4.2 Rejection sampling procedure

Sun and Durlofsky [30] applied a rejection sampling (RS) procedure to provide reference results for uncertainty quantification. Though computationally expensive, RS can provide a proper sampling of the posterior distribution of the production data. In this study, we apply a rejection sampling procedure that differs slightly from that used in [30]. The method now proceeds as follows:

1. Generate model realization \mathbf{m} consistent with prior geological information.
2. Sample a variable p from a uniform distribution within the range $[0,1]$.
3. Accept \mathbf{m} as a posterior realization if $p \leq L(\mathbf{m})/S_L$, where $L(\mathbf{m})$ is the likelihood function and S_L is a value that is greater than or equal to the maximum of the likelihood function.

The likelihood function $L(\mathbf{m})$ is defined as

$$L(\mathbf{m}) = c \exp\left(-\frac{1}{2}O(\mathbf{m})\right), \tag{13}$$

where c is a normalization constant. The quantity $O(\mathbf{m})$ is the weighted mismatch between observed and simulated data, written here as

$$O(\mathbf{m}) = \frac{1}{N_{\text{obs}}}(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}})^T C_D^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}). \tag{14}$$

Because $O(\mathbf{m})$ is nonnegative, we have $L(\mathbf{m}) \leq c$ for any model \mathbf{m} . Therefore, we set S_L to be c in this study. It is important to observe that the mismatch $O(\mathbf{m})$ in Eq. 14 is computed as the mean of the weighted data mismatch, rather than as the sum of the squared data mismatch. The latter representation was used in [30], meaning that the factor of $1/N_{\text{obs}}$ in Eq. 14 did not appear in that work. The use of the $1/N_{\text{obs}}$ factor decreases the value of the mismatch function $O(\mathbf{m})$ and thus results in the acceptance of more prior models within the RS procedure (in the absence of this scaling, few if any of the 10,000 prior model results would be accepted).

For the purposes of strict uncertainty quantification, we do not believe it is appropriate to introduce this $1/N_{\text{obs}}$

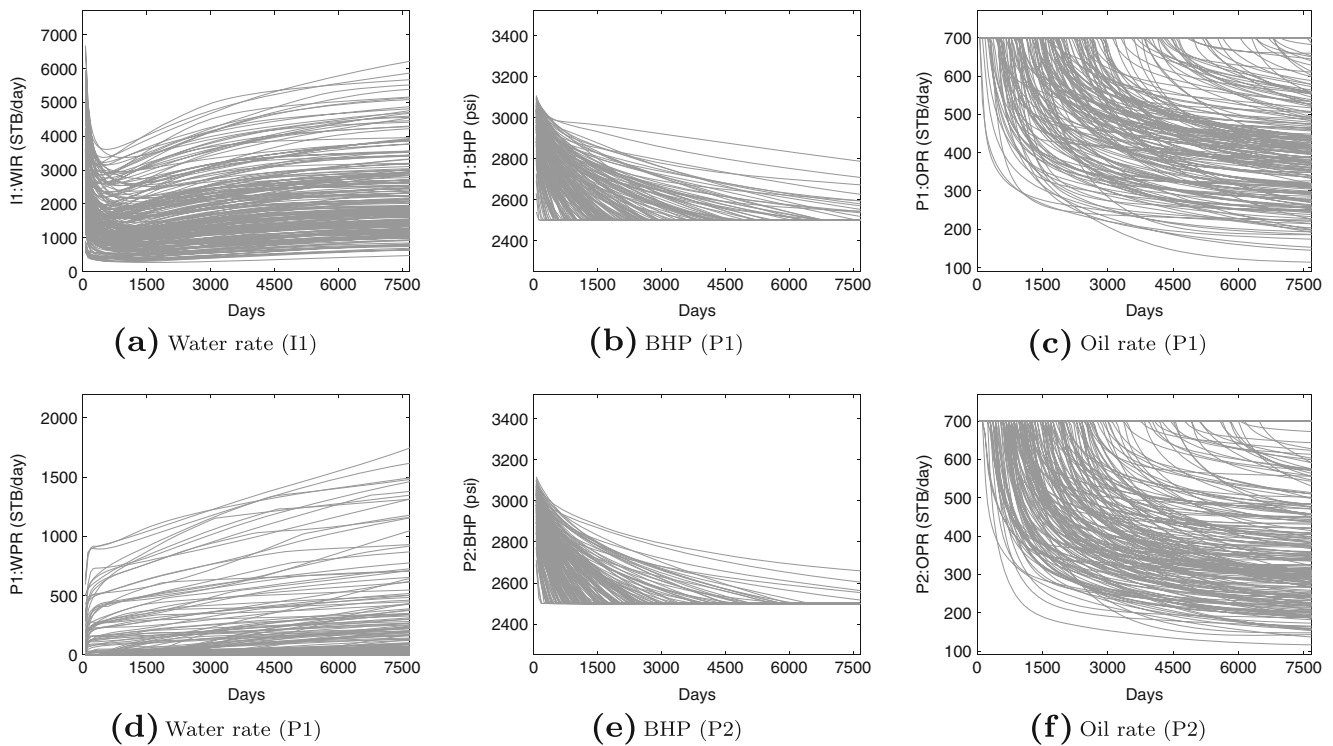


Fig. 4 Simulated production results from prior models (two-dimensional synthetic case)

factor when we have zero model error and Gaussian measurement error (as we have here). Thus, we cannot claim that the resulting RS predictions provide the actual uncertainty assessment. However, the comparison between the DSI and RS results will still be valid as long as the likelihood functions used for DSI and RS are consistent. This means we can still use the RS results to assess the DSI procedure. To enable a valid comparison with RS, the DSI objective function (Eq. 11) is now written as

$$\xi_{\text{rml}} = \arg \min_{\xi} \left[(\xi - \xi^*)^T (\xi - \xi^*) + (Hf(\xi) - \mathbf{d}_{\text{obs}}^*)^T \times (N_{\text{obs}} C_D)^{-1} (Hf(\xi) - \mathbf{d}_{\text{obs}}^*) \right], \quad (15)$$

where we have $(N_{\text{obs}} C_D)^{-1}$ instead of C_D^{-1} in the data mismatch term. In addition to the use of Eq. 15 in place of Eq. 11, in the DSI RML procedure we must also sample (perturbed) observations from the Gaussian distribution $N[\mathbf{d}_{\text{obs}}, N_{\text{obs}} C_D]$ rather than from $N[\mathbf{d}_{\text{obs}}, C_D]$.

