ORIGINAL PAPER

Evaluation of Gaussian approximations for data assimilation in reservoir models

Marco A. Iglesias · Kody J. H. Law · Andrew M. Stuart

Abstract The Bayesian framework is the standard approach for data assimilation in reservoir modeling. This framework involves characterizing the posterior distribution of geological parameters in terms of a given prior distribution and data from the reservoir dynamics, together with a forward model connecting the space of geological parameters to the data space. Since the posterior distribution quantifies the uncertainty in the geologic parameters of the reservoir, the characterization of the posterior is fundamental for the optimal management of reservoirs. Unfortunately, due to the large-scale highly nonlinear properties of standard reservoir models, characterizing the posterior is computationally prohibitive. Instead, more affordable ad hoc techniques, based on Gaussian approximations, are often used for characterizing the posterior distribution. Evaluating the performance of those Gaussian approximations is typically conducted by assessing their ability at reproducing the truth within the confidence interval provided by the ad hoc technique under consideration. This has the disadvantage of mixing up the approximation properties of the history matching algorithm employed with the information content of the particular observations used, making it hard to evaluate the effect of the ad hoc approximations alone. In this paper, we avoid this disadvantage by comparing the ad hoc techniques with a fully resolved state-of-the-art probing of the Bayesian posterior distribution. The ad hoc techniques whose performance we assess are based on (1) linearization around the maximum a posteriori estimate, (2) randomized maximum likelihood, and (3) ensemble Kalman filter-type methods. In order to fully resolve the posterior distribution, we implement a state-of-the art Markov chain Monte Carlo (MCMC) method that scales well with respect to the dimension of the parameter space, enabling us to study realistic forward models, in two space dimensions, at a high level of grid refinement. Our implementation of the MCMC method provides the gold standard against which the aforementioned Gaussian approximations are assessed. We present numerical synthetic experiments where we quantify the capability of each of the ad hoc Gaussian approximation in reproducing the mean and the variance of the posterior distribution (characterized via MCMC) associated to a data assimilation problem. Both single-phase and two-phase (oil–water) reservoir models are considered so that fundamental differences in the resulting forward operators are highlighted. The main objective of our controlled experiments was to exhibit the substantial discrepancies of the approximation properties of standard ad hoc Gaussian approximations. Numerical investigations of the type we present here will lead to the greater understanding of the cost-efficient, but ad hoc, Bayesian techniques used for data assimilation in petroleum reservoirs and hence ultimately to improved techniques with more accurate uncertainty quantification.

A. M. Stuart · M. A. Iglesias (✉)
University of Warwick, Coventry, UK
e-mail: m.a.iglesias-hernandez@warwick.ac.uk

A. M. Stuart
e-mail: A.M.Stuart@warwick.ac.uk

K. J. H. Law
CEMSE, King Abdullah University of Science and Technology,
Thuwal, Saudi Arabia
e-mail: K.J.H.Law@warwick.ac.uk

1 Introduction

Simulating the dynamics of a reservoir involves solving a large-scale numerical model that depends on parameters related to petrophysical properties of the reservoir. These properties need to be known at each discretization point of the physical domain of the reservoir. Unfortunately, direct measurements of petrophysical properties are only available at a small number of locations within the domain of interest. Therefore, a statistical description of the reservoir parameters is required to properly account for the uncertainty in the petrophsyical properties caused by the lack of information. A *prior distribution* of geologically consistent reservoir parameters can be generated, for example, from a variogram analysis conducted on static data from core samples. Geostatistical techniques can then be used to generate realizations of reservoir parameters conditioned to static data [11]. In some cases, information concerning geologic facies may also be incorporated [3]. On the other hand, with the aid of downhole permanent sensors, measurements of reservoir flow can be continuously acquired. In a Bayesian framework, the prior distribution of the petrophysical properties is combined with these flow measurements, using a forward reservoir model linking the parameter and data spaces, to characterize the posterior distribution of the reservoir parameters given the dynamic (flow) data. This posterior distribution quantifies the uncertainty in the reservoir predictions, and it is essential for assessing the economical and environmental risk of oil recovery procedures.

Markov chain Monte Carlo (MCMC) methods are the standard techniques for sampling the posterior distribution described above. In particular, the Metropolis–Hasting variant of MCMC has been typically used for data assimilation in reservoir models [4, 13, 14, 16, 23, 25, 28]. In general, the posterior distribution that arises from Bayesian data assimilation does not admit a finite-dimensional parametrization, with the exception of very few particular cases such as the linear and Gaussian case. Therefore, strictly speaking, an infinite number of samples are required to define it. This implies, in practice, that hundreds of thousands or even millions of reservoir simulations may be required for standard MCMC methods to accurately characterize the posterior distribution. This computational disadvantage of standard MCMC approaches has given rise to the development of more computationally efficient MCMC techniques [9, 13, 16, 25]. With the increasing advancement of computational power, the aforementioned MCMC methods may potentially become viable tools for reservoir management in the decades to come. At the present time, however, it is essential to develop techniques that provide a reasonable characterization of the posterior with a computational cost that involves a limited number of reservoir model runs. It then comes as no surprise that, in recent years, research on data assimilation for uncertainty quantification applications has mainly focused on improving the efficiency and accuracy of ad hoc ensemble-based techniques that provide approximations of the posterior distribution and are typically based on invocation of a Gaussian assumption at some point in their derivation.

As pointed out in the recent literature review of [27], there are three main approaches that have been consistently adopted for sampling approximations of the posterior distribution: (1) linearization around the maximum a posteriori (MAP) estimate (linearization around the MAP, LMAP), (2) randomized maximum likelihood (RML) method, and (3) ensemble Kalman filter (EnKF). Under a Gaussian prior and a linear model, it can be shown that all these techniques provide samples of the posterior distribution [22, 26]. We therefore refer to those methods as *Gaussian approximations* of the posterior. For standard (nonlinear) reservoir models, the mathematical structure of the approximation provided by the three approaches is still unknown. Nonetheless, the aforementioned methods are widely applied for generating model parameters conditioned to dynamics data, which are then used for statistical analysis of reservoir performance. Consequently, in the Bayesian framework, optimal decision making and risk management depend on the quality of the underlying Gaussian approximations of the posterior. It is therefore fundamental to understand the accuracy and convergence properties of those Gaussian approximations in order to interpret predictions of uncertainty made using them and in order to develop improved methodologies from them. Rigorous numerical studies of these Gaussian approximate algorithms can shed light on these issues and can point us towards theoretical analyses. In this paper, we therefore provide a numerical evaluation of the performance of the ad hoc Gaussian approximate algorithms LMAP and RML and the variants of the EnKF methodology, by using an expensive, but fully resolved, MCMC simulation as our gold standard. This approach is analogous to the recent study of similar Gaussian approximate algorithms arising in the context of atmospheric data assimilation [21].

1.1 Literature review

Although the theoretical aspects of the approximation properties of LMAP, RML, and EnKF are unknown for the case of nonlinear models, some numerical investigations have been performed [4, 23]. To the best of our knowledge, only the work in [23] has the purpose of assessing approximate evaluation of approximate methods for sampling the posterior. To accomplish this goal, Liu and Oliver [23] used a standard random walk MCMC method to generate accurate samples of the posterior. These, in turn, are used as the gold standard against which the performance of the approximate methods, LMAP and RML, and of the pilot point

methods is compared. In their evaluation, Liu and Oliver use synthetic data from a single-phase one-dimensional reservoir discretized with 20 gridblocks. The main conclusion of [23] is that RML provides the best uncertainty quantification when compared against the MCMC gold standard. In particular, RML outperforms pilot point methods whose application to reservoir data assimilation problems has been lately abandoned.

Within the context of evaluating the uncertainty quantification properties of data assimilation techniques, it is relevant to mention the work of [4] where several methods were compared for the synthetic PUNQ-S3 reservoir model. The main goal of [4], however, is to evaluate the ability of the corresponding techniques to provide confidence interval that contains the truth estimate. Among those techniques, MCMC is the only one that can potentially provide accurate samples from the Bayesian posterior. Unfortunately, as stated in [4], the MCMC results are not conclusive due to the small chain used for their experiments. Thus, the evaluation in [4] does not provide an evaluation in the strict sense of a Bayesian framework. It is worth mentioning that the work of [4] and [23] appeared almost a decade ago when EnKF had just been introduced to the history matching community [1], and so EnKF was not assessed in [4, 23]. Almost a decade later and after hundreds of publications, EnKF-based methods are now perhaps the only computationally feasible technique for real-time data assimilation in petroleum reservoirs. For a comprehensive review of EnKF for reservoir applications, we refer the reader to [1].

In the recent work of [16], the EnKF is combined with an MCMC algorithm to improve the efficiency of standard MCMC methods for sampling the posterior in a Bayesian framework. Although the analysis of the approximate properties of EnKF is not the main goal of [16], an implicit evaluation of EnKF is displayed. Indeed, under the assumption that the MCMC samples of [16] provide an accurate characterization of the posterior, then [16] provides a partial assessment of EnKF for approximating the posterior. It is also worth mentioning the work of [17] where the history matching and uncertainty quantification properties of several variants of Kalman-based methods are compared with respect to the gold standard generated with an MCMC approach similar to the one used in [23]. It is important to note that the experiments [17] are conducted on a rather small problem (30, one-dimensional), with a limited amount of observations (12), for which the standard MCMC approach used in [17] can be computed. While the aforementioned work reveals advantages and limitations of EnKF for sampling the posterior, an evaluation in the setup of the more stringent model problems that we consider in this paper, which are of high dimension, would not be possible using the standard MCMC techniques used in [17, 23]

due to their computational limitations. Our mesh-invariant MCMC methods, however, *will* allow for the study to be undertaken under these more stringent conditions.

In the context of evaluating the uncertainty quantification properties of Gaussian approximations of the posterior, we highlight the work of [18, 34]. For the PUNQ-S3 model mentioned above, [18] compares the performance of EnKF and RML. In [34], a new singular value decomposition (SVD)-based RML is introduced and compared against EnKF. It is important to mention, however, that [18, 34] evaluate the capability of these Gaussian approximations for reproducing the truth within the confidence interval of the technique under consideration. While this is a natural strategy for assessing uncertainty quantification properties, it is an insufficient evaluation from the perspective of the Bayesian framework. In other words, capturing the truth within the spread of model predictions obtained with a Gaussian approximation does not ensure that the spread correctly represents the uncertainty quantified by the posterior distribution of Bayesian data assimilation. It is therefore essential to develop a controlled experiment where standard Gaussian approximation can be tested against the solution to the Bayesian data assimilation problem: the posterior distribution.

1.2 The proposed work

In this paper, we propose the numerical evaluation of LMAP, RML, and some standard versions of the ensemble Kalman filter-type methods for approximating the posterior distribution within the Bayesian framework of data assimilation. We characterize the posterior distribution by using a state-of-the art MCMC method that provides the gold standard against which the aforementioned Gaussian approximations is compared. In this sense, our work has a similar goal to the one of [23]. However, there are two recent algorithmic developments which motivate our desire to revisit the perspective introduced in [23]. The first, discussed in detail below, is that MCMC methodology has evolved significantly, enabling the study of considerably more sophisticated forward models and more highly dimensional parameterizations of the unknown petrophysical quantities, leading to greater realism. The second new aspect of our work, in comparison with [23], is the assessment of ensemble Kalman filter-type methods. In particular, we consider the most standard EnKF implementations, namely, (1) the perturbed observation EnKF and (2) the square root filter (ensemble square root filter, EnSRF) of [14, 30]. In both cases, we also evaluate the effect of performing distance-based localization [6, 15].

We emphasize that, in contrast to other approaches [4, 6, 15, 18, 34] where the aim is to recover the truth within the confidence interval of relevant quantities and to

history match data, here we are interested in assessing the performance for characterizing the posterior distribution. Evaluation of algorithms by their ability to recover the truth within a confidence interval has the disadvantage of entangling the approximation properties of the history matching algorithm employed with the information content of the particular observations used, making it hard to evaluate the effect of the ad hoc approximations alone. Our assessment of the ability to probe the Bayesian posterior distribution is conducted by quantifying the capability of each Gaussian approximation in reproducing the mean and the variance of the posterior distribution associated to a data assimilation problem. Two prototypical reservoir models are used, both in two spatial dimensions: (1) slightly compressible single-phase Darcy flow model and (2) incompressible oil–water reservoir model. In both models, the unknown is the logarithm of the absolute permeability of the reservoir $u = \log K$. For the single-phase model, pressure data are collected from production wells. For the oil–water model, total flow rate is measured at the production wells while bottom-hole pressure is collected at the injection wells. For both models considered here, the corresponding parameter-to-output map $G(u)$ is nonlinear. Thus, even when the prior distribution is Gaussian, the Bayesian posterior is non-Gaussian. This constitutes the ideal scenario to evaluate approximation properties of the techniques of interest, provided that a gold standard is obtained from accurately sampling the posterior as we described below.

As we indicated earlier, some MCMC methods have been used for sampling the Bayesian posterior distribution in reservoir models. However, some of these methods [13, 25] rely on reducing the parameter space (e.g., via truncating Karhunen–Loève expansion) and/or upscaling the model to reduce the computational cost of the algorithm. In the present experiment, however, we are interested in the more general case where no reduction of the parameter space is possible. In other words, we assume that the petrophysical property is unknown at each location of the physical domain of the reservoir. In this case, a standard MCMC technique like the one used in [23] is computationally prohibitive for larger size problems like the ones considered here. To overcome this difficulty, we take advantage of recent developments in MCMC methodology and sample the posterior by applying the preconditioned Crank–Nicolson MCMC (pCN-MCMC) method described in [9] and derived from the infinite-dimensional Bayesian framework developed in [33]. In contrast to standard MCMC methods, the acceptance probability in the pCN-MCMC method is invariant with respect to the dimension of the parameter space, therefore making pCN-MCMC ideal for large-scale problems like the one studied here. The advantage of using pCN-MCMC over standard MCMC for data assimilation in some geophysical problems has been shown in [9]. For petroleum

reservoir applications, the computational efficiency of pCN-MCMC with respect to other existing methods deserves further investigation. Nevertheless, in the present work, we apply pCN-MCMC and provide numerical evidence of convergence so that the corresponding realizations generated with pCN-MCMC are samples from the Bayesian posterior. These, in turn, provide a gold standard against which LMAP, RML, and standard versions of ensemble Kalman filter-type of methods are compared. The proposed numerical evaluation of the approximation techniques has two concrete goals: (1) assess the capability to recover the mean and variance of the posterior on the geological properties of interest and (2) evaluate the performance for reproducing the uncertainty (quantified by the posterior) in the reservoir model predictions.

It is important to mention the recent literature on other ad hoc ensemble methods for history matching and uncertainty quantification in petroleum reservoirs [7, 12, 22, 31, 32]. These techniques have been shown to outperform standard ensemble methods in the sense that they produce an ensemble of geologic parameters whose corresponding model predictions capture the one from the truth (or reference) parameter within the ensemble spread. However, the uncertainty quantification properties of those techniques with respect to characterizing the posterior distributions are still an open problem. Assessing these new techniques is beyond the scope of the present paper. Nonetheless, we remark that the aforementioned techniques are either EnKF-based or RML-based techniques for which the proposed approach of benchmarking and assessing the posterior distribution can be extended. Undertaking such extensions would form an interesting research direction, building on the work herein, but is beyond the scope of this paper itself.

In Section 2, we describe the prototypical reservoir models that define the forward operators that we use for data assimilation. The Bayesian framework for data assimilation as well as the MCMC methodology for sampling the posterior is introduced in Section 3. Methodologies based on Gaussian approximations of the posterior are described in Section 4. In Section 5, we report and discuss the numerical results and comparisons of our synthetic experiments. The summary and final remarks are presented in Section 6.

2 Forward reservoir models

In this section, we briefly outline the forward (reservoir) models that we use for the evaluation of Gaussian approximations of the posterior. On the one hand, we consider simplified two-dimensional models for which a forward model run is computationally inexpensive and therefore feasible for the highly computationally challenging MCMC

method. On the other hand, by sharing the mathematical structure of more sophisticated models, the models we describe below are ideal for prototyping and evaluating performance in a controlled fashion. For each of the following models, we consider a two-dimensional reservoir whose physical domain, absolute permeability, and porosity are denoted by D, K, and ϕ, respectively. The interval $[0,T]$ $(T > 0)$ is the time interval of interest for the flow simulation. For each reservoir model, we define the forward operator $G:X \rightarrow \mathbb{R}^N$ that maps the parameter space X into the observation space $\mathbb{R}^N$. In other words, $G(u)$ is the model predictions corresponding to the parameter $u \in X$. For simplicity, we assume that the only unknown parameter is $u = \log K$. Nevertheless, all the techniques and implementations that we describe in subsequent sections can be extended to include additional parameters, such as porosity which is routinely estimated alongside permeability in many practical scenarios.

2.1 Single-phase Darcy flow

We consider a single-phase reservoir where oil is produced at N_w production wells operated under prescribed production rates $\{q^l(t)\}_{l=1}^{N_w}$ ($t \in [0, T]$). The flow in the reservoir is described in terms of the (state variable) fluid pressure $p(x, t)$ ($(x, t) \in D \times [0, T]$) which is governed by the following equation [8]:

$$c\phi \frac{\partial p}{\partial t} - \nabla \cdot \nu^{-1} e^u \nabla p = \sum_{l=1}^{N_w} q^l \delta(x - x^l) \text{ in } D \times (0, T] \tag{1}$$

where $u \equiv \log K$, c is the total compressibility, ν is the viscosity, and $\delta(x - x^l)$ is (a possibly mollified) Dirac delta centered at the lth well with location denoted by x^l. In addition, we consider the initial condition $p = p_0$ in $D \times \{0\}$ and no-flow boundary conditions. As we indicated earlier, the only uncertain parameter in Eq. (1) is the log permeability u. Therefore, the additional model parameters c, ϕ, ν, and p_0 and the geometry D in Eq. (1) are prescribed.

In order to construct the forward operator, we first define the model predictions of measurements. Let us then assume that N_M measurements of pressure from wells are collected at times $t_1, \ldots, t_M$. We define the measurement functional as

$$M_n^l(p) = p(x^l, t_n) \tag{2}$$

that corresponds to the fluid pressure at time t_n and well location x^l. For the exposition of subsequent sections, we also define the vector

$$M_n(p) = (M_n^1(p), \ldots, M_n^{N_w}(p)). \tag{3}$$

We finally define $N = N_w N_M$, i.e., the total number of observations from wells, and construct the forward operator

$$G(u) = (M_1(p), \ldots, M_{N_M}(p)). \tag{4}$$

Note that p in Eq. (4) depends on u via Eq. (1).

2.2 Oil–water reservoir model

We consider an oil–water reservoir model initially saturated with oil and irreducible water. Let us index by $\gamma = w$ and $\gamma = o$ the water and oil phase, respectively. We assume that both fluids and the rock are incompressible. We are interested in a waterflood process where water is injected at N_I injection wells located at $\{x_I^l\}_{l=1}^{N_I}$. Water and oil are produced at N_P production wells located at $\{x_P^l\}_{l=1}^{N_P}$. Additionally, we assume that injection wells are operated under prescribed rates $\{q^l(t)\}_{l=1}^{N_I}$ while production wells are constrained to bottom-hole pressure denoted by $\{P_{bh}^l(t)\}_{l=1}^{N_P}$. The reservoir dynamics in $[0, T]$ are described by the (state variables) water saturation and the pressure, which are denoted by $s(x, t)$ and $p(x, t)$, respectively ($(x, t) \in D \times [0, T]$). From standard arguments, it can be shown that (s, p) is the solution to the following system [8]:

$$-\nabla \cdot \lambda(s) e^u \nabla p = \sum_{l=1}^{N_I} q^l \delta\left(x - x_I^l\right)$$
$$+ \sum_{l=1}^{N_w} \omega^l \lambda(s)[P_{bh}^l - p]\delta\left(x - x_P^l\right), \tag{5}$$

$$\phi \frac{\partial s}{\partial t} - \nabla \cdot \lambda_w(s) e^u \nabla p = \sum_{l=1}^{N_I} q^l \delta\left(x - x_I^l\right)$$
$$+ \sum_{l=1}^{N_w} \omega^l \lambda_w(s)\left[P_{bh}^l - p\right] \delta\left(x - x_P^l\right), \tag{6}$$

in $D \times (0, T]$, where $\delta(x - x_P^l)$ and $\delta(x - x_I^l)$ are the (possibly mollified) Dirac deltas as defined before and $\{\omega^l\}_{l=1}^{N_P}$ is the constant related to the well model [8]. Additionally, $\lambda_w(s)$ and $\lambda(s)$ denote the water and total mobility defined by

$$\lambda_w(s) = \frac{k_{rw}(s)}{\nu_w}, \qquad \lambda(s) = \frac{k_{ro}(s)}{\nu_o} + \lambda_w(s) \tag{7}$$

where $k_{r\gamma}(s)$ and ν_γ denote the relative permeability and the viscosity of the γ-phase fluid, respectively. Furthermore, we assume that

$$k_{rw}(s) = a_w \left[\frac{s - s_{iw}}{1 - s_{iw} - s_{or}}\right]^2,$$
$$k_{ro}(s) = a_o \left[\frac{1 - s - s_{or}}{1 - s_{iw} - s_{or}}\right]^2 \tag{8}$$

where a_w, $a_o \in (0, 1]$, s_{iw} is the irreducible water saturation, and s_{or} is the residual oil saturation. We additionally prescribe initial conditions for pressure and water saturation:

$$p = p_0, \qquad s = s_0 \qquad \text{in } D \times \{0\}. \tag{9}$$

For simplicity, no-flow boundary conditions are prescribed on the reservoir boundary. Let us assume that there are N_M measurement times denoted as before $\{t_n\}_{n=1}^{N_M}$. We assume that measurements of bottom-hole pressure are collected at the injection wells at $\{t_n\}_{n=1}^{N_M}$. This, according to Peaceman's well model [8], is defined by

$$M_n^{l,I}(p, s) = P_{bh}^{l,I}(t_n) = \left[\frac{q^l(t_n)}{\omega^l \lambda \left(s \left(x_I^l, t_n \right) \right)} + p \left(x_I^l, t_n \right) \right] \tag{10}$$

for $l = 1, \ldots, N_I$ and $n = 1, \ldots, N_M$. Analogously, we consider measurements of total flow rate at the production wells:

$$M_n^{l,P}(p, s) = q^{l,P}(t_n) = \omega^l \lambda \left(s \left(x_P^l, t_n \right) \right)$$
$$\times \left[P_{bh}^l(t_n) - p \left(x_P^l, t_n \right) \right] \tag{11}$$

for $l = 1, \ldots, N_P$ and $n = 1, \ldots, N_M$. Let us denote by $N_w = N_P + N_I$ the total number of wells and define the N_w-dimensional vector

$$M_n(p, s) = \left(M_n^{1,I}(p, s), \ldots, M_n^{N_I,I}(p, s), \right.$$
$$\left. M_n^{1,P}(p, s), \ldots, M_n^{N_P,P}(p, s) \right). \tag{12}$$

The total number of measurements N is defined as before and the forward map $G : X \to \mathbb{R}^N$ is then given by the expression

$$G(u) = (M_1(p, s), \ldots, M_{N_M}(p, s)) \tag{13}$$

which in this case comprises the production data obtained from production and injection wells at the measurement times.