4.3 Results for two ‘true’ models

We now apply the DSI procedure for two different ‘true’ synthetic two-dimensional models. ‘True’ model 1 corresponds to predictions that, visually, fall around the middle of the range of predictions from the prior models (Fig. 4),

and ‘true’ model 2 corresponds to predictions toward the edge of the ranges. In both cases, observed data are generated by adding random noise to the true data (true data are the predictions from the selected ‘true’ model). The noise is sampled from the Gaussian distribution $N[0, C_D]$. The standard deviations of the noise are set to be 30 psi for BHP data, 100 STB/day for WIR data, and 5% of the corresponding true data, with a maximum of 20 STB/day for WPR and OPR data (as noted above, in the RML computations we perturb the observations with noise sampled from $N[0, N_{\text{obs}} C_D]$ rather than $N[0, C_D]$). The historical period is taken to be the first 1200 days, with well data observed every 60 days at I1, P1, and P2, which results in 70 observed data points. We focus on forecasts for production data for I1, P1, and P2. The total simulation time is 7680 days and we take uniform 60-day time steps. We thus have $N_d = 7 \times 128 = 896$, and $(\mathbf{d}_{\text{full}}) \in \mathbb{R}^{896 \times 1}$.

As discussed in Section 4.1, we generated a total of 10,000 prior realizations from a set of 100 geological scenarios. For the RS procedure, we use 9800 of the 10,000 prior models. The 200 models discarded correspond to those for the two geological scenarios from which the two ‘true’ models were generated (which means that the true scenario is not represented in the set of prior realizations). For the DSI method, we use a total of $N_r = 1000$ prior models, which are randomly selected from the set of 9800 prior models used with RS. See [30] for a discussion of the choice of

N_r . As noted there, if too few models are accepted based on the RML offset criterion, this suggests that more prior models are needed. In this paper, we use the same ‘energy’ criterion described in [30] to determine the number of principal components to retain in the Φ matrix in Eq. 8. In this case, a total of $N_l = 9$ principal components are required.

Forecasting results for ‘true’ model 1, in terms of P10, P50, and P90 statistics, from prior models, RS and DSI, are presented in Fig. 5. The red dots represent observed data, and the dashed red curves denote the true data. The gray regions indicate the P10–P90 range of forecasts from prior models, which are obtained using the ensemble of prior forecasts shown in Fig. 4. In the RS procedure, 268 models (out of the 9800 simulated) are accepted. For DSI, a total of 100 forecasts are generated. The posterior P10, P50, and P90 statistics estimated from the RS and DSI forecasts are shown as dashed blue and black curves in Fig. 5. The RS and DSI results are seen to be in generally close agreement (though some differences are evident in the P90 results in Fig. 5d), which demonstrates that DSI is able to provide uncertainty quantification results that are consistent with those from a much more expensive RS procedure. We reiterate that, for the results in this paper, this consistency is demonstrated for the likelihood function in Eq. 14, though we could also apply DSI without the $1/N_{\text{obs}}$ scaling.

The DSI forecasts for all quantities in Fig. 5 display narrower uncertainty ranges than the prior forecasts. Uncertainty is reduced dramatically for, e.g., P1 oil and water production rates. Also of interest is the fact that the true data (dashed red curves) consistently fall within the P10–P90 range of the DSI forecasts.

We now consider ‘true’ model 2. Prior and posterior flow results for this case are shown in Fig. 6. As noted earlier, the flow responses for ‘true’ model 2 are toward the edge of the prior-model predictions. In fact, the true data fall outside of the P10–P90 range of forecasts from the prior models for I1 WIR (Fig. 6a) and for P2 BHP and OPR (Fig. 6e and f). As a result, only 27 models out of the 9800 prior models considered are accepted in the RS procedure (in contrast to 268 for ‘true’ model 1). Again, a total of 100 DSI forecasts are generated. As shown in Fig. 6, despite the fact that relatively few data responses lie ‘near’ that of ‘true’ model 2, we see that the DSI results still agree reasonably well with the RS results (some differences are evident, however, for P2 OPR in Fig. 6f). It is noteworthy that the true data are always encompassed within the P10–P90 range of DSI forecasts, even when they fall outside the P10–P90 range of prior forecasts (as in Fig. 6a and f). As was the case for ‘true’ model 1, the uncertainty reduction in the posterior predictions is again significant.

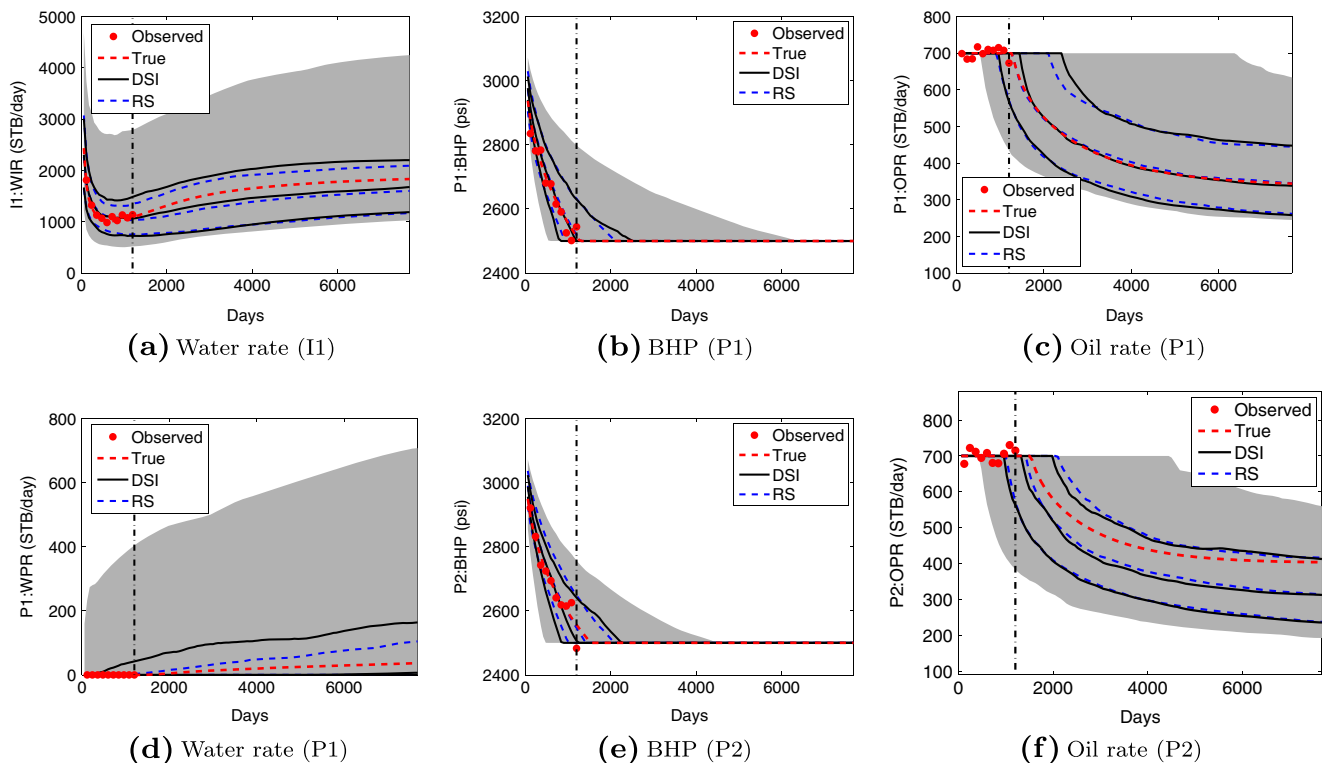


Fig. 5 Statistics of production forecasts for wells I1, P1, and P2 from prior models, RS, and DSI. Red points and red dashed curves indicate observed data and predictions from ‘true’ model 1. Vertical dot-dash

line defines the history-match period. Gray-shaded regions represent the P10 to P90 range of predictions from prior models. Black and blue curves denote the P10, P50, and P90 results from DSI and RS

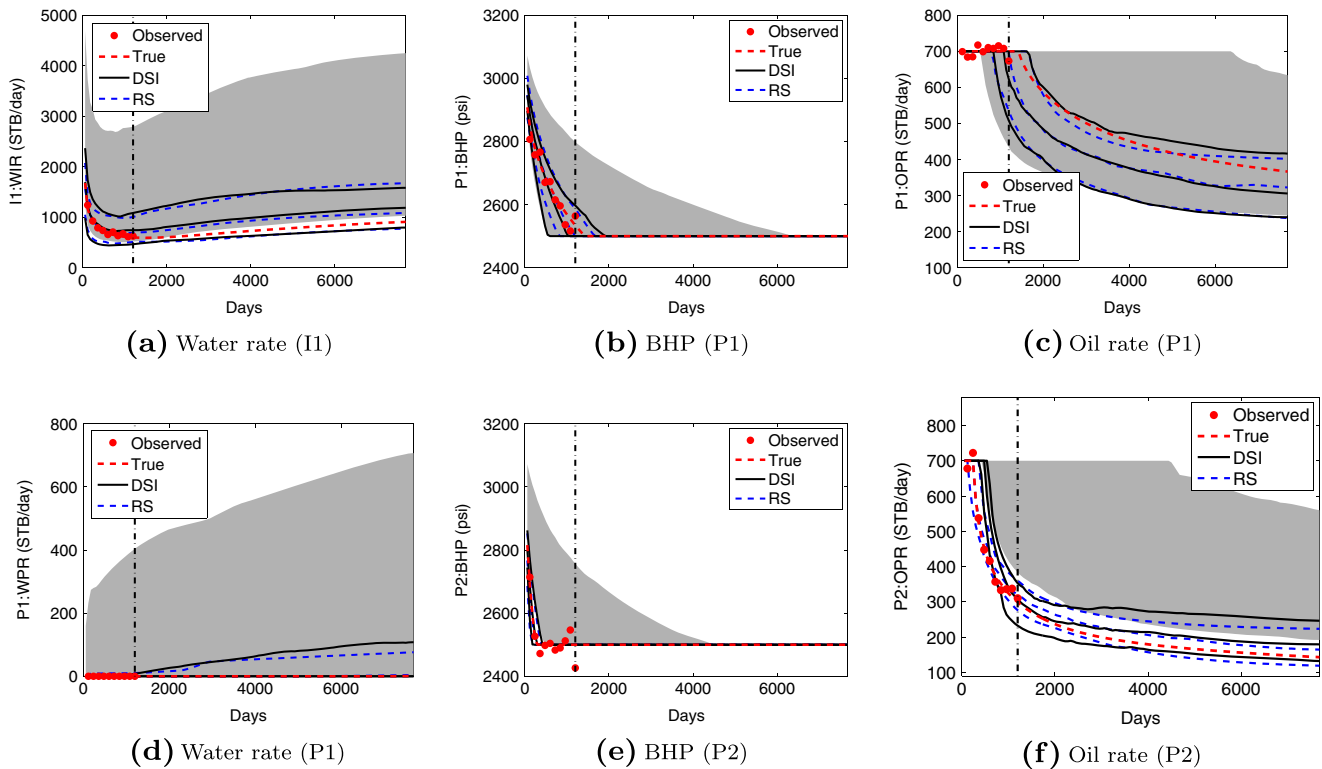


Fig. 6 Statistics of production forecasts for I1, P1, and P2 from prior models, RS, and DSI for ‘true’ model 2. Lines, shadings, and colors have the same meanings as in Fig. 5

We reiterate that neither ‘true’ models 1 and 2, nor any realizations from their corresponding scenarios, were included in the set of prior models used with DSI. This means that, at least for the cases presented above, the DSI method is able to provide reasonable uncertainty quantification, even when the true fracture/matrix property scenario is not contained in the prior ensemble. Although the systems considered here are relatively simple, history-matching this type of model can still be challenging for traditional model-inversion algorithms due to the difficulty of parameterizing fracture networks. The DSI method may indeed be attractive for such systems if the primary goal is the production forecast and associated uncertainty assessment, rather than the generation of a set of posterior models.

5 Numerical results for realistic field models

The results in Section 4 demonstrated the performance of DSI for synthetic two-dimensional cases. We now apply the method using an ensemble of three-dimensional DFM models generated to characterize a real naturally fractured carbonate oil reservoir. This example again involves uncertainty in the geological scenario, though only six different fracture-network realizations are considered

(our emphasis is now on uncertainty in other reservoir parameters, as described below). The realizations contain around 9000 discrete fractures, and the resulting DFM grids include about 900,000 total cells, ~600,000 of which correspond to matrix control volumes, and ~300,000 to fracture control volumes. The basic reservoir model is illustrated in Fig. 7. Many of the fracture and matrix properties, discussed below, are similar to those applied in [11, 13]. There are 11

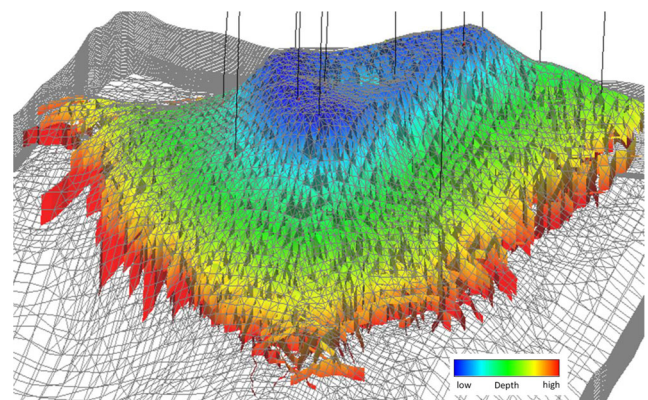


Fig. 7 One of the unstructured DFM grids generated for the naturally fractured reservoir of interest. Color indicates relative depth of discrete-fracture cells. Underlying grid indicated by the gray lines

active producers and one shut-in well that will be evaluated for use as an injector in a pilot waterflood.