With both forward models written in terms of the forward operator G, in the next section, we describe the Bayesian inverse problem of finding u, given noisy observations of $G(u)$.

3 The Bayesian framework

We assume that the unknown parameter u and the data $y \in Y$ are related by

$$y = G(u) + \eta \tag{14}$$

where G is the forward map introduced in the previous section and $\eta \in \mathbb{R}^N$ is a vector of random noise. Informally, the Bayesian approach to inversion proceeds by placing a *prior* probability distribution $\mathbb{P}(u)$ on u and assuming an independent probability distribution on η. The *likelihood*,

namely, the probability of the observed data y, given a particular instance of the unknown parameter u, is then denoted $\mathbb{P}(y|u)$. Bayes' rule then states that the *posterior* probability of the unknown parameter u, given the observed data y, denoted by $\mathbb{P}(u|y)$, is determined by the formula

$$\frac{\mathbb{P}(u|y)}{\mathbb{P}(u)} \propto \mathbb{P}(y|u). \tag{15}$$

In this section, we formulate this precisely in the case where the unknown parameter is a function.

We define the norm $|| \cdot ||_B = ||B^{-1/2}(\cdot)||$ for any covariance operator B, and we use this notation throughout the paper, in particular in the observation space, with $B = \Gamma$, and in the log permeability space with $B = C$. For simplicity and following convention in the field, we will not distinguish notationally between the random variable and its realization, except in the case of the truth, which will be important to distinguish by $u^\dagger$ in subsequent sections in which it will be prescribed and known.

3.1 An infinite-dimensional Bayesian framework

When approximating fields in finite dimensions, for example, those arising as input to or solution of a partial differential equation (PDE), the accuracy of the approximation improves as the mesh is refined and the dimension increases; practitioners often want to take advantage of this refinement to improve predictions. However, the benefits resulting from approximation of the forward PDE model are lost if it is subsequently embedded into a method for inverse problems which degenerates under such mesh refinements. Therefore, it is of interest to establish methods for solving inverse problems, whose performance is invariant with respect to increasing dimension of the underlying unknown. Confronting the function space framework head on guarantees this desirable invariance with respect to mesh refinement of the forward model approximation.

We are interested in the inverse problem of characterizing the posterior distribution of the unknown log permeability *function* u, given *finite-dimensional* observational data denoted by y. We approach this inverse problem by means of the infinite-dimensional Bayesian framework of [33] that we now briefly describe. Assume that $u \in X$ where X is a Hilbert space, and denote by μ_0 the prior probability measure on u. We assume that the unknown u and the data $y \in Y$ are related by Eq. (14). For simplicity, we assume that $\eta \sim N(0, \Gamma)$. Then, the rigorous interpretation of Eq. (15) is that the posterior distribution on $u|y$ is given by measure μ satisfying

$$\frac{d\mu}{d\mu_0}(u) = \frac{\exp(-\Phi(u, y))}{\int_X \exp(-\Phi(u, y))\mu_0(du)} \tag{16}$$

where the left-hand side of Eq. (16) is the ratio of infinitesimal probabilities at the point u of the posterior distribution $\mu(u) = \mathbb{P}(u|y)$ and the prior $\mu_0 = \mathbb{P}(u)$ and

$$\Phi(u, y) = \frac{1}{2}||y - G(u)||_\Gamma^2.$$

A sufficient condition for this to be well defined is that $\Phi(\cdot, y)$ is continuous as a mapping from X into $\mathbb{R}$ for each fixed y and that $\mu_0(X) = 1$ so that functions drawn from μ_0 are in X almost surely. Formula (16) then holds in infinite dimensions, exhibiting the posterior density with respect to the prior; in practical terms, this means that posterior expectations can be found by reweighting prior expectations by the right-hand side of Eq. (16).

The posterior distribution μ quantifies the uncertainty of the logarithm of the absolute permeability given production data, normalized by the prior. Since G is nonlinear, the posterior is non-Gaussian even when the prior μ_0 is Gaussian. Thus, there is no useful closed-form expression for the posterior distribution and it must be characterized by approximation; we do this by means of sampling.

Before we describe the approach to sampling μ that we adopt in this paper, we note that, if we assume the prior μ_0 is Gaussian with mean $\overline{u}$ and covariance C, it follows (see [10, 20]) that the MAP estimate u_{MAP} is the minimizer of the Tikhonov-regularized least squares functional

$$J(u) = \Phi(u, y) + \frac{1}{2}||u - \overline{u}||_C^2. \tag{17}$$

The MAP estimator is the typical estimate computed in standard history matching problems where the goal is to recover the truth by fitting historic production data. Note that the Bayesian approach thus subsumes this classical approach to inversion while also providing rigorous quantification of uncertainty of predictions, given clear assumptions on the prior and noise probabilities.

3.2 Sampling the posterior with Markov chain Monte Carlo (MCMC)

For PDE-based inverse problems where the unknown is a function, it is desirable that the resulting methods for sampling the posterior distribution scale well with respect to mesh refinement; this is because the ability to resolve the mesh progresses with increased computer power and practitioners will often seek to use improvements in computer resources to further refine computational meshes as they strive for greater fidelity. Designing sampling algorithms which perform well under mesh refinement may be achieved by conceiving of the algorithms on function space (the limit of infinite mesh refinement) and only then discretizing; thus, these algorithms are built directly on the infinite-dimensional Bayesian framework introduced in the preceding subsection. A review of state-of-the-art MCMC methods conceived in this way, together with numerical evidence of the mesh-invariant properties of these methods, may be found in [9]. Throughout this paper, we will use the following pCN-MCMC [9, Section 5.2] method, which may be viewed as the natural generalization of random walk Metropolis methods to the sampling of measures μ defined via Eq. (16) with μ_0 as the Gaussian $N(\overline{u}, C)$:

Algorithm 1 [pCN-MCMC] Take $u^{(0)} \sim N(\overline{u}, C)$, $n = 1$, and $\beta \in (0, 1)$. Then,

1. pcN proposal. Generate u from

$$u = \sqrt{1 - \beta^2} u^{(n)} + (1 - \sqrt{1 - \beta^2})\overline{u} + \beta\xi,$$
$$with \ \xi \sim N(0, C). \tag{18}$$

2. Set $u^{n+1} = u$ with probability $a(u^n, u)$ and $u^{n+1} = u^n$ with probability $1 - a(u^n, u)$, where

$$a(u, v) = \min\left\{1, \exp\left(\Phi(u, y) - \Phi(v, y)\right)\right\}. \tag{19}$$

3. $n \mapsto n + 1$ and repeat.

All the probabilities are generated independently of one another, leading to a Markov chain which is invariant with respect to μ. Notice that the small change in proposal, when compared with the standard random walk MCMC [23], results in an acceptance probability defined via differences of Φ and not J. Because Φ is finite with respect to μ while J is not, this leads to a considerably improved algorithm which has desirable $\dim(X)$-independent properties when implemented on a sequence of approximating problems with $\dim(X) \to \infty$. Therefore, for large $\dim(X)$ like the one considered here, pCN-MCMC provides a more robust and efficient technique than the standard MCMC approaches. In particular, the method does not degenerate in the limit like the standard random walk version does and is therefore robust to arbitrary mesh refinement. Numerical evidence of these scaling properties, together with references to relevant theory, can be found in [9].

The proposal chain in pCN-MCMC preserves the prior, and so the accept/reject step serves only to reweight the proposals so that their distribution is consistent with the likelihood. In particular, the accept/reject step depends only on the log likelihood Φ and not on the regularized log likelihood J; the latter contains terms from the prior which are not needed because the proposal is exact for the prior. There is a wealth of literature on choosing the optimal step size β (see [9] and references therein), but typically, one must do this either by trial and error or adaptively such that the acceptance rate is far from 0 and from 1. Throughout this paper, we have tuned β so that the acceptance rate is in the range 0.2–0.3.

Based on the forward models of Section 2, the purpose of the present work is to design synthetic experiments for solving the Bayesian data assimilation problem, i.e., finding the posterior distribution. By implementing the pcN-MCMC algorithm, we characterize this posterior and generate a gold standard against which the Gaussian approximations that we introduce in the following section are compared.

4 Gaussian approximations of the posterior

In this section, we introduce some standard ad hoc methods that use Gaussian approximations to sample the posterior distribution (see Eq. (16)). In particular, we consider LMAP, RML, EnKF, and EnSRF which have been typically used for history matching and uncertainty quantification in the Bayesian framework of data assimilation of petroleum reservoirs. While many variants of the aforementioned techniques can be found in the literature [27], here we focus on the most standard and typical implementations used for history matching. For each of the aforementioned techniques, the objective of the subsequent description is twofold. First, we introduce the algorithm and the associated computational cost. Second, we indicate the type of Gaussian approximation made for the definition of the technique under consideration.

4.1 Linearization around the MAP (LMAP)

As described in Section 3, the minimizer of J introduced in Eq. (17) defines the MAP estimator, i.e.,

$$u_{\text{MAP}} = \text{argmin}_u \left\{ \Phi(u, y) + \frac{1}{2} ||u - \overline{u}||_C^2 \right\}. \quad (20)$$

We can further define

$$C_{\text{MAP}} = C - CQ^T (QCQ^T + \Gamma)^{-1} QC \quad (21)$$

where $Q \equiv DG(u_{\text{MAP}})$ is the Fréchet derivative of G evaluated at $u = u_{\text{MAP}}$. The linearization around the MAP [26, Section 10.5] consists of approximating the posterior μ in Eq. (16) by $\mu \approx N(u_{\text{MAP}}, C_{\text{MAP}})$.

The LMAP algorithm approximates the posterior with an ensemble of N_e realizations from $N(u_{\text{MAP}}, C_{\text{MAP}})$ [26, Section 10.5]. This ensemble can then be used to approximate integrals with respect to the posterior of *nonlinear* functions of u. Note that when G is linear, $\mu = N(u_{\text{MAP}}, C_{\text{MAP}})$ and then u_{MAP} and C_{MAP} are the mean and covariance of the posterior. The algorithm, however, is well defined in general and may thus be applied to cases in which G is nonlinear.

Algorithm 2 [LMAP]

1. Compute u_{MAP} and C_{MAP} from Eqs. (20) and (21), respectively.
2. Compute the Cholesky factor L of C_{MAP}, i.e., $C_{MAP} = LL^T$.
3. For $j \in \{1, \ldots, N_e\}$, generate

$$u^{(j)} = u_{MAP} + L^T z^{(j)} \quad (22)$$

where $z^{(j)} \sim N(0, I)$.

Samples generated by Eq. (22) are drawn from $N(u_{\text{MAP}}, C_{\text{MAP}})$, and so the ensemble $\{u^{(j)}\}_{j=1}^{N_e}$ provides an approximation to $N(u_{\text{MAP}}, C_{\text{MAP}})$ and hence the posterior. The computational cost of LMAP depends on the cost of computing the MAP estimator (see Eq. (20)) and the factorization of C_{MAP}. For the present work, we develop implementations of the Levenberg–Marquardt algorithm of [26, Section 8.4] with the stopping criteria given by (8.82) and (8.83) from [26, Section 8.5] with $\epsilon_0 = 10^{-3}$ and $\epsilon_1 = 10^{-3}$. The selection of the LM parameter λ is chosen according to the guidelines discussed in [26, Section 8.5] (see also [34]). The mean of the prior distribution is used as the initial guess for the LM method. It is worth mentioning that, within the context of reservoir characterizations, multiple techniques for computing the MAP estimator have been widely studied (e.g., BFGS, LBFGS, Gauss–Newton) [26, Section 8]. It is of interest to evaluate the optimal minimization technique, but this is beyond the scope of our present work.

4.2 Randomized maximum likelihood (RML)

The RML technique was developed as an attempt to accelerate MCMC methods for sampling the posterior from Bayesian data assimilation in reservoir models [28]. The main idea of RML is to construct an ensemble of MAP estimators from randomized objective functions (see Eq. (20)). A standard implementation of RML is presented in the following algorithm:

Algorithm 3 [RML] For $j \in \{1, \ldots, N_e\}$:

1. Generate $u^{(j)} \sim N(\overline{u}, C)$.
2. Define $y^{(j)} = y + \eta^{(j)}$ with $\eta^{(j)} \sim N(0, \Gamma)$.
3. Compute

$$u_{RML}^{(j)} = \text{argmin}_u \left\{ \Phi(u, y^{(j)}) + \frac{1}{2} ||u - u^{(j)}||_C^2 \right\}. \quad (23)$$

In the case where G is linear, the RML algorithm can be shown to sample the posterior distribution (i.e., from $\mu = N(u_{\text{MAP}}, C_{\text{MAP}})$) [22]. For the nonlinear case of interest here, the RML algorithm provides an approximation of

the nature which, to the best of our knowledge, has not been systematically understood.

Note that, for each ensemble member, RML requires the solution to the minimization problem (see Eq. (23)). Nevertheless, since each minimization problem is independent from one another, RML is then embarrassingly parallelizable. Each of those minimization problems has the same structure as the one that we solve for the MAP estimator (see Eq. (20)). For a relatively small problem, the computational cost of computing L in LMAP, given Q which has already been constructed while solving Eq. (20), as well as the generation of Eq. (22), is negligible compared to the cost of one forward model evaluation. Thus, the computational cost of RML is roughly N_e times the computational cost of LMAP, although the effect of the multiplier N_e can be ameliorated in a parallel context.

Similarly to our implementation of the MAP, for RML, we consider the Levenberg–Marquardt method and the corresponding stopping criteria mentioned above. Each of the corresponding samples $u^{(j)}$ of the prior distribution is used as the initial guess for the LM method. Improving the optimization technique required for Eq. (23) can reduce the overall computational cost of RML. Alternative methods to reduce the computational cost of RML by means of a truncated SVD approach can be found in [34].

4.3 Ensemble Kalman filter (EnKF)

Ensemble methods based on the Kalman filter have been extensively applied for Bayesian data assimilation in petroleum reservoir applications. For a complete review of most of the EnKF implementations, we refer the reader to the monograph of [1]. In this section, we briefly discuss some relevant aspects of EnKF in the context of history matching of petroleum reservoirs. These ensemble Kalman filter-type algorithms make Gaussian approximations in a sequential manner as we described below. As a result, for the general case, those techniques do not provide correct sampling of the posterior (see Eq. (16)). Nevertheless, due to its ease of implementation and low computational cost, ensemble Kalman filter-type methods are arguably the only feasible techniques for online data assimilation in subsurface applications.

4.3.1 Introduction and main algorithm

In order to introduce the algorithms, we first consider a sequential formulation of the reservoir model. In particular, let us define v_n as the state variable at time t_n and S as the state space. For example, for the single-phase model in

Section 2, $v_n = p(x, t_n)$. We define the solution operator $\Psi_n : S \times X \to S$:

$$v_n = \Psi_n(v_{n-1}, u) \tag{24}$$

which, for a given parameter u, maps the state variable from time $t = t_{n-1}$ to $t = t_n$. In practice, Ψ_n is simply the numerical solver that arises from the time discretization of the reservoir model under consideration. In addition, we assume that data are given at each of these points in time and are correlated between times only through the state itself, i.e.,

$$y_n = M_n(v_n) + \eta_n, \tag{25}$$

where $\eta_n \sim N(0, \Gamma_n)$ and $M_n : S \to \mathbb{R}^{N_w}$ is the measurement functional acting on the state variable at time $t = t_n$. For the models in Section 2, M_n is defined by Eqs. (3) and (12) and N_w is the number of total wells. Define

$$z = \begin{pmatrix} u \\ v \\ w \end{pmatrix}, \qquad \Xi_n(z) = \begin{pmatrix} u \\ \Psi_n(v, u) \\ M_n(\Psi_n(v, u)) \end{pmatrix}. \tag{26}$$

Since the permeability in the forward reservoir model does not change in time, it follows that Eqs. (24) and (25) can be written as

$$z_n = \Xi_n(z_{n-1}), \tag{27}$$

$$y_n = H z_n + \eta_n \tag{28}$$

where $H = (0, 0, I)$. We now consider the following standard perturbed observation version of EnKF [1]:

Algorithm 4 [EnKF] Construct an initial ensemble

$$z_0^{(j,a)} = \begin{pmatrix} u_0^{(j)} \\ v_0 \\ M_0(v_0) \end{pmatrix} \tag{29}$$

where $\{u_0^{(j)}\}_{j=1}^{N_e} \sim \mu_0$ and v_0 is the initial condition for the state variable. For $n = 1, \ldots, N_M$:

1. Prediction step: Propagate the ensemble of particles forward under Eq. (27) giving

    $$z_n^{(j,f)} = \Xi_n\left(z_{n-1}^{(j,a)}\right) \qquad j \in \{1, \ldots, N_e\}. \tag{30}$$

    From this ensemble, we define a sample mean and covariance as follows:

    $$\overline{z}_n^f = \frac{1}{N_e} \sum_{j=1}^{N_e} z_n^{(j,f)} \tag{31}$$

    $$C_n^f = \frac{1}{(N_e-1)} \sum_{j=1}^{N_e} z_n^{(j,f)} \left(z_n^{(j,f)}\right)^T - \overline{z}_n^f \left(\overline{z}_n^f\right)^T. \tag{32}$$

2. Analysis step: Compute the updated ensembles

    $$z_n^{(j,a)} = z_n^{(j,f)} + K_n(y_n^{(j)} - H z_n^{(j,f)}) \tag{33}$$

    where

    $$K_n = C_n^f H^T \left(H C_n^f H^T + \Gamma_n\right)^{-1} \tag{34}$$

    and

    $$y_n^{(j)} = y_n + \eta_n^{(j)}, \qquad \eta_n^{(j)} \sim N(0, \Gamma_n).$$
    $$j \in \{1, \ldots, N_e\} \tag{35}$$

Here, the $\eta_n^{(j)}$ is chosen i.i.d. From Eq. (30) and the block structure of the matrices involved, expression (33) can be written as

$$z_n^{(j,a)} = \Xi_n\left(z_{n-1}^{(j,a)}\right) + C_n^{zw,f}(C_n^{ww,f} + \Gamma)^{-1}$$
$$\left(y_n^{(j)} - M_n\left(\Psi_n\left(v_{n-1}^{(j,a)}, u_{n-1}^{(j,a)}\right)\right)\right) \qquad (36)$$

where

$$C_n^{zw,f} = \begin{pmatrix} C_n^{uw,f} \\ C_n^{vw,f} \\ C_n^{ww,f} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{N_e}\sum_{j=1}^{N_e} u_n^{(j,f)}\left(w_n^{(j,f)}\right)^T - \overline{u}_n^f\left(\overline{w}_n^f\right)^T \\ \frac{1}{N_e}\sum_{j=1}^{N_e} v_n^{(j,f)}\left(w_n^{(j,f)}\right)^T - \overline{v}_n^f\left(\overline{w}_n^f\right)^T \\ \frac{1}{N_e}\sum_{j=1}^{N_e} w_n^{(j,f)}\left(w_n^{(j,f)}\right)^T - \overline{w}_n^f\left(\overline{w}_n^f\right)^T \end{pmatrix}.$$

We recall that $w \in \mathbb{R}^{N_w}$ where N_w is the number of wells. Typically, N_w is much smaller than the dimensions of the (discretized) parameter space X. Consequently, the computational cost of constructing C_n^{zw} and C_n^{ww} and inverting the $(C_n^{ww} + \Gamma_n)^{-1}$ in Eq. (36) is negligible compared to the cost of computing $\Xi_n(z_{n-1}^{(j,a)})$, which from Eq. (26) we can see is mainly determined by the cost of $\Psi_n(v_{n-1}^{(a,j)}, u_{n-1}^{(j,a)})$ (i.e., running the reservoir simulator in the time interval $[t_{n-1}, t_n]$). Therefore, the computational cost of the EnKF is approximately N_e times the cost of a forward model simulation.

4.3.2 Derivation by Gaussian approximation of the filtering distribution

We now indicate how a Gaussian approximation gives rise to the EnKF algorithm presented above. We start by defining the conditional measures for $n_1, n_2 \le N_M$:

$$\mu_{n_1|n_2}(z_{n_1}) = \mathbb{P}\left(z_{n_1}|\{y_k\}_{k=1}^{n_2}\right). \qquad (37)$$

In the filtering approach, the prior distribution ($\mu_{n|n-1}$) of z_n and the data up to the previous time $t = t_{n-1}$ are combined with the data provided at the current time $t = t_n$ to define the posterior distribution ($\mu_{n|n-1}$) of z_n (given data up to the current time $t = t_n$). The latter can be obtained from Bayes' rule:

$$\frac{\mu_{n|n}(z)}{\mu_{n|n-1}(z)} \propto \exp\{-\Phi_n(z)\} \qquad (38)$$

where

$$\Phi_n(z) = \frac{1}{2}||y_n - Hz||_\Gamma^2. \qquad (39)$$

The EnKF approach then assumes that $\mu_{n|n-1}(z)$ is the Gaussian measure $N(\overline{z}_n^f, C_n^f)$ where $\overline{z}_n^f$ and C_n^f are the ensemble mean and covariance defined in Eqs. (31) and

(32), respectively. Given this Gaussian assumption, it is not difficult to see that Eq. (38) implies that $\mu_{n|n}(z) = N(\overline{z}_n^a, C_n^a)$ with

$$\overline{z}_n^{(a)} = \overline{z}_n^{(f)} + K_n(y_n - H\overline{z}_n^{(f)}), \qquad (40)$$
$$C_n^a = (I - K_n)C_n^f, \qquad (41)$$

and K_n defined in Eq. (34). In [22, Appendix A], it has been shown that the ensemble updates defined in Eq. (33) are samples from $\mu_{n|n}(z) = N(\overline{z}_n^a, C_n^a)$. In fact, [22, Appendix B] shows that the analysis step (see Eq. (33)) can be derived from an application of RML under the Gaussian approximation $\mu_{n|n-1}(z) \approx N(\overline{z}_n^f, C_n^f)$. Indeed, it is straight forward to show that Eq. (33) can be obtained from

$$z_n^{(j,a)} = \text{argmin}_z \left(||\Gamma_n^{-\frac{1}{2}}\left(y_n^{(j)} - Hz\right)||^2 \right.$$
$$\left. + ||(C_n^f)^{-\frac{1}{2}}\left(z - z_n^{(j,f)}\right)||^2\right) \qquad (42)$$

which is a sequential version of Eq. (23), for the augmented state z with a prior $N\left(\overline{z}_n^f, C_n^f\right)$ and the linear measurement operator H.

For our evaluation and comparison of techniques, we consider the outcome of the EnKF algorithm after all data have been assimilated in the time interval $[0, T]$. In other words, we are interested in $\mu_{n|N_M}(z_n) = \mathbb{P}(z_n|\{y_k\}_{k=1}^{N_M})$ which corresponds to the probability of z_n after all data have been assimilated (recall N_M is the total number of assimilation times). Then, the posterior $\mu_{n|N_M}(z_n)$ computed via the EnKF algorithm provides an approximation to μ defined in Eq. (16).

4.3.3 Further modifications

While the standard version of EnKF has been successfully applied for some history matching problems, several shortcomings due to sampling error have been identified. In particular, when a small ensemble is used, spurious correlations often cause gross overestimation of the physical variables that EnKF aims at recovering (e.g., permeability). In addition to the issues caused by small sample size, standard EnKF with a small ensemble is suboptimal when a large amount of data are assimilated. This can be easily observed from the two following properties of EnKF. First, the ensemble updates Eq. (33), when a linear combination of the initial ensemble members is projected into the parameter space [22, 24]. Second, the ensemble updates minimize Eq. (42), which involves fitting data at each assimilation time. Therefore, when the prior ensemble is small, the EnKF updates cannot fit large amount of data within the subspace generated by the prior ensemble. These shortcomings of using standard EnKF have given rise to several EnKF variants designed to reduce the spurious correlations

described above as well as increasing the number of degrees of freedom. In this work, we focus on the application of distance-based covariance localization which has recently been investigated in [6, 15]. In particular, the EnKF with localization that we implement for the forward models of Section 2 is given by the same EnKF algorithm described before, except that Eq. (34) is replaced by

$$K_n = \rho \circ C_n^f H^T \left(H(\rho \circ C_n^f)H^T + \Gamma_n \right)^{-1}. \tag{43}$$

Here, ρ, to be defined below, is a positive-definite matrix which induces localization and the matrix $\rho \circ C_n$ is the Schur product between ρ and C_n with entries defined by $[\rho \circ C_n]_{ij} = [\rho]_{ij}[C_n]_{ij}$. Due to the spurious correlations described above, matrix C_n^f may become positive semi-definite and the parameter update then lies in smaller subspace than the one generated by the prior ensemble. With properly chosen ρ, the matrix $\rho \circ C_n^f$ has full rank, and replacing C_n^f with $\rho \circ C_n^f$ increases the dimension of the linear subspace where the parameter update is sought. This, in turn, results in a better estimation. In terms of the block structure previously described, covariance localization becomes

$$z_n^{(j,a)} = z_n^{(j,f)} + \rho_{zw} \circ C_n^{zw} \left(\rho_{ww} \circ C_n^{ww} + \Gamma_n \right)^{-1}$$
$$\left(y_n^{(j)} - M_n \left(\Psi_n \left(v_{n-1}^{(j,a)}, u_{n-1}^{(j,a)} \right) \right) \right). \tag{44}$$

As in the covariance localization approach of [6], we consider only localization in the u-component (e.g., for the log permeability updates). In other words,

$$u_n^{(j,a)} = u_n^{(j,f)} + \rho_{uw} \circ C_n^{uw} \left(\rho_{ww} \circ C_n^{ww} + \Gamma_n \right)^{-1}$$
$$\left(y_n^{(j)} - M_n \left(\Psi_n \left(v_{n-1}^{(j,a)}, u_{n-1}^{(j,a)} \right) \right) \right). \tag{45}$$

Following the implementation of [15], each column of the localization matrix ρ_{uw} is defined as the fifth-order compact function of Gaspari–Cohn [19] localized at the corresponding measurement location. Each row of the matrix ρ_{ww} in Eq. (45) is obtained from ρ_{uw} by projecting it on the corresponding measurement location. By construction, ρ_{uw} and ρ_{ww} are positive definite.

Recent publications [6, 15] have investigated optimal choices for the critical length of the correlation function used for distanced-based localization. The focus of those investigations is to improve the ability of the EnKF with localization to recover the truth within the confidence interval provided by the ensemble. In contrast to [6, 15], our goal is to assess the performance of EnKF with localization for reproducing the uncertainty quantified by the posterior. However, for the present work, we consider a fixed critical length obtained from a simple trial–error procedure, which enables us to observe a significant effect of covariance localization in characterizing the posterior distribution. While the optimal choice of covariance localization

is beyond the scope of the present work, we recognize the importance for assessing optimal choices of covariance localization for providing better Gaussian approximation of the posterior distribution at a reasonable computational cost. Moreover, additional forms of covariance regularization (e.g., covariance inflation) should also be assessed.

4.3.4 Ensemble square root filter (EnSRF)

Sampling error that arises from perturbing the observations in standard EnKF has been often associated with a poor performance of history matching data. In order to avoid the aforementioned sampling error, an EnSRF is often used. Here, we consider the following EnSRF [14]:

Algorithm 5 [EnSRF] Construct an initial ensemble as in Eq. (29). For $n = 1, \ldots, N_M$:

1. Prediction step: Propagate the ensemble of particles forward under Eq. (27) yielding Eq. (30). Construct the sample mean and covariance from Eqs. (31)–(32). Additionally, define the deviations from the mean

$$\Delta z_n^{(j,f)} = z_n^{(j,f)} - \overline{z}_n^f. \tag{46}$$

2. Analysis step: Compute the updated mean $\overline{z}_n^{(a)}$ via formula (40) with K_n given by Eq. (34). Consider the matrices $\Delta Z_n^f := \left[\Delta z_n^{(1,f)} \ \Delta z_n^{(2,f)} \ \cdots \ \Delta z_n^{(N_e,f)} \right]$, with jth column $\Delta z_n^{(j,f)}$, and ΔZ_n^a defined analogously. Compute the matrix with updated deviations:

$$\Delta Z_n^a = (I - \tilde{K}_n H)\Delta Z_n^{(f)}\Theta \tag{47}$$

with

$$\tilde{K}_n = C_n^f H^T \left[H_n C_n^f H_n^T + \Gamma_n \right]^{-T/2}$$
$$\times \left[\left(H_n C_n^f H_n^T + \Gamma_n \right)^{1/2} + \Gamma_n^{1/2} \right]^{-1}. \tag{48}$$

The updated ensemble is then obtained from the expression

$$z_n^{(j,a)} = \overline{z}_n^{(a)} + \Delta z_n^{(j,a)}.$$
$$j \in \{1, \ldots, N_e\} \tag{49}$$

In expression (47), Θ is a $N_e \times N_e$ mean-preserving orthogonal random matrix constructed as suggested in [30]. The mean-preserving property of Θ ensures that $\sum_{j=1}^{N_e} \Delta z_n^{(j,a)} = 0$ and so the analyzed ensemble (see Eq. (49)) has mean $\overline{z}_n^{(a)}$ as required. In contrast to the EnKF, where Eq. (41) is only exactly satisfied in the limit of arbitrarily large ensemble size, the sample covariance computed from the EnSRF (finite) ensemble updates exactly satisfies Eq. (41), therefore providing a better approximation

of $\mu_{n|n}(z)$; it is then hoped that this will lead to a better approximation to the posterior distribution (see Eq. (16)) itself. A block structure similar to the one introduced before applies to the EnSRF. From this structure, it is easy to appreciate that only small matrices are involved in the square root computations. Therefore, the computational cost of EnSRF is essentially the same as EnKF, i.e., N_e times the number of forward model evaluations.

Even though the implementation of the EnSRF avoids the sampling error due to perturbing the observations, limitations related to the small ensemble size still apply. Distance-based localization can then be applied to the EnSRF as suggested in [14]. Concretely, we replace K_n in Eq. (34) with Eq. (43) and $\tilde{K}_n$ in Eq. (46) with

$$\tilde{K}_n = \left(\rho \circ C_n^f\right) H^T \left[H_n\left(\rho \circ C_n^f\right)H_n^T + \Gamma_n\right]^{-T/2}$$
$$\times \left[\left(H_n\left(\rho \circ C_n^f\right)H_n^T + \Gamma_n\right)^{1/2} + \Gamma_n^{1/2}\right]^{-1}. \quad (50)$$

Similar to the localization procedure for the EnKF, EnSRF is localized only in the u-component (log permeability) of Eq. (40) with the localization matrix ρ described above.

5 Numerical results

In this section, we present the results of three numerical experiments for assessing the Gaussian approximations defined in Section 4. These experiments are described on the three subsections which follow, each of which is organized as follows: (i) details of the forward model and the generation of synthetic data are provided, (ii) numerical results from the gold standard MCMC implementation described in Section 3 are discussed, (iii) the numerical results of Gaussian approximations are presented and the results in (iii) are compared against the results from (ii) in terms of their ability to reproduce mean and variance; and (iv) we assess the

performance of the Gaussian approximation at quantifying the uncertainty in reservoir model forecast.

5.1 Single-phase flow

For the first experiment, we consider the single-phase reservoir model of Section 2 on a square domain $D = [0, L] \times [0, L]$ with the production wells located at the points labeled by $P_1, \ldots, P_9$ in Fig. 1 (middle). The spatial and temporal discretization of Eq. (1) is conducted with cell-centered finite differences [29] and the backward Euler method, respectively. Relevant information of this model is displayed in the first column of Table 1. For each well term in the right-hand side of Eq. (1), we prescribe a production rate constant of 85 m^3/day during the total simulation time of 50 days. We recall that the measured variable in this model is the pointwise evaluation of the pressure at the production wells defined by Eq. (2). The reservoir domain and the time interval $[0, T]$ are discretized on a 60×60 grid and 20 time steps, respectively.

We consider a Gaussian prior distribution of log permeability:

$$\mu_0(u) = N(\overline{u}, C) \quad (51)$$

where the covariance is defined by $C = \kappa A^{-\alpha}$ with the operator $A = -\Delta$ defined on

$$D(A) = \left\{v \in H^2(D) | \nabla v \cdot \mathbf{n} = 0 \text{ on } \partial D, \int_D v = 0\right\}, \quad (52)$$

i.e., A is the negative Laplacian with no-flow boundary conditions and restricted to the domain of functions that have square-integrable second derivatives and average to zero. Covariance operators from a class including the one defined above have been used in the analysis of a range of Bayesian inverse problems (see [33] for an overview). The tunable parameters in Eq. (51) are defined as follows: $\overline{u}(x, y) = \log(5 \times 10^{-13} \text{ m}^2)$ for all $(x, y) \in D$, $\kappa = 2.0$, and $\alpha = 1.3$. In Fig. 2, we show some realizations of the
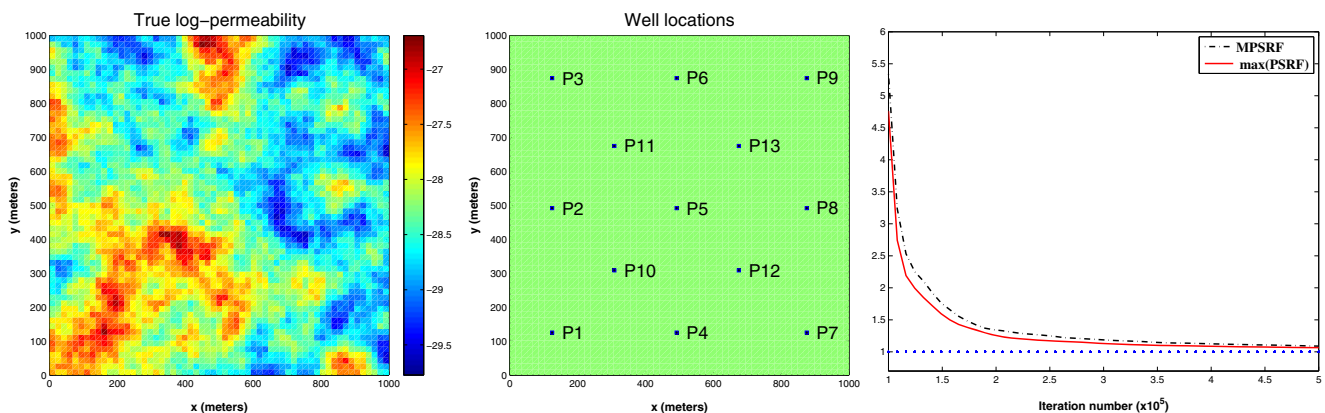

Fig. 1 Single-phase model. *Left*: True log permeability [log m^2]. *Middle*: Well configuration. *Right*: Gelman–Rubin diagnostic

Table 1 Reservoir model description

Variable	Single-phase reservoir	Water–oil reservoir (small number of wells)	Water–oil reservoir (large number of wells)
L [m]	10^3	2×10^3	5×10^3
c [Pa^{-1}]	10^{-8}	0.0	0.0
ν_o [Pa s]	10^{-2}	10^{-2}	10^{-2}
T [years]	0.13	5	3.5
p_0 [Pa][a]	3.5×10^7	2.5×10^7	2.5×10^7
s_0[a]	Not applicable	0.2	0.2
ν_w [Pa s]	Not applicable	5×10^{-4}	5×10^{-4}
s_{iw}	Not applicable	0.2	0.2
s_{ro}	Not applicable	0.2	0.2
P_{bh}^l [Pa][b]	Not applicable	2.7×10^7	2.0×10^7
q_w^l [m^3/day][b]	Not applicable	2.6×10^3	1.8×10^2
a_w	Not applicable	0.3	0.3
a_o	Not applicable	0.9	0.9

[a]Constant in Ω

[b]Constant in $[0, T]$

prior distribution (see Eq. (51)). It is important to mention that other choices of C can also be used. In particular, C can be defined in terms of a standard correlation function (e.g., spherical, exponential, etc.). Our choice, however, has the advantage that C is diagonalized by cosine functions, so that the fast Fourier transform can be used to efficiently sample from the Gaussian $N(0, C)$ as required to implement Algorithm 1, the pCN-MCMC. Sampling from the prior on the spectral domain is straightforward and computationally inexpensive. This is a desirable property since at each iteration of MCMC (see Eq. (18)), a draw from the prior is generated for computing the proposal. The spatial correlation function of draws from the prior is, of course, simply the Green's function of C^{-1}

since the Green's function of an operator is defined as the kernel of the spatial integral operator representation of its inverse.

For the generation of synthetic pressure data, we first define the "true log permeability" denoted by $u^\dagger$ and displayed in Fig. 1 (left). This true log permeability is generated from the prior distribution defined above. Synthetic data are generated by first solving Eq. (1) for p with $u = u^\dagger$. Then, $G(u^\dagger)$ is calculated from Eq. (4). Finally, we add a random error, i.e., $y = G(u^\dagger) + \eta$ with $\eta \sim N(0, \sigma^2 I)$ with $\sigma = 4 \times 10^5$ Pa. The measurement times used in Eq. (2) are $t_1 = 5$, $t_n = 10n$ days, $n = \{2, \ldots, 5\}$.

Synthetic data are used in the pCN-MCMC Algorithm 1 with $\beta = 0.015$. Our MCMC results consist of 110 chains

Fig. 2 Single-phase model. Samples from the prior distribution [log m^2]
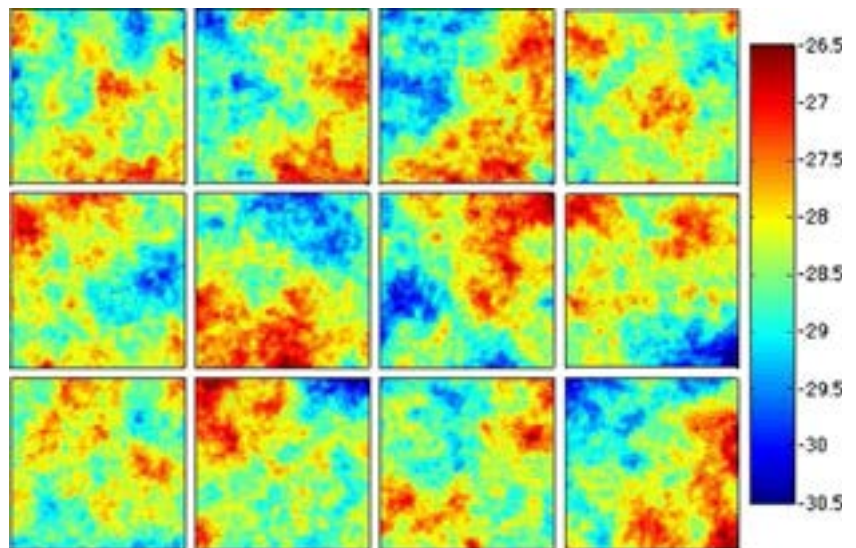
starting from independent draws from the prior distribution. After a burn-in period of 1×10^4, each chain generates 5×10^5 samples. Therefore, a total of 5.5×10^7 samples are generated with our MCMC implementations. For assessing the convergence of our chains, we consider the diagnostics suggested by Gelman and Rubin in [5]. In Fig. 1 (right), we display the maximum of the potential scale reduction factor (PSRF) and the multivariate potential scale reduction factor (MPSRF) for the smallest $J = 16$ frequencies that account for 76 % of the total prior energy defined by $e(J) = \sum_{j=1}^{J} \lambda_k / \sum_{j=1}^{\infty} \lambda_k$ where λ_k is the eigenvalues of the prior covariance C (ordered as $\lambda_1 \geq \lambda_2 \geq \dots$). Convergence of the chains is achieved when the maximum PSRF and the MPSRF are close to 1. From Fig. 1 (right), the max(PSRF) and the MPSRF have dropped below 1.1 after 5×10^5 iterations where we establish the convergence of our MCMC chains. From the numerical evidence of convergence of our chains, we conclude that the MCMC provides samples from the posterior. The associated mean and variance fields (computed from our 5.5×10^7 MCMC samples), denoted by $u_{\mathrm{pos}}(x)$ and $\sigma_{\mathrm{pos}}(x)$, are used as the gold standard for the assessment of the Gaussian approximations that we discuss below. In Fig. 3, we display some samples from the independent chains (i.e., uncorrelated) obtained after convergence was achieved. Although there are substantial differences among those realizations, some common spatial features can be observed. For example, note the high permeability region around wells P_1 and P_6. Furthermore, the variability is considerably lower than under the prior, as exhibited in Fig. 2; this indicates that the data used are quite informative.