We now describe how the DFM model properties are assigned. We use an experimental design involving four uncertain parameters, shown in Table 2, to create an ensemble of ~400 DFM models. In addition to the base-case geologic concept, two other possibilities are considered: one with additional fractures within the aquifer, and one with additional fractures in the target pilot waterflood area. Two stochastic fracture networks are considered for each of these three concepts, giving rise to a total of six fracture realizations. For fracture permeabilities, we specify values that are scaled in accordance with fracture apertures (which are in turn correlated with fracture length) and are calibrated to well test data. Matrix properties such as porosity and irreducible water saturation vary spatially and are derived from a well-calibrated, structured reservoir model of the field. The base-case mean matrix and fracture permeability values differ by 5 orders of magnitude. The fracture and matrix cells are assigned different relative permeabilities, capillary pressures, and rock compressibilities. We specify straight-line relative permeabilities, zero capillary pressure, and higher rock compressibility for fractures. To reflect uncertainty in fracture permeability, fracture pore volume, and matrix permeability, global multipliers are applied for the low and high cases (see Table 2 for multiplier values). Finally, uncertainty in the degree of water imbibition in the matrix is captured through the use of three different water-oil capillary pressure curves.

The actual reservoir has undergone primary depletion for 5400 days. A water-injection pilot, which entails injection for 60 days and monitoring for 800 days, is being considered to evaluate the improved-oil-recovery potential for the field. The water is of viscosity slightly higher than that of the reservoir oil at existing temperature and pressure. Here we will consider a (typical) scenario in which the history-match period is shorter than the prediction period. We thus specify a 2400-day history-match period during which data are collected (except where otherwise indicated), and then consider a prediction period that involves 3000 days of additional primary depletion followed by the 800-day pilot-evaluation

period (for a total prediction period of 3800 days). During the initial 5400-day period, the 11 producers are constrained to operate at their individual historical oil production rates (these wells experience multiple shut-ins for operational reasons). The observed data include static well pressures (SWPs) and water production rates (WPRs).

In the pilot-evaluation period, the shut-in well is used to inject water for 60 days (the time during which water is actually injected is brief to minimize operational disruptions). This well operates at a target rate subject to a maximum bottom-hole pressure (BHP). During the 800-day evaluation period, the producers are specified to operate at prescribed target oil rates subject to minimum BHP constraints. Again, SWPs and WPRs are monitored at all producers during this period.

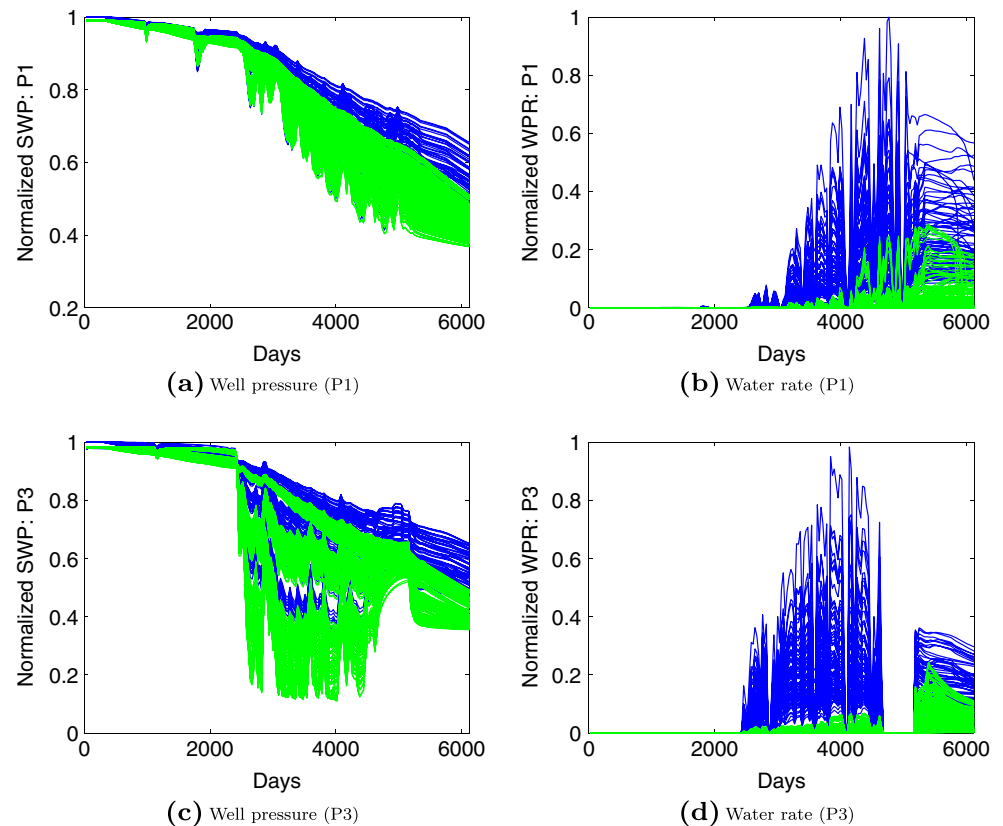
Figure 8 shows simulation results for producers P1 and P3 (which are two wells that display relatively high WPRs), for 398 prior models. The SWP data are normalized by the maximum static well pressure observed for any well at any time. The data plotted in blue correspond to 79 prior models under the geologic concept ‘base + aquifer fractures,’ and the data plotted in green correspond to 319 prior models from the other two geologic concepts (defined above). Because wells are controlled by oil production rate for most of the simulation period (as noted above), we focus on the forecasts for SWP and WPR. In the DSI procedure, the data vectors $(\mathbf{d}_{full})_i, i = 1, 2, \dots, N_r$, include SWP and WPR data at 11 wells (the shut-in water-injection well is not considered). The simulation data at all wells are reported monthly for the total simulation time frame of 6200 days (or 17 years), resulting in $N_d = 17 \times 12 \times 2 \times 11 = 4488$, and $(\mathbf{d}_{full})_i \in \mathbb{R}^{4488 \times 1}$. Though the dimension of the data vector is large, the total number of principal components retained for the basis matrix (Eq. 8) is only 13. This large amount of reduction results from the strong cross-correlation between the data variables in \mathbf{d}_{full} .

Rather than use actual field data, we will evaluate the performance of the DSI procedure for two different geo-statistical ‘true’ models that correspond to different water production behavior. One of these ‘true’ models is selected from the ‘green’ set of models and one is from the ‘blue’ set of models. The true data are then generated by performing flow simulation on the ‘true’ model. Observed data (which are used in DSI) are obtained by adding random Gaussian noise, sampled from $N[0, C_D]$, to the true data. The standard deviation of measurement noise is now specified to be 2% of the corresponding pressure/rate data in all cases. Well SWPs and WPRs are observed every 2 months at all wells. This results in 902 observed data points for a 2400-day history-match period. It is important to emphasize that neither of the ‘true’ models is included in the N_r prior models used in the DSI procedure. Due to the limited number (398) of prior models available, the rejection sampling procedure

Table 2 Set of uncertain parameters considered for six different stochastic DFM realizations. These various combinations define the ensemble of prior DFM models for the realistic field case

Parameter	Low	Medium	High
Fracture permeability	× 0.2	× 1	× 10
Fracture pore volume	× 0.7	× 1	× 0.7
Matrix permeability	× 0.05	× 1	× 20
Matrix water imbibition level	Low	Mid	High

Fig. 8 Simulated production results from prior models for realistic field case. *Green and blue curves* represent forecasts corresponding to different geological concepts



described in Section 4.2 will not be applied here. Therefore, the objective function shown in Eq. 11, instead of that in Eq. 15, is used to generate the DSI forecasts.

5.1 DSI results for ‘true’ model 3

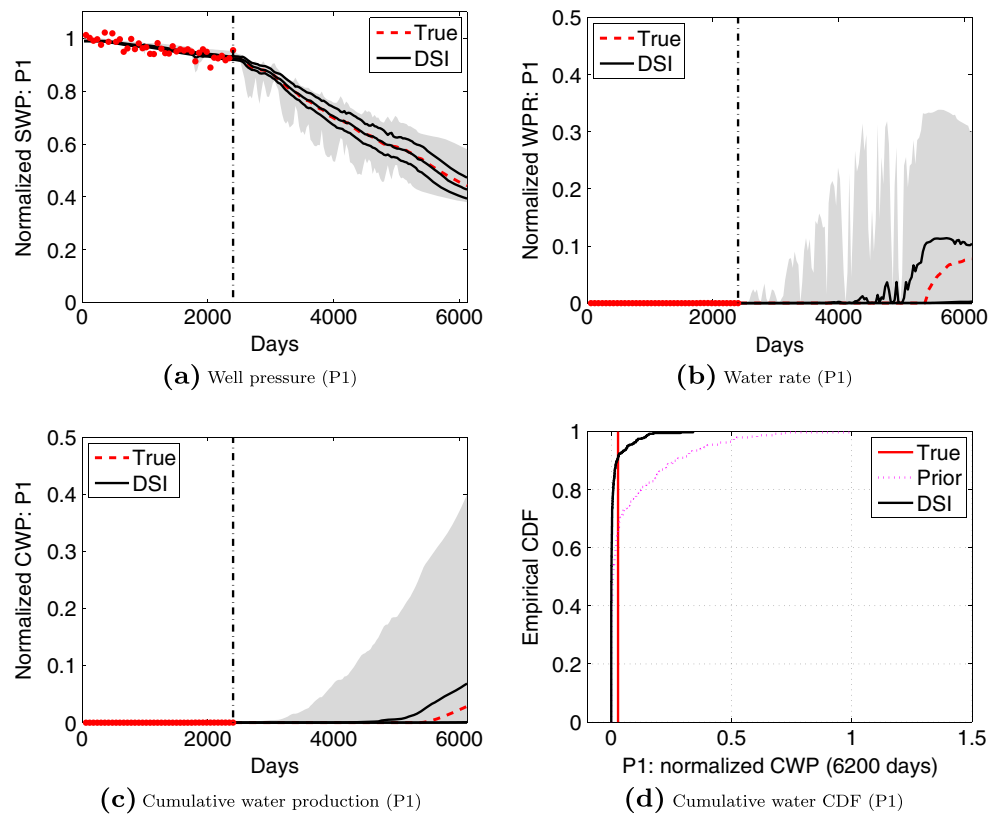
‘True’ model 3 is selected from the ‘green’ set of models. These models generally lead to less water production than models from the ‘blue’ set (as is evident in Fig. 8), though ‘true’ model 3 was specifically chosen because it leads to non-negligible WPRs. This case should therefore be somewhat challenging for the DSI approach since many of the forecasts within the ‘green’ scenario are far from the true results. Figure 9a and b presents DSI results for well P1. The red points again correspond to the observed data (these points lie to the left of the vertical dot-dash line at 2400 days, which indicates the history-match period), and the dashed red curves represent the forecast from the ‘true’ model. The gray-shaded regions represent the P5 to P95 forecast range from the prior models (note that in Section 4 we considered P10 to P90 ranges). The solid black curves represent the P5, P50, and P95 results obtained using DSI. In Fig. 9b and c, only one curve (P95) is evident, since the P5 and P50 curves in these figures correspond to zero water production.

For both SWP and WPR, the DSI results in Fig. 9a and b display narrower uncertainty ranges (P5–P95) compared

with those from the prior models. It is also apparent that the true forecasts (dashed red curves) consistently lie within the P5–P95 range of the DSI predictions. Because WPR displays large variations due to the frequent well shut-ins, for improved clarity, we present forecasts of cumulative water production (CWP) in Fig. 9c. The cumulative distribution function (CDF) of CWP at the end of the forecasting period (6200 days) is shown in Fig. 9d. Substantial reduction in forecast uncertainty through use of the DSI procedure is evident in both plots.

Forecasts for well P3 are shown in Fig. 10. The SWP at P3 displays larger variations in the forecasting period than were observed for P1. Nonetheless, the DSI procedure still narrows the range of uncertainty considerably compared with that from the prior models. In addition, the true forecasts consistently lie within the P5–P95 DSI interval. CWP forecasts (and the degree of uncertainty reduction) for well P3 (Fig. 10b and c) are generally similar to those for P1, and the true data are again captured within the P5–P95 range. It is noteworthy that even though neither well experiences water breakthrough during the history-match period, forecast uncertainty in CWP (and WPR) is still reduced substantially after assimilating the observed data. This demonstrates that the DSI approach captures correlations between data of different types (SWP and WPR in this case) and at different wells.

Fig. 9 Statistics of production forecasts for well P1 from prior models and DSI. Red points and red dashed curves indicate observed data and predictions from ‘true’ model 3. Vertical dot-dash line defines history-match period. Gray-shaded regions represent the P5 to P95 range of predictions from prior models. Black curves denote the P5, P50, and P95 results from DSI



5.2 DSI results for ‘true’ model 4

We now present results for ‘true’ model 4. This model was selected from the ‘blue’ set of models in Fig. 8, which correspond to higher WPRs. DSI results for ‘true’ model 4 are shown in Fig. 11. The SWP forecasts for the two wells (Fig. 11a and d) are comparable to those for ‘true’ model 3, even though the fluctuations are greater here. The CWP results (Fig. 11b and e) are quite interesting in this case. For both wells, the P5–P95 uncertainty intervals are actually expanded, rather than reduced as we might expect, after application of the DSI procedure. However, the true data

now consistently fall within the P5–P95 range, while they lie slightly above the P95 prior from around 3000 days to 5000 days in Fig. 11b. This lack of uncertainty reduction is also evident in the CWP CDFs shown in Fig. 11c and f.