The numerical implementation of the following Gaussian approximations are conducted: LMAP, RML, EnKF, EnKF with localization, EnSRF, and EnSRF with localization. We consider two sets of experiments corresponding to ensemble sizes $N_e = 50$ and $N_e = 150$. For the localization

matrix, we chose a critical length equal to the length of the edge of the reservoir domain L. This choice of the critical length is sufficient to appreciate the effect of localization for characterizing the posterior distribution. However, we remark that optimal choices of this critical length should be further studied in future work. It is also important to note that the mean and variance for (the sequential methods) EnKF, EnKF with localization, EnSRF, and EnSRF with localization are computed after all the measurements have been assimilated. In other words, exactly the same data used to sample the posterior via MCMC are also used for all the Gaussian approximations under consideration. From each of these approximations, we compute the mean and variance which we compare against the mean u_{pos} and variance σ_{pos} of the posterior distribution generated with the MCMC method. However, since the results from each of the Gaussian approximations depend on the particular choice of initial (prior) ensemble [35, 36], we conduct a set of 50 experiments corresponding to 50 different initial ensembles. The mean and variance for one of those experiments in the case $N_e = 50$ are shown in Figs. 4 and 5, respectively. In Table 2, we display the relative errors of the deviation of the mean with respect to $\overline{u}$ (the prior mean) and the relative error of the variance defined by

$$\epsilon_u = \frac{||(\hat{u} - \overline{u}) - (u_{\mathrm{pos}} - \overline{u})||_{L^2(D)}}{||(u_{\mathrm{pos}} - \overline{u})||_{L^2(D)}}$$

and

$$\epsilon_\sigma = \frac{||\hat{\sigma} - \sigma_{\mathrm{pos}}||_{L^2(D)}}{||\sigma_{\mathrm{pos}}||_{L^2(D)}}, \tag{53}$$

averaged over 50 experiments associated with different initial ensembles. In the previous expression, $\hat{u}$ and $\hat{\sigma}$ are the mean and variance of the Gaussian approximation under consideration. We reiterate that $\hat{u} - \overline{u}, \hat{\sigma}$, etc.

Fig. 3 Single-phase model. Samples from the posterior distribution (characterized with MCMC) [log m^2]
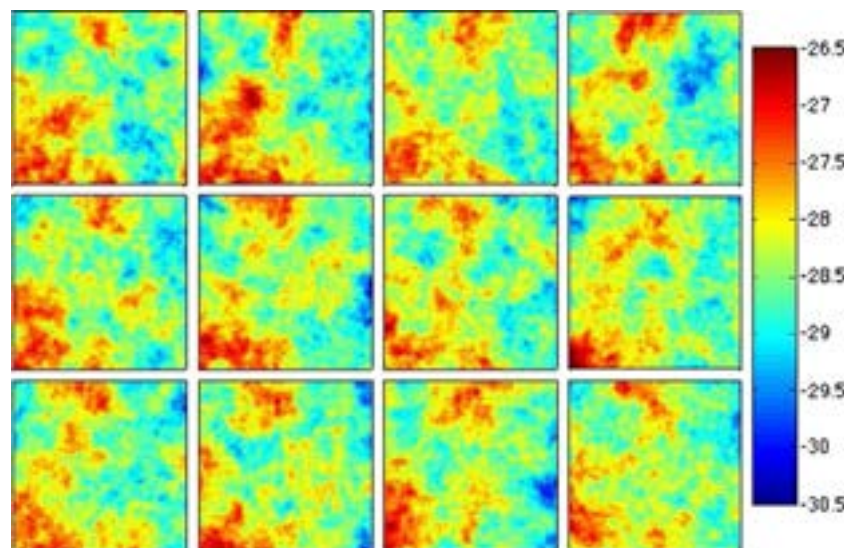
Fig. 4 Single-phase model. Mean of the posterior distribution (characterized with MCMC) and Gaussian approximations [$\log m^2$]
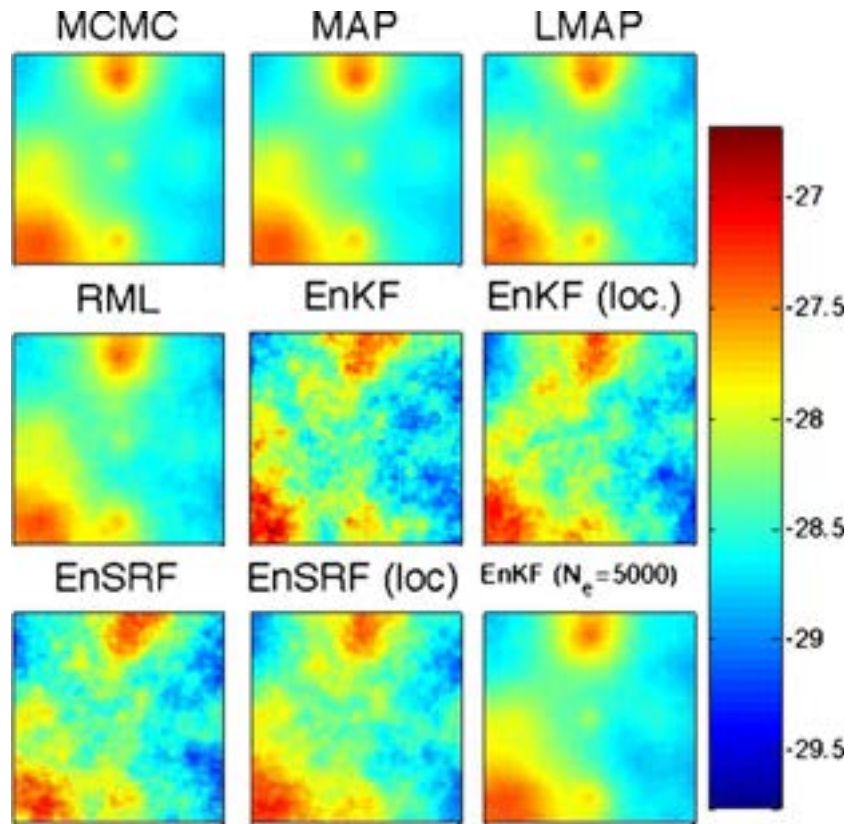

Fig. 5 Single-phase model. Variance of the posterior distribution (characterized with MCMC) and Gaussian approximations [$(\log m^2)^2$]
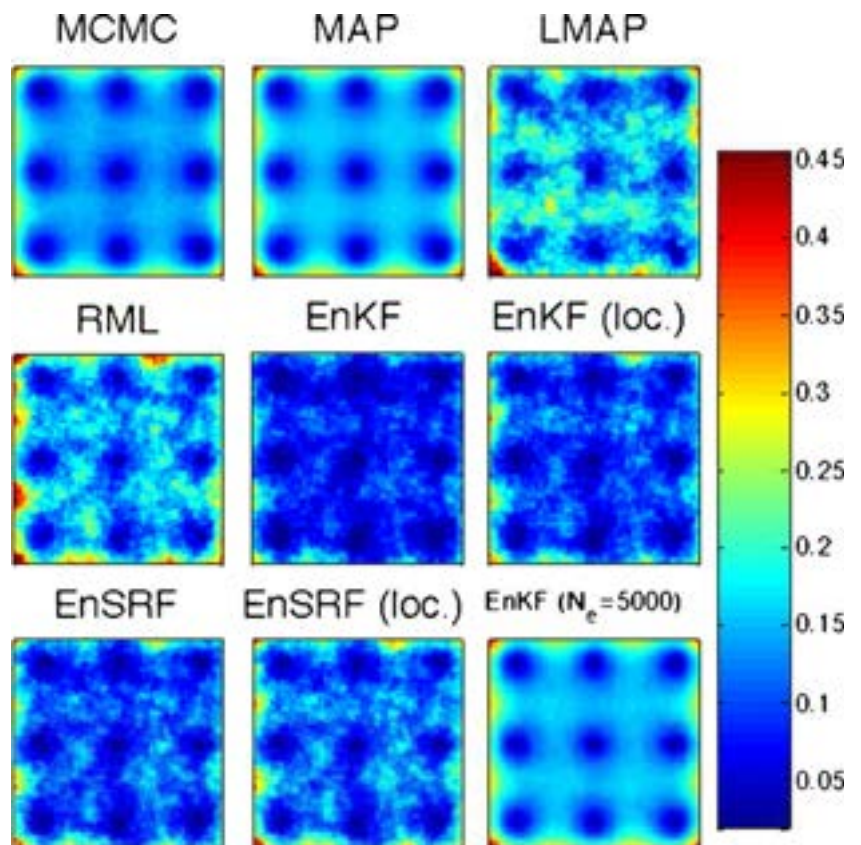
Table 2 Evaluation of Gaussian approximations on the single-phase model

Method	Relative error in the mean ϵ_u	Relative error in the variance ϵ_σ	Computational cost (forward model runs)
MCMC	0.000	0.000	5.5×10^7
MAP	0.030	0.094	2.5×10^1
LMAP ($N_e = 50$)	0.145	0.239	2.5×10^1
RML ($N_e = 50$)	0.148	0.240	1.25×10^3
EnKF ($N_e = 50$)	0.839	0.448	5.0×10^1
EnKF (localization, $N_e = 50$)	0.621	0.333	5.0×10^1
EnSRF ($N_e = 50$)	0.665	0.306	5.0×10^1
EnSRF (localization, $N_e = 50$)	0.518	0.231	5.0×10^1
LMAP ($N_e = 150$)	0.088	0.158	2.5×10^1
RML ($N_e = 150$)	0.087	0.159	3.75×10^3
EnKF ($N_e = 150$)	0.472	0.154	1.5×10^2
EnKF (localization, $N_e = 150$)	0.453	0.123	1.5×10^2
EnSRF ($N_e = 150$)	0.370	0.121	1.5×10^2
EnSRF (localization, $N_e = 150$)	0.403	0.130	1.5×10^2
EnKF ($N_e = 1,250$)	0.192	0.075	1.25×10^3
EnKF ($N_e = 2,500$)	0.120	0.089	2.5×10^3
EnKF ($N_e = 5,000$)	0.102	0.094	5.0×10^3

are (discrete approximations of) fields in the space of square-integrable functions $L^2(D)$ and the distance is therefore measured with a standard $L^2(D)$ (Euclidean) norm. The error of both mean and variance is therefore quantified by relative root-mean-square error over the domain D. The right column of Table 2 indicates the computational cost for computing each of the techniques (per experiment) in terms of forward model runs. As stated in Section 4, while the computational cost of the Kalman filter-type methods is stable with respect to the details of the implementation, the cost of RML and LMAP depends crucially on the optimization technique used for solving Eqs. (20) and (23). Our implementation of the Levenberg–Marquardt technique costs around five forward model runs per iteration. Furthermore, in average, each of optimization problems (see Eq. (23)) converged into five iterations and so the average computational cost of Eq. (23) is 25 forward model runs. This computational cost can be potentially reduced by applying a more efficient optimization technique.

Since we are assessing the Gaussian approximations only in terms of mean and variance, we can additionally measure the error (with respect to the posterior) of the exact mean and variance of $N(u_{MAP}, C_{MAP})$ given directly in Eqs. (20) and (21), respectively. The corresponding relative errors are provided in the "MAP" row in Table 2. From construction, it is clear that the mean and covariance of LMAP are u_{MAP} and C_{MAP} for

sufficiently large N_e. The results for this experiment indicate that $N(u_{MAP}, C_{MAP})$ provides a good approximation to the posterior in terms of mean and variance. Therefore, more samples from LMAP can be generated at a negligible cost so that its mean and variance approach u_{MAP} and C_{MAP}, respectively.

Table 2 also indicates that, among the all the Gaussian approximation with $N_e = 50$ and $N_e = 150$, RML provides the best approximation in terms of the mean. In those cases, the worst performance in terms of mean and variance was obtained with EnKF. However, for the case of small ensemble size $N_e = 50$, considerable improvement was obtained by applying the localization approach described in the preceding section (see Eq. (45)). Additionally, note that the application of localization in the EnSRF (see expression (46)) further reduces the relative errors in the mean and variance. In fact, among all the techniques with $N_e = 50$, the best approximation in terms of variance is given by the EnSRF with localization. It is clear that sampling error causes severe limitations in the performance of EnKF and EnSRF for small ensemble sizes. It is worth mentioning that issues of the EnKF and EnSRF due to sampling error have been often reported [6, 27] and used as motivation for covariance regularization. However, this existing work is focused on (1) history matching production data, (2) recovering the true permeability, and (3) recovering data generated with the true permeability within the estimated confidence interval. In contrast, here we assess the

performance in terms of the posterior distribution of the Bayesian framework.

For the case of larger ensemble size $N_e = 150$, sampling error is reduced and all Kalman-based techniques provide similar results in terms of mean and variance. Although the ability to use large ensembles is constrained by the computational resources at hand, it is instructive to consider cases with large ensembles. Therefore, we additionally consider three implementations of EnKF for larger ensembles: $N_e = 1,250$, $N_e = 2,500$, and $N_e = 5,000$. These results appear at the end of Table 2. Note that $N_e = 1,250$ corresponds to the case where the computational cost of EnKF coincides with the cost of our implementation of RML for $N_e = 50$. While the performance of EnKF improved significantly for $N_e = 1,250$, RML still provides a better approximation in terms of the mean. In addition, we observe that although increasing the size of the ensemble may reduce the sampling error, this is not associated to the convergence to the posterior. Actually, it is clear from this experiment that the variance of EnKF for large N_e seems to diverge from the variance of the posterior.

With the previous results, we are able to appreciate and evaluate the differences in the approximations provided by each of the Gaussian approximations of the posterior. This posterior is the conditional probability of the unknown (permeability) given production data collected during the 50 days of simulation. For practical applications, it is of particular interest to assess how different Gaussian approximations fare at reproducing the probability distribution of various predicted quantities with respect to the posterior, in other words to assess how the approximate algorithms fare in the quantification of uncertainty in these predictions. The assessment of performance in terms of the distribution of predictions under the posterior is conducted by creating a new flow scenario as we now describe. Assume that after the initial 50 days of simulation (that we used to generate synthetic data), we now drill new wells labeled by P_{10}, P_{11}, P_{12}, and P_{13} in Fig. 1. These new wells are operated at a constant production rate of 60 m^3/day during 100 days. During this 100 days of forecast, the old wells $P_1, \ldots, P_9$ are first shut down for a pressure build-up time window of 50 days, followed by a constant production rate of 60 m^3/day during the remaining 50 days. In Fig. 6, we show, as a function of time, the pressure at the well locations P_1, P_6, and P_{11}. The first 50 days corresponds to the data assimilation phase, and the subsequent 100 days is the prediction. In the first row of Fig. 6, we display the pressure obtained from the reservoir simulation with 100 permeabilities obtained from the prior distribution (see Eq. (51)). The second row corresponds to the pressure obtained from the simulation with the samples from the posterior obtained from independent MCMC chains. Subsequent rows of Fig. 6 correspond to the pressure obtained from

simulating the permeabilities obtained from some of the Gaussian approximations under consideration (obtained from one of the 50 experiments with $N_e = 50$). The vertical line divides the data assimilation phase from the forecast. Additionally, since we are interested in the performance with respect to the posterior, for each curve presented in Fig. 6, we include a red curve of the pressure at the corresponding well location obtained by simulating the mean of the posterior distribution (i.e., top-left field of Fig. 4).

We expect the uncertainty in the model predictions to be considerably smaller at the old wells (where measurements were assimilated) than the one at the wells drilled after the assimilation. Indeed, note from Fig. 6 that in P_{11}, the posterior under the model predictions is visually closer to the prior than the ones at P_1 and P_6 (where data were assimilated). This, in turn, indicates the uninformative effect of the data at the location of P_{11} where measurements were not collected. For the new wells where uncertainty is larger, we can appreciate the performance on the Gaussian approximations. Note from Fig. 6 that the EnKF (without localization) at well P_{11} underestimates the uncertainty in the model predictions.

Substantial differences in the performance of the Gaussian approximation with respect to the posterior is observed in Table 2. Interestingly, these differences are not always reflected in terms of the model predictions under the posterior. In Fig. 7, we display the distribution of the pressure at wells P_1, P_6, P_{11}, and P_{13} at the final time of $t = 150$ days. The horizontal line corresponds to the value of the pressure at the corresponding location obtained from simulating the model with the posterior mean. In Figs. 8 and 9, we display the probability density estimate for the pressure at $t = 150$ days measured at wells P_1 and P_{11}. From this figures, we confirm that model predictions (pressure) under the posterior is a peaked distribution with small uncertainty for those wells where measurements were collected (e.g., P_1). In contrast, for the new wells (e.g., P_{11}), the pressure under the posterior is visually close to the prior, indicating the uninformative effect of the data for the location of the new wells.

5.2 Oil–water reservoir: small number of wells

In this subsection, we consider the oil–water reservoir model described in Section 2. The reservoir is a square $D = [0, L] \times [0, L]$ with a five-spot well configuration consisting of production wells $P_1, \ldots, P_4$ and one injection well I_1 as displayed in Fig. 10 (middle, well P_5 will play a role in later discussions). For the spatial discretization of Eqs. (5) and (6), we consider cell-centered finite differences and the upwind method. The resulting approximations are then solved implicitly with the backward Euler method as presented in [2, Section 6.2]. Relevant information
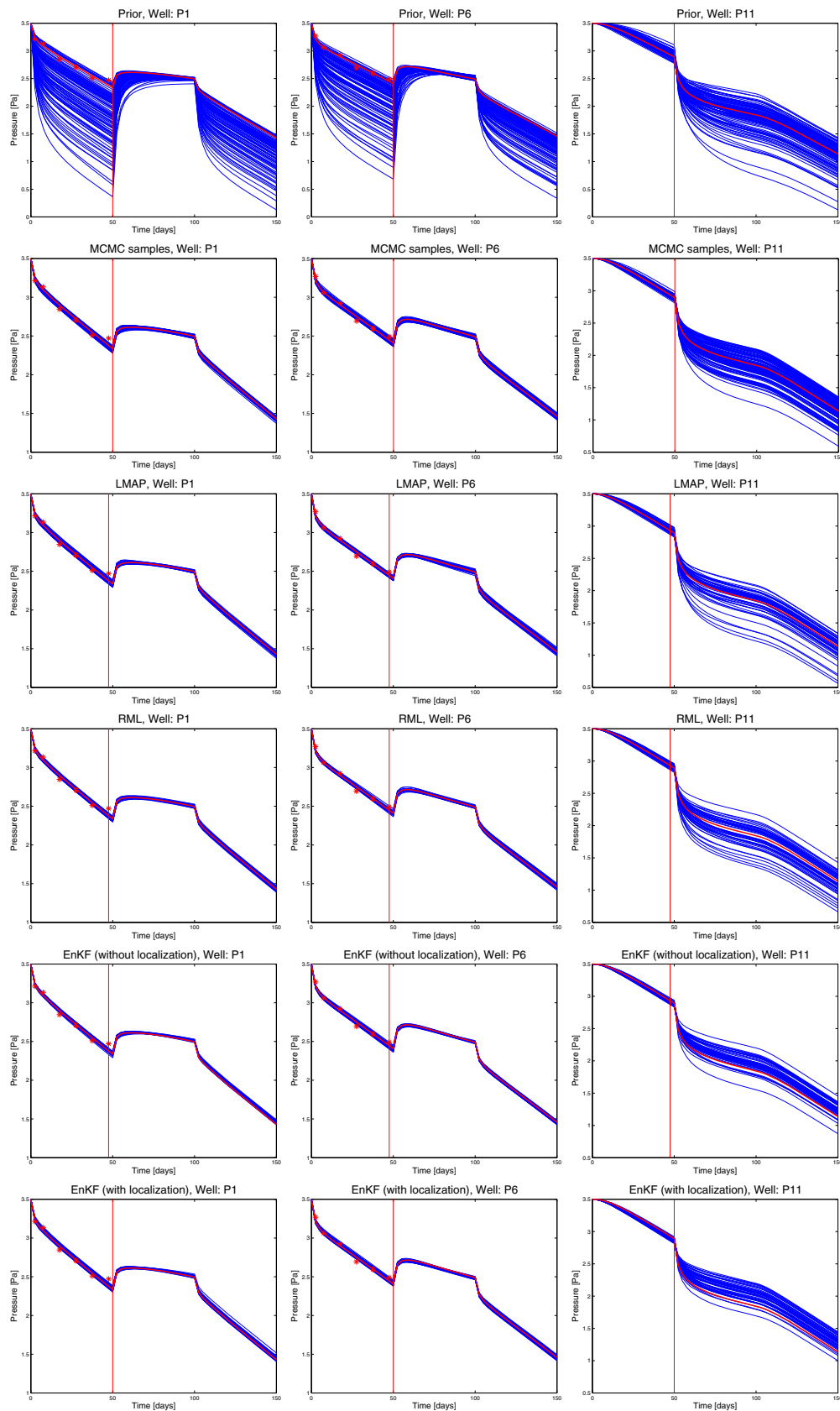
Fig. 6 Single-phase model. Pressure from wells P_{11} (*right column*), P_6 (*middle column*), and P_1 (*left column*) simulated with permeabilities sampled from (*top to bottom rows*) the prior, the posterior, LMAP, RML, EnKF, and EnKF with localization
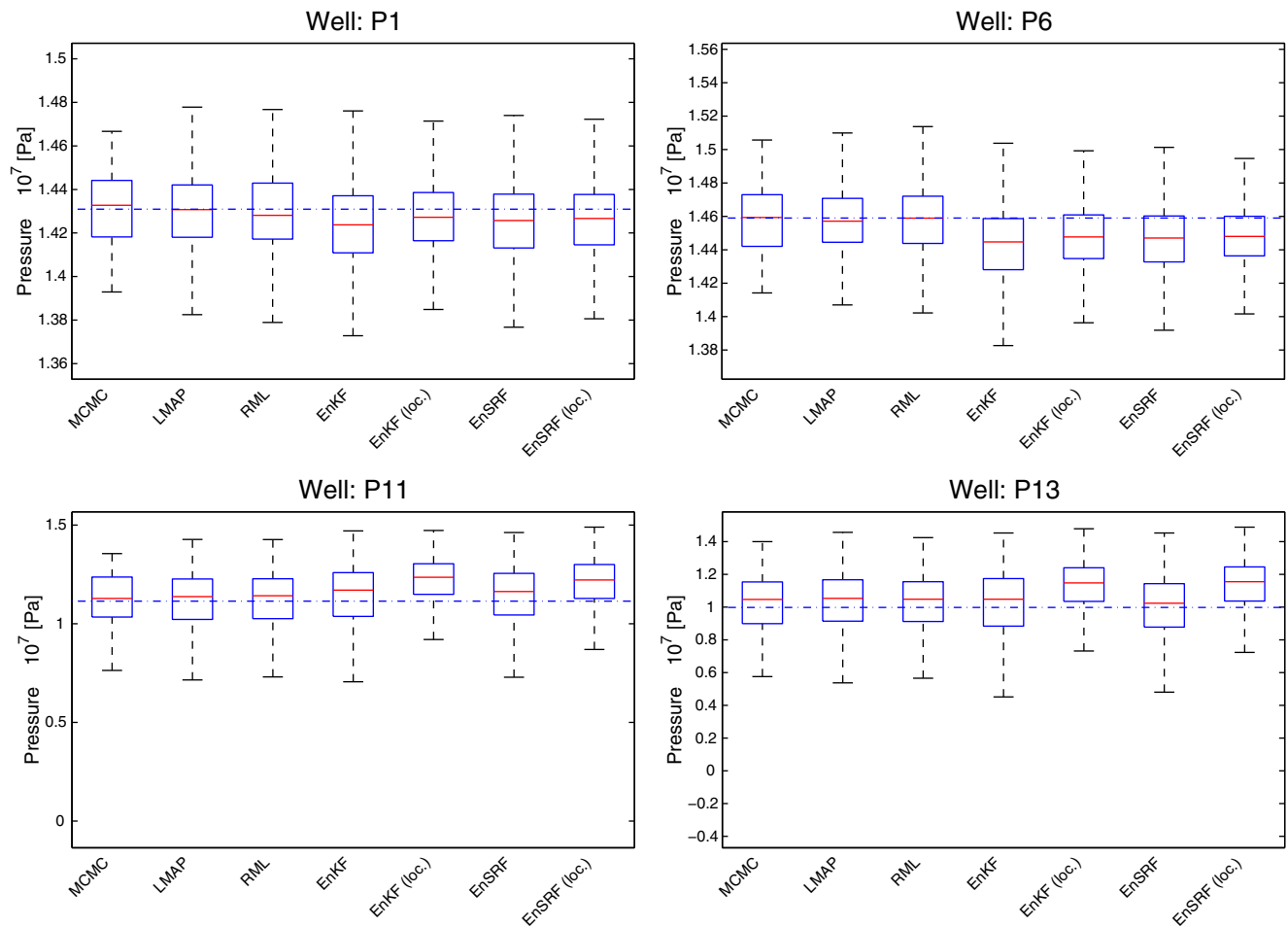
Fig. 7 Single-phase model. Distribution of pressure at wells P_1, P_6, P_{11} and P_{13} at the final simulation time of $t = 150$ days

concerning this model is displayed in Table 1. The reservoir domain and the time interval $[0, T]$ are discretized on a 60×60 grid and 30 time steps, respectively. We recall that the production and injection wells are operated under prescribed bottom-hole pressure and flow rate, respectively. The measured variables for this model are bottom-hole pressure at the injection wells and total flow rate at the production wells. The prior distribution of log permeability is defined by the Gaussian measure defined in Eq. (51) with same parameters except κ which for this experiment is $\kappa = 4.0$. With these parameters, similar realizations to those of Fig. 2 are obtained but with a range of variability with respect to the prior mean $\overline{u}$ increased by a factor of $\sqrt{2}$.

Similarly to the previous example, for the generation of synthetic data, we define the true log permeability $u^\dagger$ displayed in Fig. 10 (left). This true log permeability is generated from the prior distribution described in the preceding paragraph. Note that $u^\dagger$ is the same as the one used for the previous experiment (see Fig. 1, left) but with the magnitude of $u^\dagger - \overline{u}$ multiplied by $\sqrt{2}$. The generation of

synthetic data is now conducted by computing (p, s) from Eqs. (5) and (6) with $u = u^\dagger$. Equation (13) is then used to compute $G(u^\dagger)$, and zero mean Gaussian random error η is added to generate data $y = G(u^\dagger) + \eta$. The measurement times used in Eqs. (10) and (11) are $t_n = 0.67n$ years, $n = \{1, \ldots, 7\}$. According to the structure of Eqs. (12) and (13), η has the following form:

$$\eta = (\eta_1, \ldots, \eta_{N_M}),$$
$$\eta_n = \left(\eta_n^{1,I}, \cdot, \eta_n^{N_I,I}, \eta_n^{1,P}, \ldots, \eta_n^{N_P,P} \right). \tag{54}$$

We generate the components of η as follows:

$$\eta_n^{1,I} \sim N \left(0, \left(3.2 \times 10^4 \text{ Pa} \right)^2 \right)$$

$$\eta_n^{1,P} \sim N \left(0, \left(0.25 \text{ m}^3/\text{day} \right)^2 \right)$$

$$\eta_m^{2,P}, \eta_n^{3,P}, \eta_n^{4,P} \sim N \left(0, \left(0.02 \text{ m}^3/\text{day} \right)^2 \right). \tag{55}$$
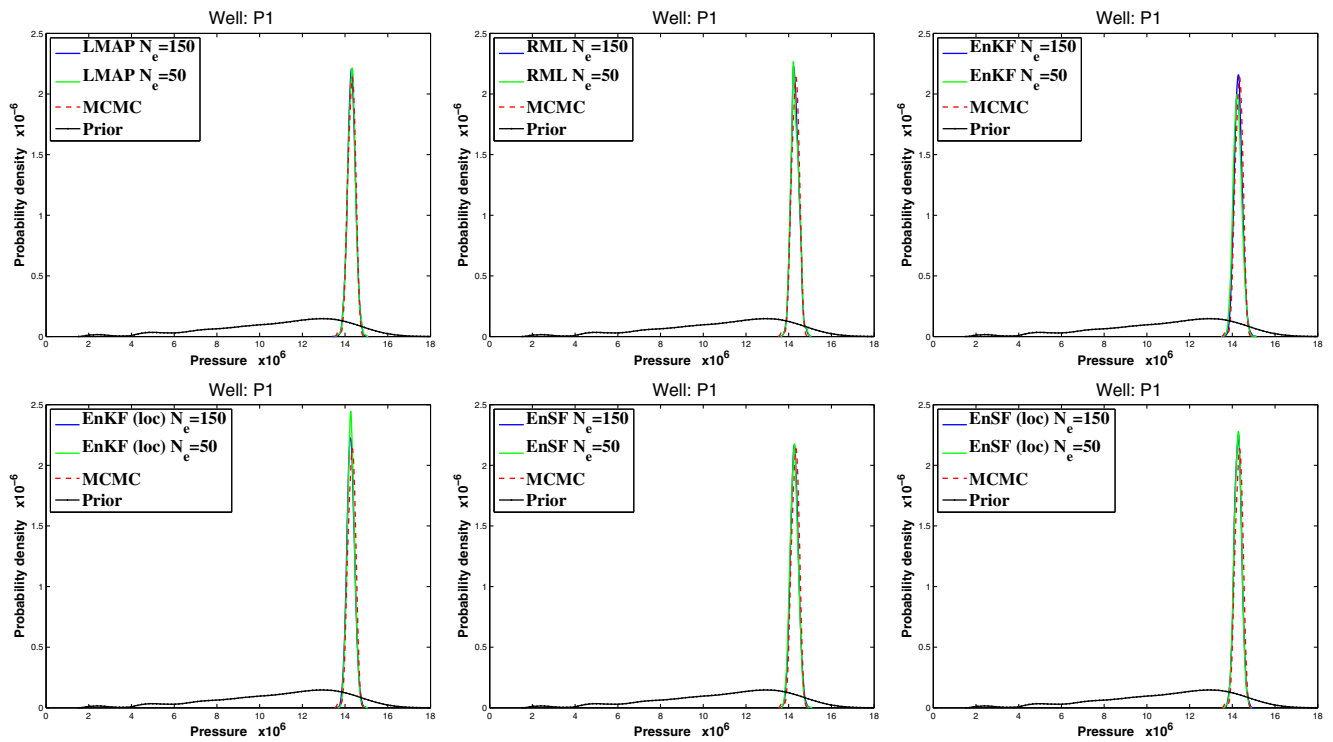
Fig. 8 Single-phase model. Probability density estimate of pressure at well P_1 at the final simulation time of $t = 150$ days, computed with different Gaussian approximations
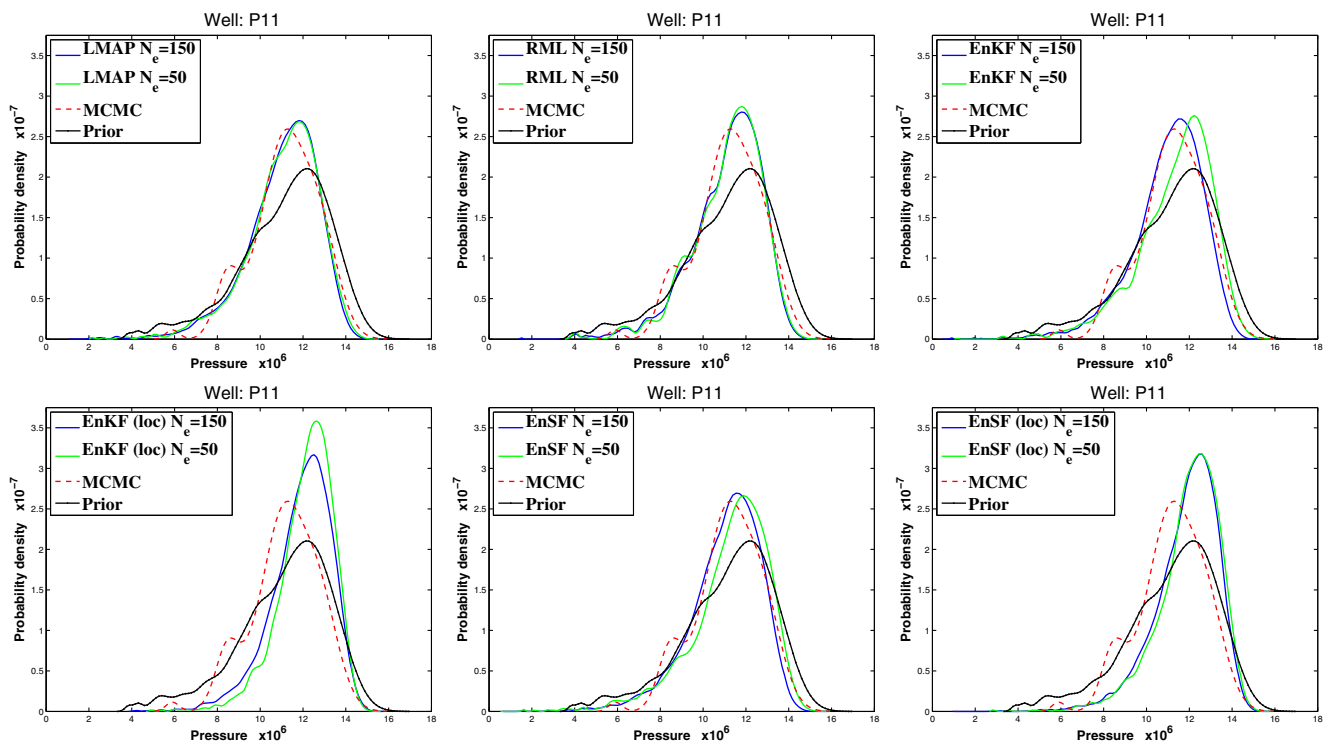

Fig. 9 Single-phase model. Probability density estimate of pressure at well P_{11} at the final simulation time of $t = 150$ days, computed with different Gaussian approximations
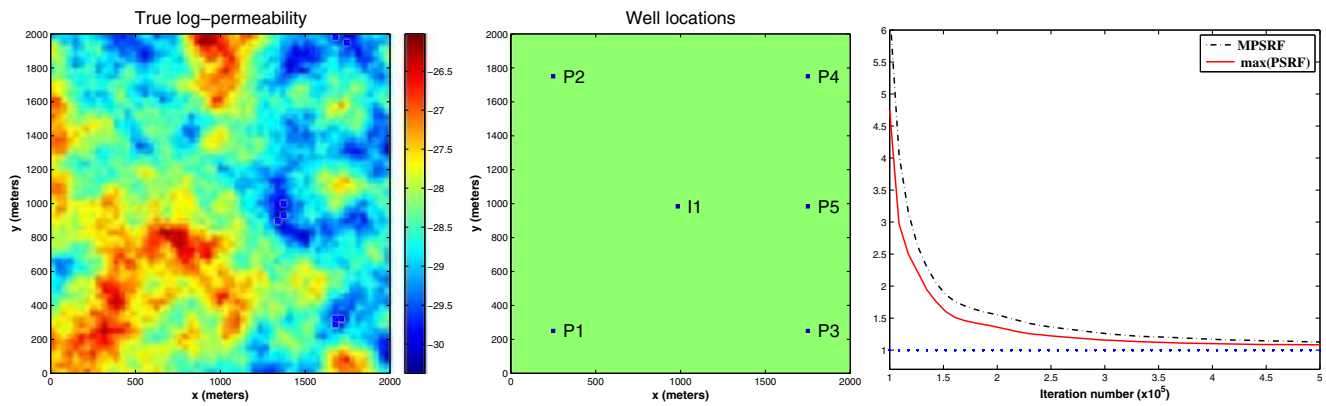
Fig. 10 Two-phase model (small number of wells). *Left*: True log permeability [log m^2]. *Middle*: Well configuration. *Right*: Gelman–Rubin diagnostic

for all $n \in \{1, \ldots, 7\}$. In the previous definitions, we consider larger measurement error at the production well P_1 since larger total flow rates are obtained. This is caused by the larger permeability region around P_1 and so the early water breakthrough at this well.

The pCN-MCMC Algorithm 1 with $\beta = 0.015$ is applied to this problem with the synthetic data previously described. Starting from independent draws from the prior distribution, 110 chains are generated. A burn-in period of 1.5×10^4 was observed after which the chains were run 5×10^5 iterations. By using the same Gelman–Rubin diagnostic indicated in the previous example, we determine convergence of our chains. This numerical evidence is presented in Fig. 10 where, after 5×10^5 iterations, both the maximum PSRF and MPSRF for the highest energy models converge to 1. Some

uncorrelated samples (from independent chains) are shown in Fig. 11.

The 5.5×10^7 samples of the posterior generated with MCMC define our gold standard. These are then used to compare the performance of Gaussian approximations in terms of mean and variance. Analogous to the previous example, we use two sets of experiments with ensemble sizes of $N_e = 50$ and $N_e = 150$. Then we compute the mean and variance with LMAP, RML, EnKF, EnKF with localization, EnSRF and EnSRF with localization, for 50 experiments corresponding to 50 different choices of the initial ensemble. The averaged relative errors of the mean and variance (see expression (53)) with respect to the posterior distribution are provided in Table 3. For one of the 50 experiments with $N_e = 50$,

Fig. 11 Two-phase model (small number of wells). Samples from the posterior distribution (characterized with MCMC) [log m^2]
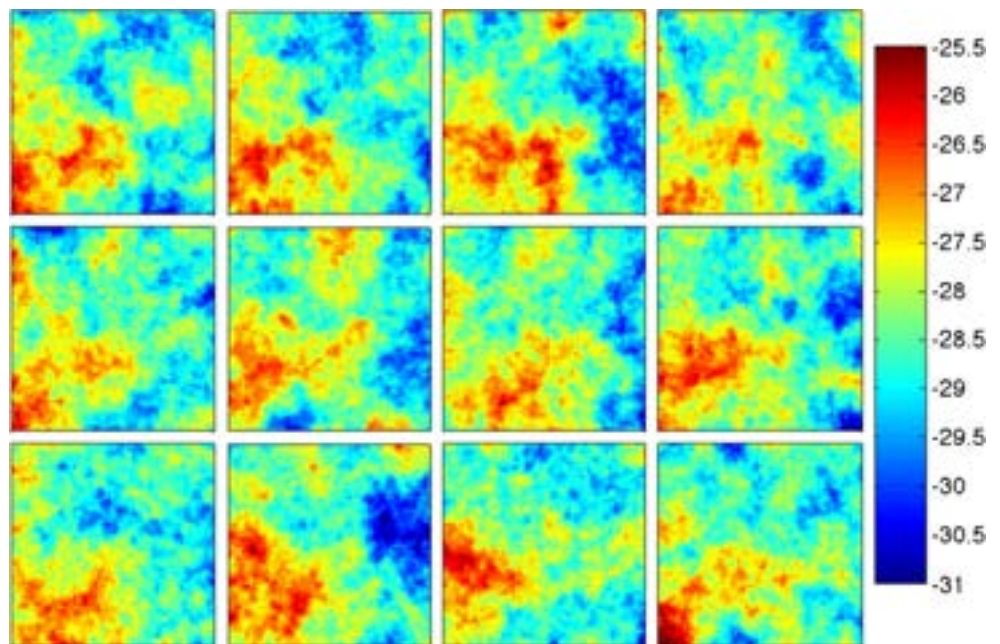
Table 3 Evaluation of Gaussian approximations for the two-phase model (the case with small number of wells)

Method	Relative error in the mean ϵ_u	Relative error in the variance ϵ_σ	Computational cost (forward model runs)
MCMC	0.000	0.000	5.5×10^7
MAP	0.277	0.143	6.0×10^1
LMAP ($N_e = 50$)	0.316	0.271	6.0×10^1
RML ($N_e = 50$)	0.235	0.415	3.0×10^3
EnKF ($N_e = 50$)	0.834	0.391	5.0×10^1
EnKF (localization, $N_e = 50$)	0.563	0.252	5.0×10^1
EnSRF ($N_e = 50$)	0.701	0.354	5.0×10^1
EnSRF (localization, $N_e = 50$)	0.508	0.234	5.0×10^1
LMAP ($N_e = 150$)	0.290	0.191	6.0×10^1
RML ($N_e = 150$)	0.197	0.351	9.0×10^3
EnKF ($N_e = 150$)	0.571	0.151	1.5×10^2
EnKF (localization, $N_e = 150$)	0.447	0.179	1.5×10^2
EnSRF ($N_e = 150$)	0.478	0.150	1.5×10^2
EnSRF (localization, $N_e = 150$)	0.397	0.210	1.5×10^2
EnKF ($N_e = 1,000$)	0.353	0.209	1.0×10^3
EnKF ($N_e = 3,000$)	0.301	0.222	3.0×10^3
EnKF ($N_e = 8,000$)	0.337	0.216	8.0×10^3