Water breakthrough does not occur over the 2400-day history-matching period in Fig. 11. To demonstrate the uncertainty reduction that can be achieved once water breakthrough is observed, we now extend the history-matching period to 3000 days, at which time water is produced in well P3. We now observe significant uncertainty reduction relative to the prior forecasts for CWP in both wells (Fig. 12b and e), with the true data again falling within

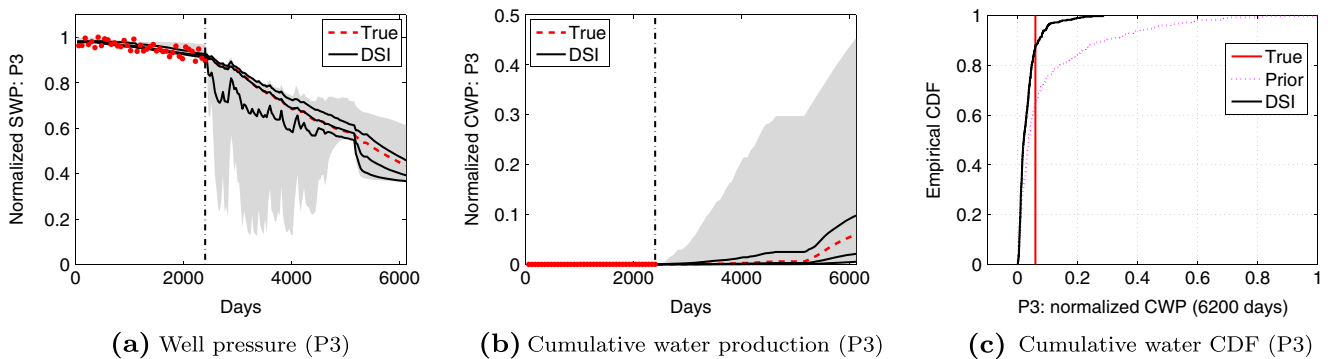


Fig. 10 Statistics of production forecasts for well P3 from prior models and DSI for ‘true’ model 3. Lines, shadings, and colors have the same meanings as in Fig. 9

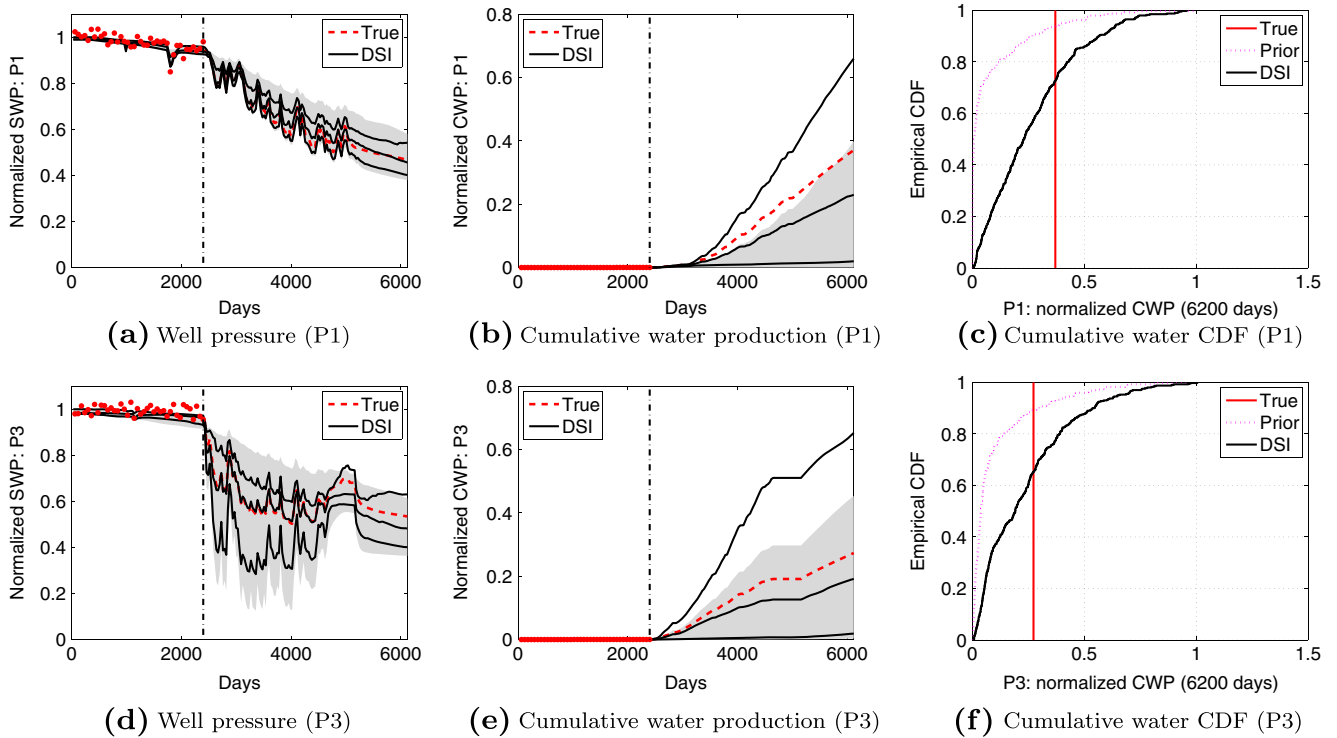


Fig. 11 Statistics of production forecasts, for wells P1 and P3, from prior models and DSI for ‘true’ model 4. Lines, shadings, and colors have the same meanings as in Fig. 9