Fig. 12 Two-phase model (small number of wells). Mean of the posterior distribution (characterized with MCMC) and Gaussian approximations [log m^2]
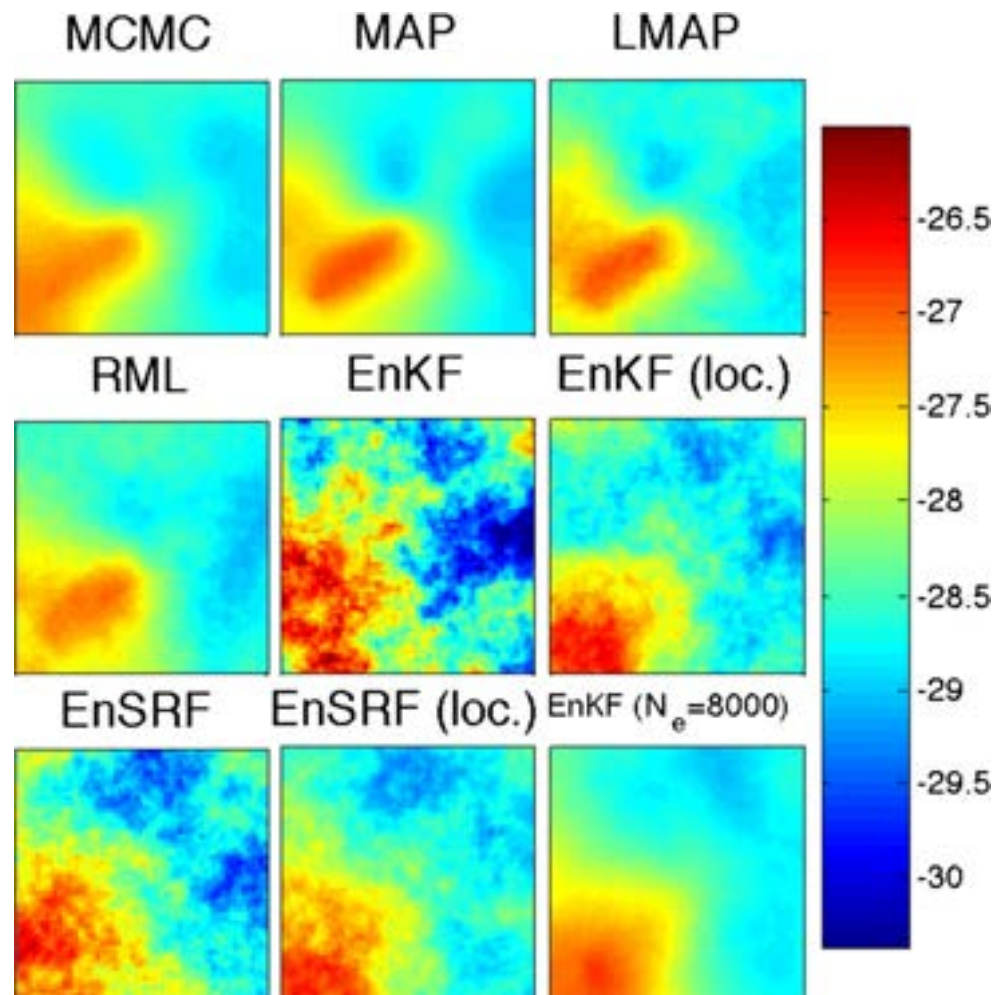
Fig. 13 Two-phase model (small number of wells). Variance of the posterior distribution (characterized with MCMC) and Gaussian approximations $[(\log m^2)^2]$
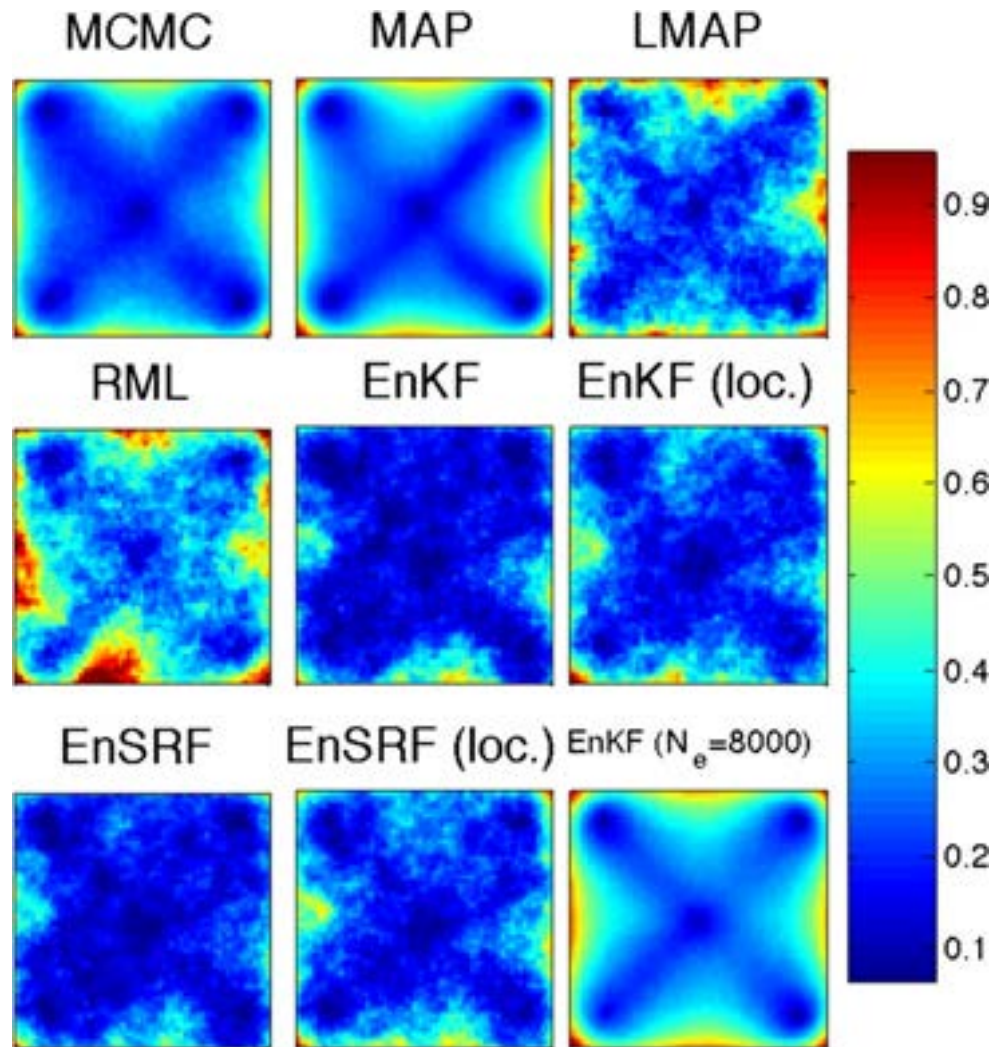
in Figs. 12 and 13 we present the mean and variance, respectively.

In contrast to the previous experiment, the error in the approximation given by $N(u_{\mathrm{MAP}}, C_{\mathrm{MAP}})$ is relatively large. The relative errors of mean and variance are 27 and 14 %, respectively. Similar to the previous experiment, RML provided the best approximation of the posterior in terms of the mean. However, the variance of the posterior was significantly overestimated by RML. The performance of the standard EnKF for $N_{\mathrm{en}} = 50$ was very poor. From Fig. 12, we observe large values of the estimated field, which are typically found in standard EnKF applications with small ensemble sizes. Nevertheless, for the same size, covariance localization has a positive effect by reducing the error in the mean and variance with respect to the posterior. Similar to the previous experiment, EnKF and EnKF with localization were outperformed by the corresponding EnSRF implementations. As in the previous experiment, EnSRF with localization provided, among the

techniques with $N_{\mathrm{e}} = 50$, the best approximation in terms of variance.

Unlike the preceding experiment, here we observe that increasing the size of the ensemble does not result in a decrease of the error with respect to the mean. This can be observed at the end part of Table 3 where we report the results of EnKF implementations for $N_{\mathrm{en}} = 1,000$, $N_{\mathrm{en}} = 3,000$, and $N_{\mathrm{en}} = 8,000$. Both in terms of mean and variance, EnKF for large ensembles does not exhibit convergence to the posterior. Note that EnKF with $N_{\mathrm{en}} = 3,000$ corresponds to the same computational cost of RML. Yet the latter provides a better approximation in terms of the mean. Among all the techniques, LMAP provides reasonable approximations of both the mean and variance of the posterior.

In order to assess the performance of the approximate posterior samples at reproducing the predicting distribution, we consider an additional simulation period of 5 years of forecast. For this additional 5 years, a new well labeled as
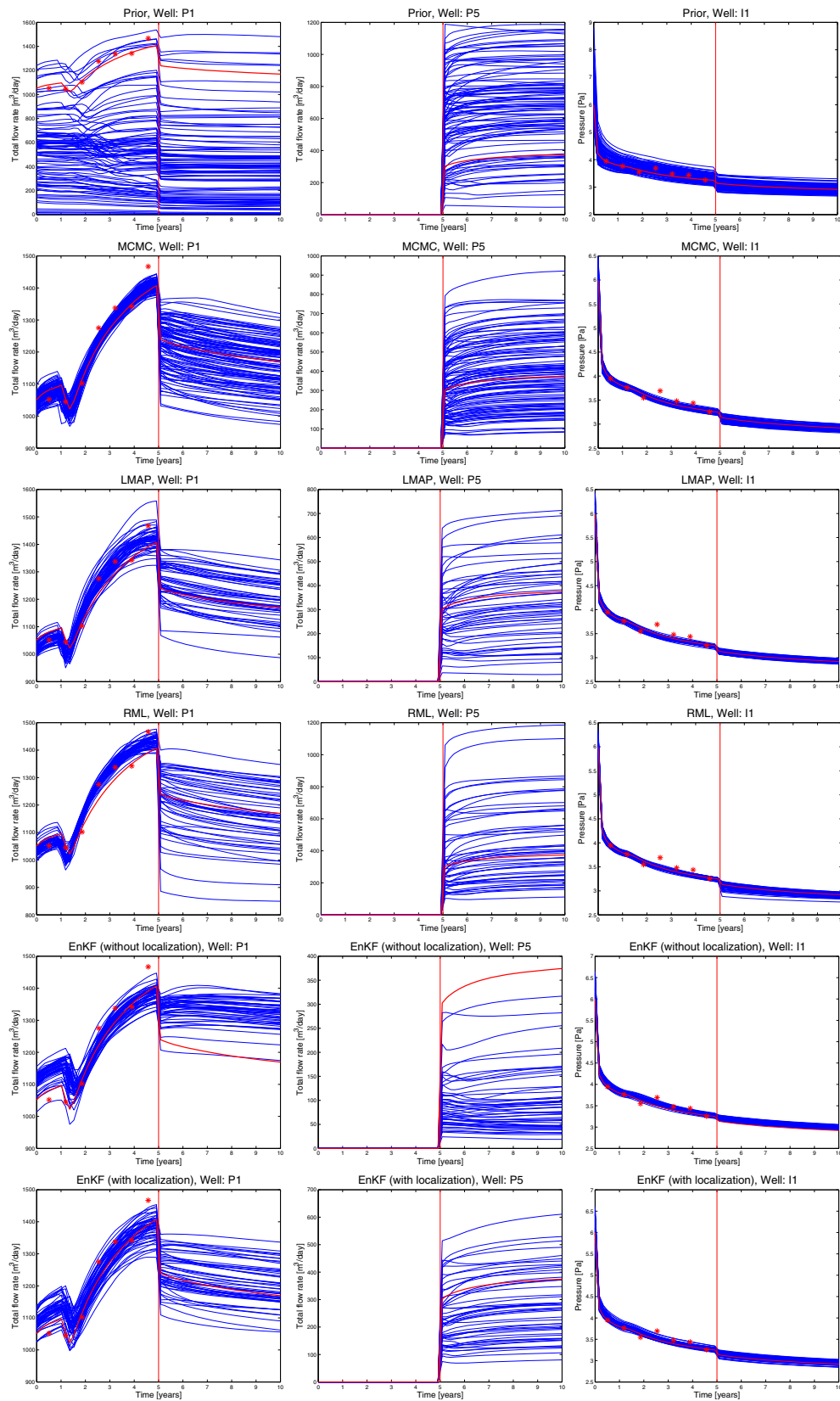
Fig. 14 Two-phase model (small number of wells). Total flow rates from P_1 (*left column*), P_5 (*middle column*), and bottom-hole pressure from I_1 (*right column*) simulated with permeabilities sampled from (*top to bottom rows*) the prior, the posterior, LMAP, RML, EnKF, and EnKF with localization
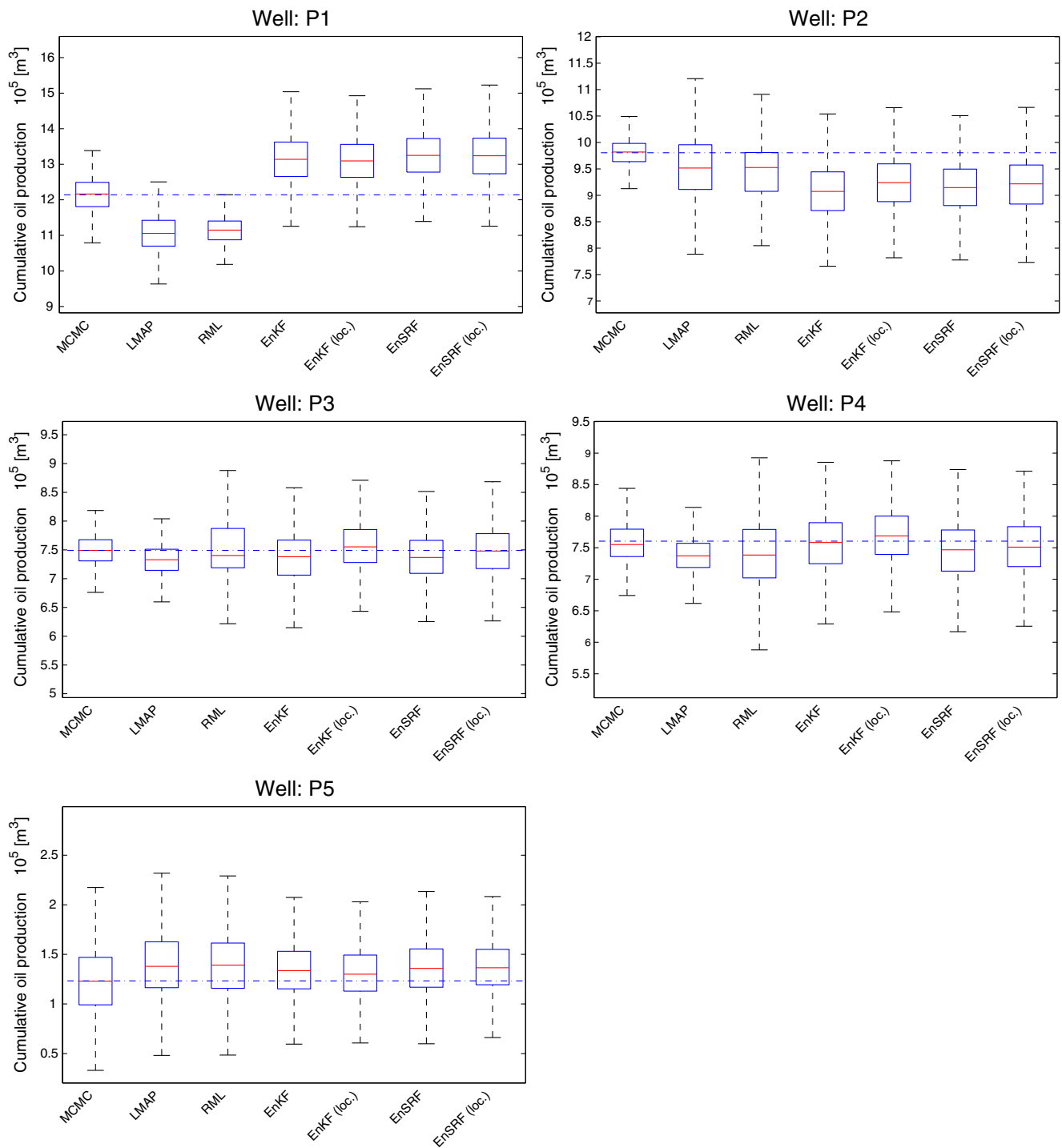
Fig. 15 Two-phase model (small number of wells). Distribution of cumulative oil production at wells $P_1, \ldots, P_5$ at the final simulation time of $t = 10$ years

P_5 in Fig. 10 (middle) is drilled and operated under constant fixed bottom-hole pressure of $P^5_{bh} = 2.7 \times 10^7$ Pa. In Fig. 14, we present the total flow rates (from P_1 and P_5) and bottom-hole pressure (from I_1) simulated with permeabilities from the prior (first row), the posterior (second row), and some of

the Gaussian approximations under analysis (third to sixth row). The vertical line divides the assimilation from the prediction. The red curve is computed from the posterior mean at the corresponding location. We note that the poor performance of EnKF is reflected in the model predictions
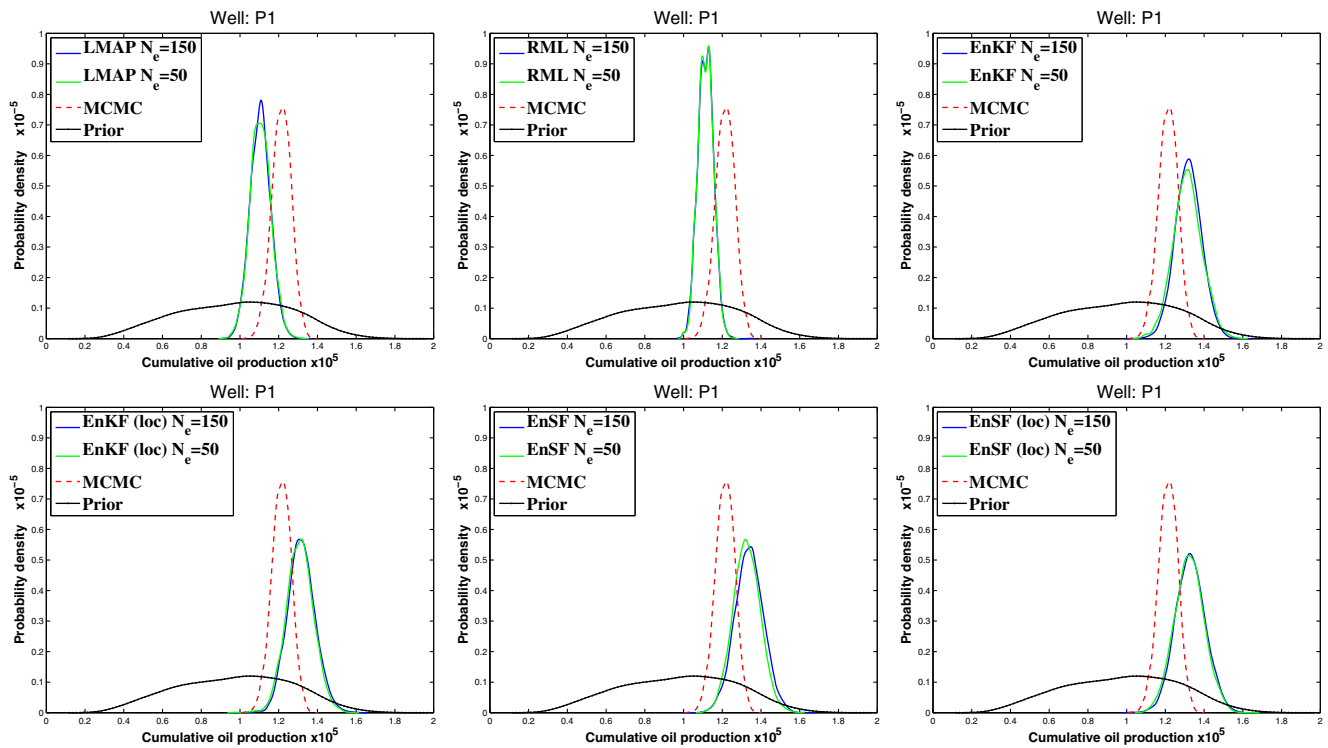
Fig. 16 Two-phase model (small number of wells). Probability density estimate of cumulative oil production at well P_1 at the final simulation time of $t = 10$ years
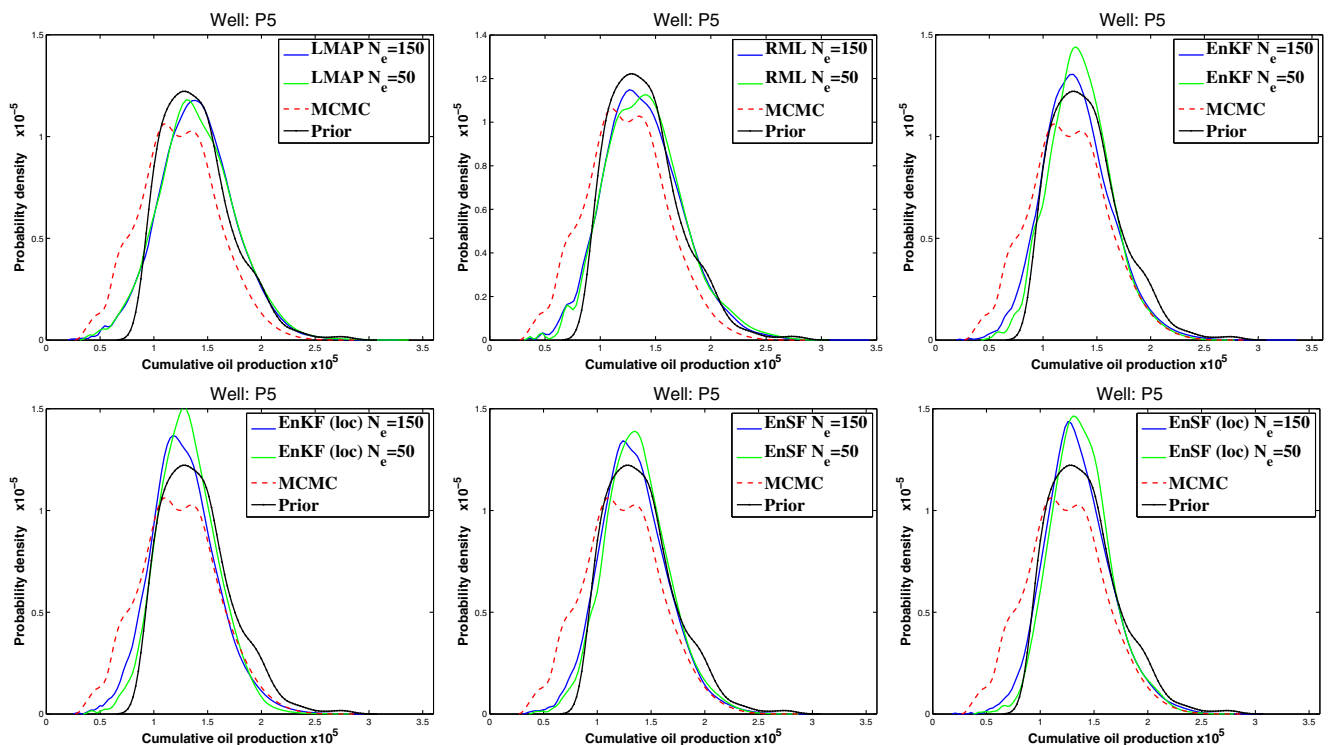
Fig. 17 Two-phase model (small number of wells). Probability density estimate of cumulative oil production at well P_5 at the final simulation time of $t = 10$ years
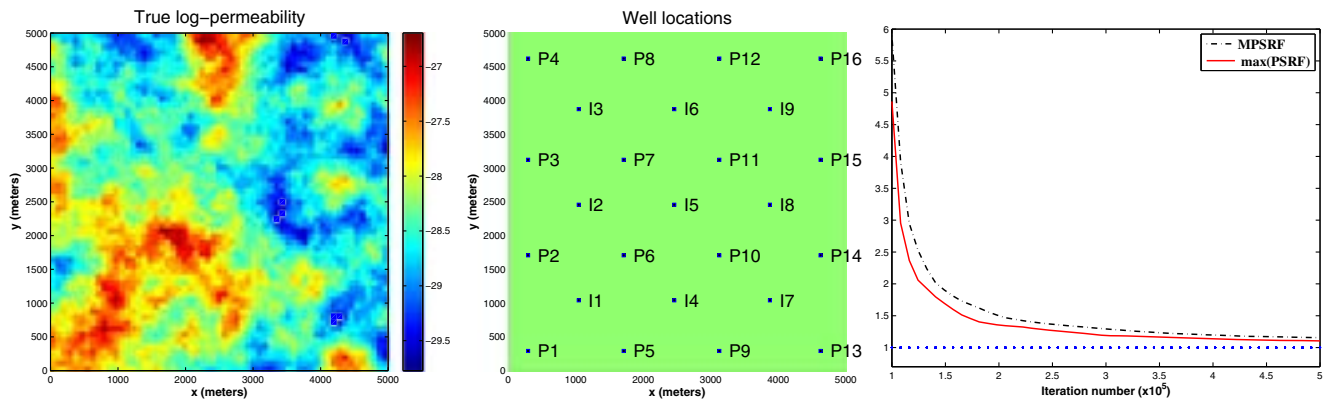
Fig. 18 Two-phase model (large number of wells). *Left*: True log permeability [log m^2]. *Middle*: Well configuration. *Right*: Gelman-Rubin diagnostic