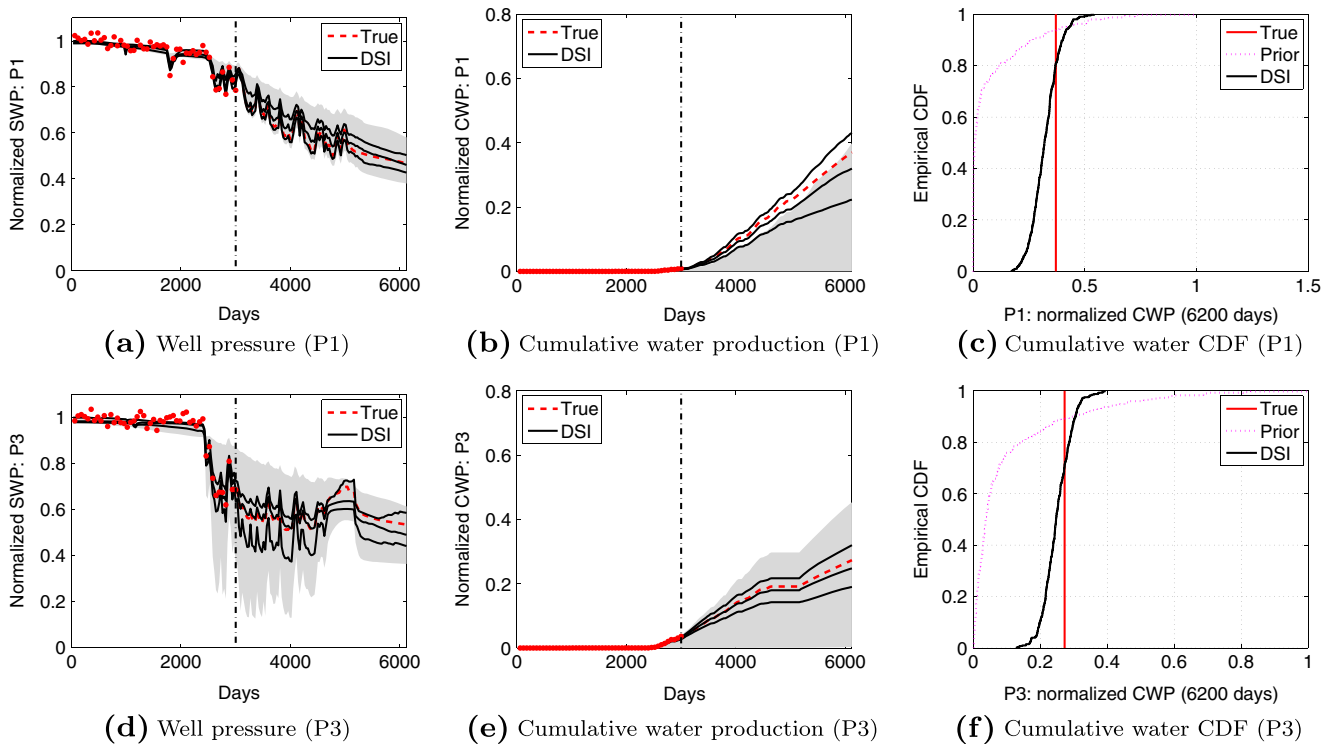


Fig. 12 Statistics of production forecasts, for wells P1 and P3, from prior models and DSI for ‘true’ model 4. History-match period here is 3000 days. Lines, shadings, and colors have the same meanings as in Fig. 9

the P5–P95 DSI interval. This high degree of uncertainty reduction is also apparent in the CWP CDFs in Fig. 12c and f, as well as in the SWP forecasts, especially for well P3 (Fig. 12d).

It is noteworthy that there are only 79 prior models in the ‘blue’ set of models, from which ‘true’ model 4 was selected. The other 319 prior models used by DSI provide quite different forecasts (e.g., low WPRs, as seen in Fig. 8). Nonetheless, the DSI procedure is still able to provide reasonable uncertainty quantification results for this case. This again demonstrates that DSI can be applied with prior models that derive from different geological scenarios/concepts. This is in contrast to many existing history-matching approaches, such as the PCA-based methods described in [32, 33], which require one scenario (i.e., training image) at a time to be considered. It is also important to emphasize that no additional simulation runs were performed when more data were included in the inversion. Again, this is in contrast to most existing approaches, which require additional simulations when new data are introduced.

6 Concluding remarks

In this paper, we extended the data-space inversion (DSI) procedure developed by Sun and Durlafsky [30] to enable its application to a realistic naturally fractured carbonate reservoir. In that case, due to the complex well controls (frequent shut-ins) employed, the production responses from the prior models do not display clear stages/patterns. This motivated the development of a new data reparameterization technique that entails the use of PCA along with histogram transformation. These treatments were implemented into a data-space RML procedure in order to generate multiple production forecasts.

We first applied the new DSI treatments to a synthetic two-dimensional DFM system. This case involved realizations drawn from a wide range of geological scenarios characterized by different fracture densities, orientations and apertures, and different matrix properties. Two different ‘true’ models were considered, and DSI predictions were compared with rejection sampling results (using the mean of the weighted data mismatch) for both cases. Acceptable agreement between DSI and RS was achieved, suggesting that DSI can be used for cases that include a variety of geological scenarios. We then applied the DSI procedure to a realistic three-dimensional naturally fractured system. Two different ‘true’ models, which corresponded to different fracture scenarios, were again considered. In both cases, we observed reasonable uncertainty quantification results using the DSI method, even when the forecasts from the ‘true’ model were outside the P5 to P95 range of

predictions from prior models. The impact of varying amounts of historical production data on uncertainty reduction was also demonstrated for one of the cases.

The DSI procedure has the advantage of requiring flow simulations only for prior models, and these can all be run in parallel. Once these simulations are performed, the data-space RML procedure can be run in a matter of minutes. Another advantage of DSI is that additional simulation runs need not be performed if new data become available. In addition, the impact of quantities such as measurement error (appearing in C_D), which may also be used to approximate model error, can be efficiently studied. In this case, only data-space computations, not additional flow simulations, are required if different C_D matrices are considered. And finally, DSI is well-suited for studies in which different realizations are represented on different grids, as is typically the case with DFM models (many traditional data-assimilation approaches are limited in this regard).

The significant limitation with DSI, however, is the fact that it does not provide posterior (history-matched) geological models. Such models are of course required in some applications. Other limitations of the DSI methodology are discussed in [30]. These include the need for a sufficiently broad set of prior-model scenarios, as were considered in Section 4 for the synthetic case. The DSI procedure in its current form cannot be used for cases in which future well locations or settings are not known when the DSI model is constructed. This means that the method is not yet suitable for use in optimizing well controls (e.g., time-varying BHPs) or the placement of new wells. Finally, the procedure may require the simulation of a large number of prior realizations in problems involving complicated data responses and/or a large amount of data.

In future work, we plan to extend DSI treatments to address several of these issues. For practical cases with many wells and large data volumes, it may be beneficial to decompose the reservoir into multiple regions and to then apply DSI region by region. It will also be useful to establish linkages between DSI and traditional model-based inversion methods. This may allow us to identify a small number of realizations, and associated weightings, that provide flow responses in agreement with those from DSI. This could facilitate the application of DSI for optimization. We also plan to consider the use of DSI for the design of a reservoir surveillance system (as in [8]) or to investigate the value of information. For both problems, the impact on uncertainty reduction of different data types, and data measured at various locations, can be quantified very efficiently using DSI.