whose ensemble does not capture the prediction based on the mean of the posterior. These issues are alleviated by localization as what we have observed in Fig. 14 (last row).

In Fig. 15, we display the distribution of the final time cumulative oil production simulated from the posterior and the Gaussian approximation. Note that even though LMAP provides a reasonable approximation in terms of both mean and variance, the approximation provides a deficient characterization of the predicting distribution. In general, all the Gaussian approximations exhibit poor performance at reproducing the predicting distribution. In Figs. 16 and 17, we display the probability density estimates of cumulative oil production at the final time of simulation at wells P_1 and P_5. Similar to the experiment in the previous

section, for the model predictions under the posterior, a peaked distribution with small variance is obtained at the wells where measurements were collected. In contrast to the previous experiment, however, the Gaussian approximations performed poorly at characterizing the distribution of the model under the posterior.

5.3 Oil–water reservoir: large number of wells

In this subsection, we consider again the oil–water reservoir model described in Section 2. The well configuration for this case is displayed in Fig. 22 (middle), and relevant information can be found in Table 1. The aim of this experiment was to evaluate the performance of Gaussian

Fig. 19 Two-phase model (small number of wells). Samples from the posterior distribution (characterized with MCMC) [log m^2]
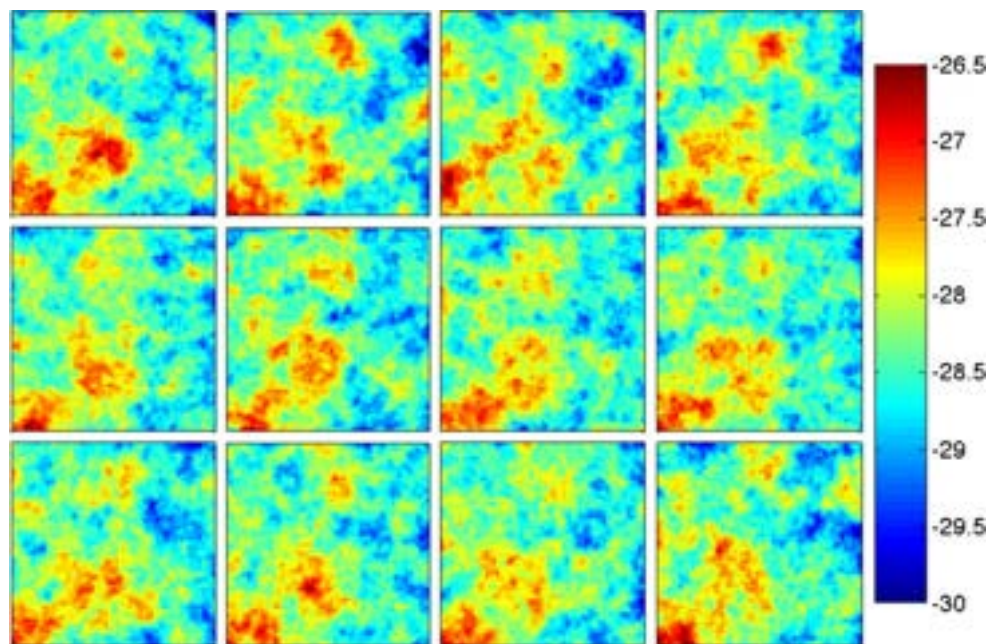
Table 4 Evaluation of Gaussian approximations for the two-phase model (the case with large number of wells)

Method	Relative error in the mean ϵ_u	Relative error in the variance ϵ_σ	Computational cost (forward model runs)
MCMC	0.000	0.000	5.5×10^7
MAP	0.131	0.165	3.5×10^2
LMAP ($N_e = 50$)	0.185	0.286	3.5×10^2
RML ($N_e = 50$)	0.175	0.299	1.85×10^4
EnKF ($N_e = 50$)	0.910	0.796	5.0×10^1
EnKF (localization, $N_e = 50$)	0.543	0.621	5.0×10^1
EnSRF ($N_e = 50$)	0.818	0.643	5.0×10^1
EnSRF (localization, $N_e = 50$)	0.517	0.471	5.0×10^1
LMAP ($N_e = 250$)	0.143	0.195	3.5×10^2
RML ($N_e = 250$)	0.130	0.215	9.25×10^4
EnKF ($N_e = 250$)	0.432	0.171	2.5×10^2
EnKF (localization, $N_e = 250$)	0.296	0.119	2.5×10^1
EnSRF ($N_e = 250$)	0.347	0.115	2.5×10^2
EnSRF (localization, $N_e = 250$)	0.255	0.102	2.5×10^2
EnKF ($N_e = 1,000$)	0.243	0.101	1.0×10^3
EnKF ($N_e = 3,000$)	0.161	0.137	3.0×10^3
EnKF ($N_e = 6,000$)	0.127	0.148	6.0×10^3
EnKF (large $N_e = 18,500$)	0.111	0.154	1.85×10^4

approximations when measurements from many wells are available. The prior distribution of log permeability is defined by the Gaussian measure defined in Eq. (51) with the same tunable parameters used in the experiment in Section 5.1.

Generation of synthetic data was conducted with the same procedure described before. The true log permeability $u^\dagger$ is displayed in Fig. 18 (left). The random error η added to $G(u^\dagger)$ has the form of Eq. (54) with $N_I = 9$ and $N_P = 16$ and

$$\eta_n^{j,I} \sim N\left(0, 2.7 \times 10^4 \,\mathrm{Pa}\right), \qquad j \in \{1, \ldots, N_I\},$$

$$\eta_n^{j,P} \sim N\left(0, 0.06 \,\mathrm{m^3/day}\right), \qquad j \in \{1, \ldots, N_P\} \quad (56)$$

for all $n \in \{1, \ldots, 7\}$. Measurement times are $t_n = 0.467n$ years, $n = \{1, \ldots, 7\}$. pCN-MCMC Algorithm 1 is applied to generate 110 chains starting from independent draws from the prior distribution. After a burn-in period of 1.5×10^4, the chains are run for 5×10^5 iterations. The Gelman-Rubin diagnostic is conducted as described in the preceding sections. Figure 18 (right) shows the PSRF and MPSRF as defined previously. Uncorrelated samples (from independent chains) are shown in Fig. 19.

Similar to the previous section, 5.5×10^7 samples of the posterior generated with our converged chains provide the gold standard against which the performance of Gaussian approximations in terms of mean and variance is compared.

The first part of Table 4 provides the results when an ensemble of size $N_e = 50$ is used. These results, similar to the ones from the previous sections, are averaged over 50 experiments with different initial ensembles. In Figs. 20 and 21, we display the mean and variance, respectively.

Among all the ensemble methods with $N_e = 50$, RML provides the best approximation in terms of the mean. Note that the approximation provided by $N(u_{MAP}, C_{MAP})$ provides the best approximation in terms of combined mean and variance. Additionally, even with localization, both EnKF and EnSRF provide a very poor approximation in terms of mean and variance. It is worth mentioning that RML and LMAP provided a better approximation (in terms of mean and variance) than the ensemble Kalman filter-type methods for $N_e = 50$. In Fig. 22, we show the total flow rates (from P_1 and P_5) and bottom-hole pressure (from I_1) simulated with permeabilities from the prior (first row), the posterior (second row), and some of the Gaussian approximations under analysis (third to sixth row). The vertical line divides the assimilation from the prediction. In this case, prediction is performed by simulating an additional 3.5 years under the same well configuration. The red curve is computed from the posterior mean at the corresponding location. In Fig. 23, we display the distribution of the final time cumulative oil production simulated from the posterior and the Gaussian approximation. The poor performance of EnKF and

Fig. 20 Two-phase model (large number of wells). Mean of the posterior distribution (characterized with MCMC) and Gaussian approximations [log m^2]
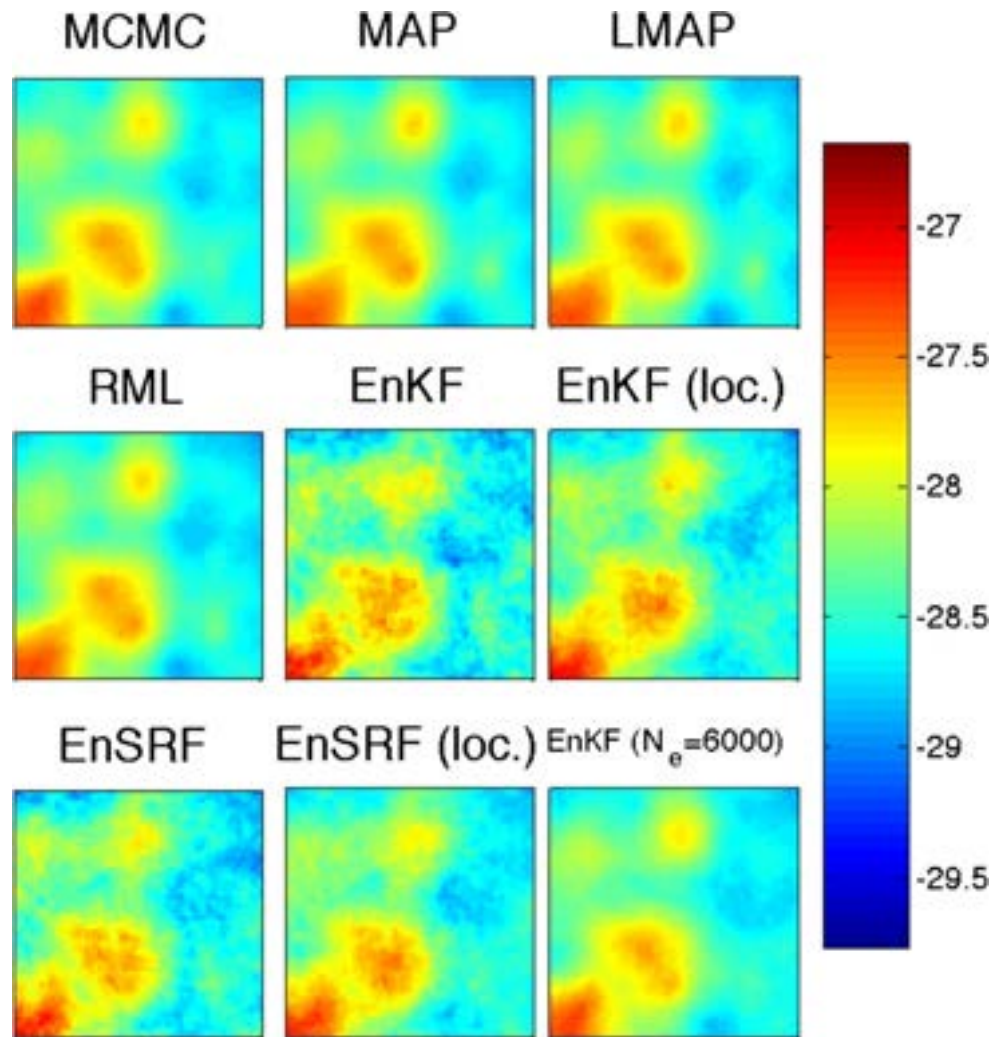
EnSRF is reflected in the poor performance at characterizing the predicting distribution. In Figs. 24 and 25, we display the probability density estimate of the cumulative oil production, at the final simulation time, at wells P_1 and P_{12}. Measurements for all wells have been collected. Therefore, substantial differences between the posterior and the prior are reflected in the model predictions under these distributions.

As we mentioned earlier, limitations of the EnKF and EnSRF arise when a large number of measurements are assimilated. In the present work, we are interested in the associated detrimental effect on the approximation of the posterior distribution. In order to observe that effect, we now consider application of our Gaussian approximation on a larger ensemble $N_e = 250$. These results are presented in the second part of Table 4. For $N_{en} = 250$, Table 4 indicates that RML provides again the best approximation in terms of the mean. The performance of EnKF, EnSRF, and their localizations is considerably improved with respect to the ones for $N_e = 50$. As in previous examples, EnSRF

with localization provides the best approximation in terms of variance. Also, similarly to the previous experiments, increasing the size of the standard EnKF does not improve the approximation in terms of variance. On the other hand, in this case, the variance increases with the size of ensemble. This can be observed from the last part of Table 4 where EnKF was implemented for $N_{en} = 1,000$, $N_{en} = 3,000$, $N_{en} = 6,000$, and $N_{en} = 18,500$. Note that the computational cost of EnKF for $N_{en} = 18,500$ coincides with the cost of our implementation of RML with $N_{en} = 50$.

6 Conclusions

The controlled experiments from the preceding sections enable us to numerically assess the performance of widely used ad hoc Gaussian approximations with respect to their ability to correctly reproduce the mean and variance under the true Bayesian posterior distribution. The true posterior is obtained by use of the expensive, but accurate,

Fig. 21 Two-phase model (large number of wells). Variance of the posterior distribution (characterized with MCMC) and Gaussian approximations [$(\log m^2)^2$]
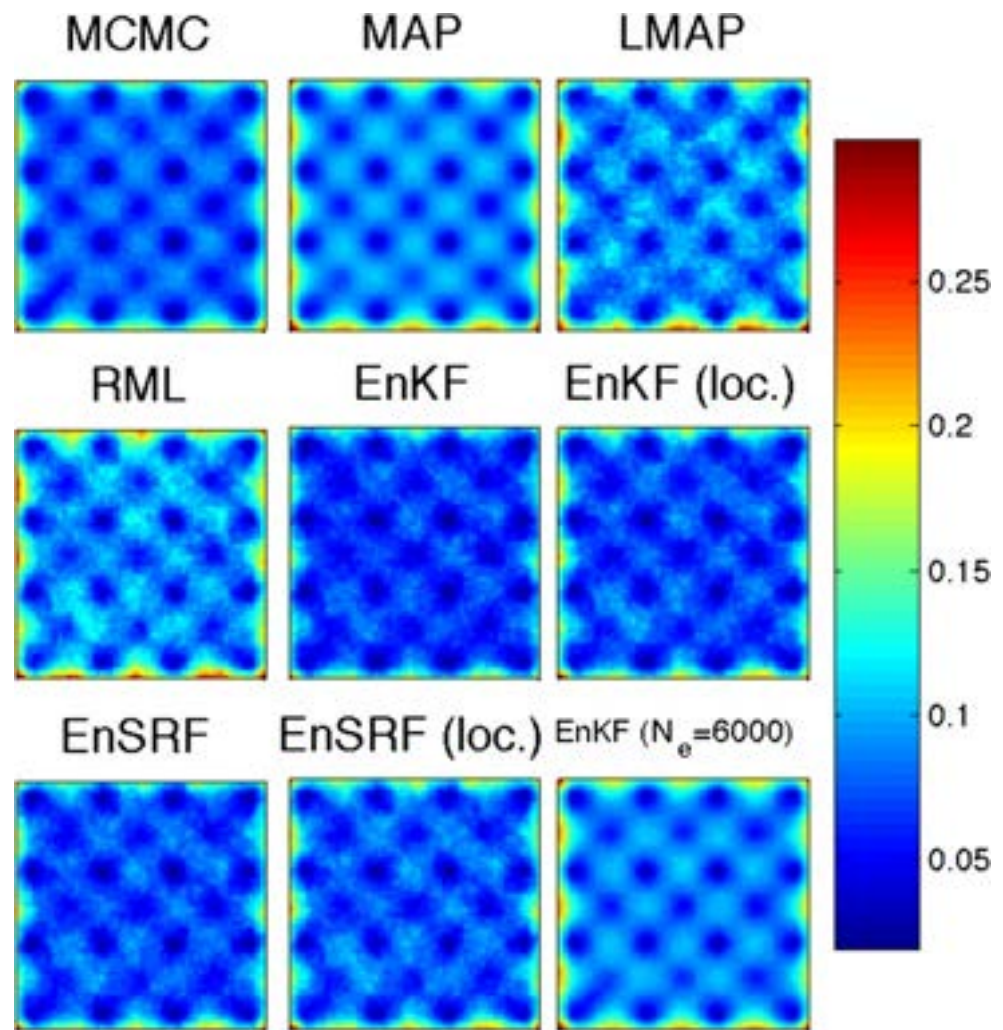
gold standard pCN-MCMC method. The forward operators associated to the reservoir models under consideration are nonlinear. Therefore, even though a Gaussian prior is considered, the posterior distribution associated to each experiment is non-Gaussian. Indeed, for all our experiments, a significant discrepancy between the MAP estimate and the mean of the posterior distribution was obtained, giving clear numerical evidence that the posteriors for the Bayesian data assimilation problems under consideration are not Gaussian. Even in the single-phase reservoir whose associated forward operator is "less" nonlinear (the PDE (Eq. (1) is linear with respect to p), we find that the relative error in the MAP estimator with respect to the mean of the posterior was 3 %. For the two-phase reservoirs, this relative error was 27 and 13 % in the case of small and large number of wells, respectively. Thus, the problems that we study demonstrate a range of deviations from Gaussianity in the posterior. This makes them a suitable range of test problems for the ad hoc algorithms, all of which can be systematically derived in the linear Gaussian scenario, but

whose accuracy in the non-Gaussian case is, in general, unclear.