Acknowledgments We thank Chevron ETC and the Stanford Smart Fields Consortium for financial support.

References

- Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., Vallès, B.: The ensemble Kalman filter in reservoir engineering—a review. *SPE J.* **14**(3), 393–412 (2009)
- Chen, Y., Oliver, D.S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci.* **44**(1), 1–26 (2012)
- Cherpeau, N., Caumon, G., Caers, J., Lévy, B.: Method for stochastic inverse modeling of fault geometry and connectivity using flow data. *Math. Geosci.* **44**(2), 147–168 (2012)
- Dehghani, K., Fischer, D., Skalinski, M.: Application of integrated reservoir studies and techniques to estimate oil volumes and recovery—Tengiz field, Republic of Kazakhstan. *SPE Reserv. Eval. Eng.* **11**(2), 362–378 (2008)
- Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**(4), 343–367 (2003)
- Evensen, G., van Leeuwen, P.J.: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.* **128**(6), 1852–1867 (2000)
- Geiger, S., Matthäi, S.K., Niessner, J., Helmig, R.: Black-oil simulations for three-component, three-phase flow in fractured porous media. *SPE J.* **14**(2), 338–354 (2009)
- He, J., Xie, J., Sarma, P., Wen, X.H., Chen, W.H., Kamath, J.: Model-based a priori evaluation of surveillance programs effectiveness using proxies. Paper SPE 173229 presented at the SPE Reservoir Simulation Symposium, Houston, Texas, USA 23–25 February (2015)
- Hoteit, H., Firoozabadi, A.: Multicomponent fluid flow by discontinuous Galerkin and mixed methods in unfractured and fractured media. *Water Resour. Res.* **41**(11) (2015). doi:[10.1029/2005WR004339](https://doi.org/10.1029/2005WR004339)
- Hu, L., Jenni, S.: History matching of object-based stochastic reservoir models. *SPE J.* **10**(3), 312–323 (2005)
- Hui, M.H., Heidary-Fyrozjaee, M., Kamath, J.: Scaling gravity-drainage oil recovery from fractured reservoirs using 3D gravity-drainage scaling relationships. Paper SPE 172295-MS, presented at the SPE Annual Caspian Technical Conference and Exhibition, Astana, Kazakhstan 12–14 November (2014)
- Hui, M.H., Kamath, J., Narr, W., Gong, B., Fitzmorris, R.E.: Realistic modeling of fracture networks in a giant carbonate reservoir. Paper IPTC 11386-MS, presented at the International Petroleum Technology Conference, Dubai, United Arab Emirates 4–6 December (2007)
- Hui, M.H., Mallison, B., Heidary-Fyrozjaee, M., Narr, W.: The upscaling of discrete fracture models for faster, coarse-scale simulations of IOR and EOR processes for fractured reservoirs. Paper SPE 166075-MS, presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, USA 30 September–2 October (2013)
- Jenni, S., Hu, L., Basquet, R., De Marsily, G., Bourbiaux, B.: History matching of a stochastic model of field-scale fractures: Methodology and case study. *Oil Gas Sci. Technol.* **62**(2), 265–276 (2007)
- Karimi-Fard, M., Durlofsky, L.J.: A general gridding, discretization, and coarsening methodology for modeling flow in porous formations with discrete geological features. *Adv. Water Resour.* **96**, 354–372 (2016)
- Karimi-Fard, M., Durlofsky, L.J., Aziz, K.: An efficient discrete fracture model applicable for general purpose reservoir simulators. *SPE J.* **9**(2), 227–236 (2004)
- King, G.R., Jones, M., Tankersley, T., Flodin, E., Jenkins, S., Zhumagulova, A., Eaton, W., Bateman, P., Laidlaw, C., Fitzmorris, R., Ma, X., Dagistanova, K.: Use of brown-field experimental design methods for post-processing conventional history match results. Paper SPE 159341-MS, presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA 8–10 October (2012)
- Kitanidis, P.K.: Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resour. Res.* **22**(4), 499–507 (1986)
- Krishnamurti, T.N., Kishtawal, C., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S., Surendran, S.: Multimodel ensemble forecasts for weather and seasonal climate. *J. Clim.* **13**(23), 4196–4216 (2000)
- Mallet, V., Stoltz, G., Mauricette, B.: Ozone ensemble forecast with machine learning algorithms. *J. Geophys. Res.* **114**(D5), 148–227 (2009)
- Matthäi, S., Mezentsev, A., Belayneh, M.: Control-volume finite-element two-phase flow experiments with fractured rock represented by unstructured 3D hybrid meshes. Paper SPE 93341-MS, presented at the SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA 31 January–2 February (2005)
- Mosegaard, K., Tarantola, A.: Monte Carlo sampling of solutions to inverse problems. *J. Geophys. Res.* **100**(B7), 12,431–12,447 (1995)
- Oliver, D.S.: On conditional simulation to inaccurate data. *Math. Geosci.* **28**(6), 811–817 (1996)
- Oliver, D.S., Chen, Y.: Recent progress on reservoir history matching: a review. *Computat. Geosci.* **15**(1), 185–221 (2011)
- Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse theory for petroleum reservoir characterization and history matching. Cambridge University Press, Cambridge (2008)
- Reynolds, A.C., He, N., Oliver, D.S.: Reducing Uncertainty in Geostatistical Description with Well-Testing Pressure Data. In: Reservoir Characterization—Recent Advances, pp. 149–162. American Association of Petroleum Geologists, Tulsa (1999)
- Satija, A., Caers, J.: Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space. *Adv. Water Resour.* **77**, 69–81 (2015)
- Scheidt, C., Renard, P., Caers, J.: Prediction-focused subsurface modeling: investigating the need for accuracy in flow-based inverse modeling. *Math. Geosci.* **47**(2), 173–191 (2015)
- Sun, W.: Data Driven History Matching for Reservoir Production Forecasting. Master's thesis, Stanford University (2014)
- Sun, W., Durlofsky, L.J.: A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. *Math Geosci.* (2017). doi:[10.1007/s11004-016-9672-8](https://doi.org/10.1007/s11004-016-9672-8)
- Tarantola, A.: Inverse problem theory and methods for model parameter estimation. SIAM (2005)
- Vo, H.X., Durlofsky, L.J.: A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models. *Math. Geosci.* **46**(7), 775–813 (2014)
- Vo, H.X., Durlofsky, L.J.: Data assimilation and uncertainty assessment for complex geological models using a new PCA-based parameterization. *Computat. Geosci.* **19**(4), 747–767 (2015)
- Wen, X.H., Chen, W.H.: Real-time reservoir model updating using ensemble Kalman filter with confirming option. *SPE J.* **11**(4), 431–442 (2006)