We clearly observe substantial differences in the approximation properties of the posterior distribution with respect to the choice of method, reservoir model, and well configuration. For all the experiments conducted here, we conclude that, among all the Gaussian approximations under analysis, LMAP is arguably the best technique at reproducing the posterior distribution in terms of combined variance and mean. It is interesting to speculate why this might be so and what it tells us about the posterior. The LMAP algorithm is the only Gaussian approximation which samples from $N(u_{MAP}, C_{MAP})$ where u_{MAP} is the MAP estimate and C_{MAP} is the associated covariance matrix defined by Eqs. (20) and (21), respectively. This suggests that, out of all the Gaussian approximations considered, the posterior distribution in all our examples can be best approximated, in terms of mean and variance, by $N(u_{MAP}, C_{MAP})$. We emphasize that this does not imply that the posterior distribution is near Gaussian since the errors of this best

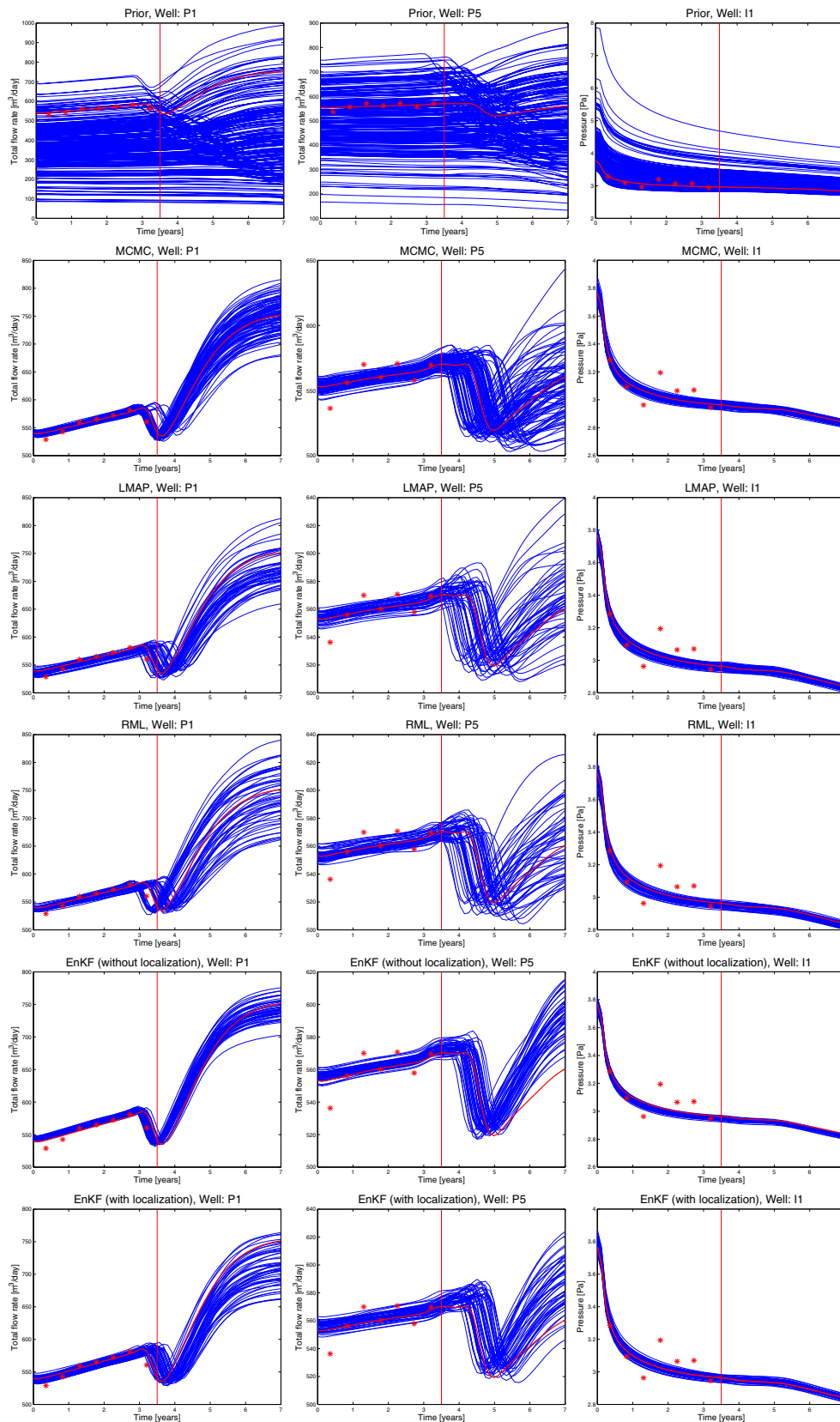**Fig. 22** Two-phase model (large number of wells). Total flow rates from P_1 (*left column*), P_5 (*middle column*), and bottom-hole pressure from I_1 (*right column*) simulated with permeabilities sampled from (*top to bottom rows*) the prior, the posterior, LMAP, RML, EnKF, and EnKF with localization
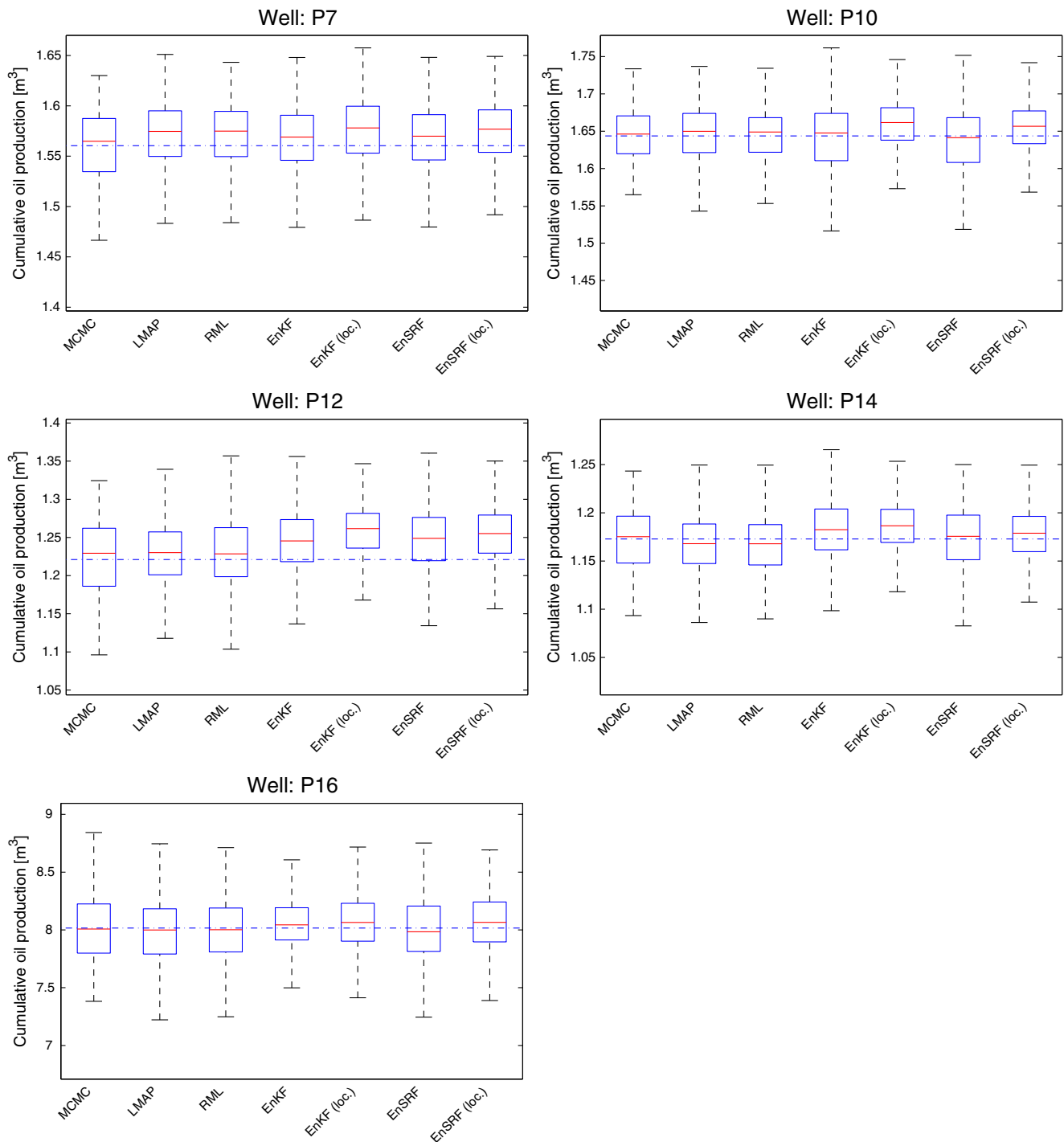
Fig. 23 Two-phase model (large number of wells). Distribution of cumulative oil production at wells P_7, P_{10}, P_{12}, P_{14}, and P_{16} at the final simulation time of $t = 10$ years

approximation are still large, i.e., the best approximation is still not a good approximation in terms of mean and variance.

We recall that all the techniques described in Section 4 produce samples of the posterior distribution in the linear Gaussian case. In other words, they sample from the exact posterior distribution $N(u_{\mathrm{MAP}}, C_{\mathrm{MAP}})$. In our experiments, however, we observe clear differences in the approximations obtained with each of the techniques under consideration. For example, note from all our experiments that the RML provided the best approximation of the posterior in terms of the mean. In addition, in the case of the single-phase
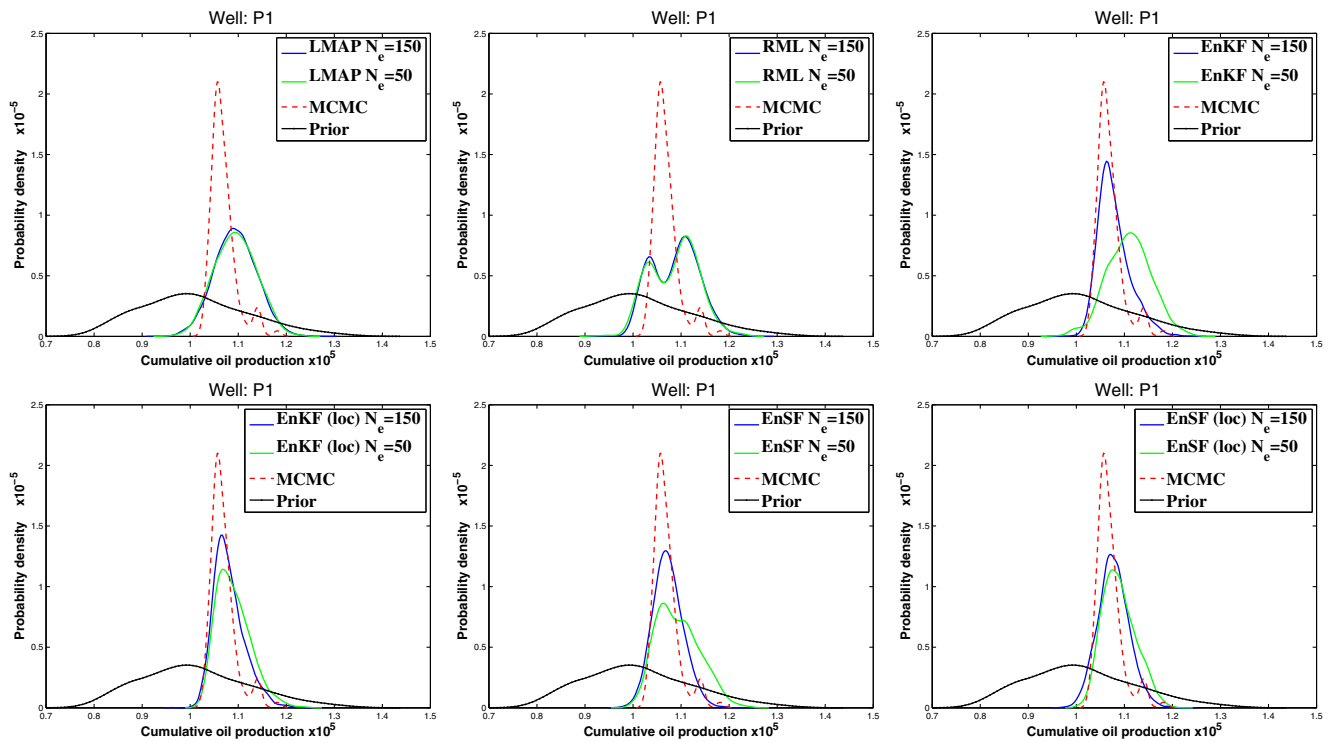
Fig. 24 Two-phase model (small number of wells). Distribution of cumulative oil production at well P_1 at the final simulation time of $t = 10$ years

model, RML provides a reasonable approximation of the posterior variance (like the one obtained with LMAP). It is worth mentioning that favorable RML results for

single-phase reservoirs are also reported in [23]. In contrast, for the two-phase model, we find examples where the error of the RML variance is the largest compared
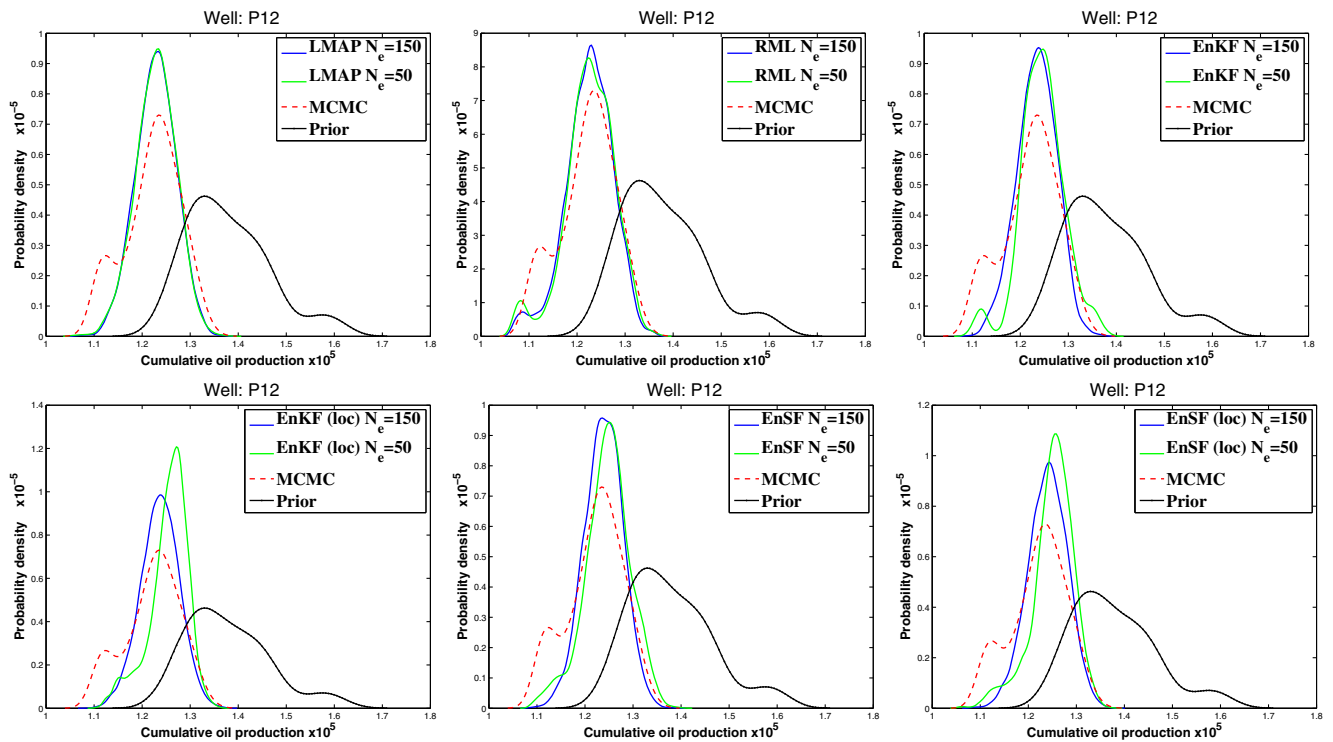

Fig. 25 Two-phase model (small number of wells). Distribution of cumulative oil production at well P_{12} at the final simulation time of $t = 10$ years

to other Gaussian approximations. These observations are likely to be related to the higher nonlinearity in the two-phase forward operator (see Eq. (13)) that results from the nonlinear PDE system (see Eqs. (5) and (6)). Due to the aforementioned higher nonlinearity, large changes in the absolute values of the log permeability field may not necessarily correspond to large changes in the production data. In fact, production data may typically have smaller sensitivity to the permeability values far from (or in between) the well locations. On the other hand, we recall from the RML algorithm that each ensemble member $u_{\text{RML}}^{(j)}$ (see Eq. (23)) produces a model output $G(u_{\text{RML}}^{(j)})$ that is close to the perturbed data $y^{(j)}$ while keeping $u_{\text{RML}}^{(j)}$ close to the corresponding sample from the prior $u^{(j)}$. Due to the aforementioned small sensitivity of production data to the log permeability in some regions of the domain, it may be possible that the penalty term $||u_{\text{RML}}^{(j)} - u^{(j)}||$ in Eq. (23) may not provide sufficient constraint to avoid possible large values of $|u_{\text{RML}}^{(j)}|$ in the aforementioned regions for which the (perturbed) production data are minimally affected by the large value of permeability. Although in our controlled experiment LMAP outperformed the RML in terms of combined mean and variance, it has been reported that RML has the advantage of approximating multimodal distribution for which $N(u_{\text{MAP}}, C_{\text{MAP}})$, and therefore LMAP, is suboptimal. The assessment of techniques where multimodal posterior distribution arises deserves further investigation; however, it has not formed part of our studies which have been confined to problems with unimodal posteriors. Moreover, we recall that the computational cost of RML can be amortized if each ensemble member is computed in parallel. Therefore, the cost of the parallel implementation of RML equals the cost of LMAP. Note also that, for very large problems, the factorization of C_{MAP} used in Eq. (22) may be computationally prohibitive while the covariance of RML is computed directly from the ensemble at a negligible cost.

For each of our experiments, very poor approximations of the posterior distribution are obtained with EnKF with a small ensemble size. However, covariance localization leads to a significant reduction in the relative error of the mean and variance with respect to the posterior. Note for example that, in the two-phase model with a small number of wells, localization reduces the relative error in the mean and the variance by a factor of 2. Additionally, EnSRF provides better approximations of the posterior (in terms of mean and variance) than the ones obtained with EnKF. Furthermore, in all our experiments, the ensemble generated with EnSRF with localization provides the best approximation of the posterior in terms of variance. The advantage of using covariance localization as well as using EnSRF instead of EnKF has been widely investigated in terms of

reconstructing the truth and/or recovering the truth within the confidence intervals provided by the ensemble approximations. Our results offer now numerical evidence of the advantage of using covariance localization and square root filters for reconstructing the posterior distribution. The choice of covariance localization that provides optimal approximation of the posterior distribution must be further investigated.

Reducing the detrimental effect of sampling error due to the small ensemble size and the possible large amount data is essential in practical applications where a small number of ensemble members is required to avoid high computational cost in data assimilation. Nonetheless, our results indicate that even for a large ensemble size where presumably sampling error issues are attenuated, we find that EnKF does not converge to the posterior distribution. In fact, as the ensemble sizes increased, the converged Gaussian approximation provided by EnKF resulted in errors of at least 10 % both in mean and variance. In addition, the approximations provided with those large-size ensembles do not coincide with the approximations provided by either LMAP or RML.

In summary, our study sheds light on various aspects of the ad hoc Gaussian approximate filters used in practice to approximate high-dimensional posterior distributions on geological reservoir properties. The study has been made possible by use of a fully resolved gold standard MCMC computation which allows for a clear and well-founded evaluation of the ad hoc algorithms. In our opinion, more evaluations of this kind will be beneficial in guiding the future evolution of the ad hoc Gaussian approximate filters that are so widely used in practice. Further evaluation of recently proposed methods [7, 12, 22, 31, 32] may be useful to understand and improve the uncertainty quantification properties of these techniques under more general conditions. For example, it is relevant to consider the case where non-Gaussian priors characterize the geology for which ensemble methods such as the Gaussian mixture filters [12, 31, 32] have been proposed.

References

1. Aanonsen, S.I., Naevdal, G., Oliver, D.S., Reynolds, A.C., Valles, B.: The ensemble Kalman filter in reservoir engineering—a review. SPE J. **14**(3), 393–412 (2009)
2. Aarnes, J.E., Gimse, T., Lie, K.-A.: An introduction to the numerics of flow in porous media using Matlab. In: Hasle, G., Lie, K.-A., Quak, E. (eds.) Geometric Modelling, Numerical Simulation, and Optimization, pp. 265–306. Springer, Berlin (2007)

3. Arpat, B., Caers, J., Strebelle, S.: Feature-based geostatistics: an application to a submarine channel reservoir. In: Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, SPE 77426 (2002)

4. Barker, J.W., Cuypers, M., Holden, L.: Quantifying uncertainty in production forecasts: another look at the PUNQ-S3 problem. SPE J. **6**, 433–441 (2001)

5. Brooks, S.P., Gelman, A.: General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. **7**, 434–455 (1998)

6. Chen, Y., Oliver, D.: Cross-covariances and localization for EnKF in multiphase flow data assimilation. Comput. Geosci. **14**, 579–601 (2010). doi:10.1007/s10596-009-9174-6

7. Chen, Y., Oliver, D.S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math. Geosci. **44**, 1–26 (2012)

8. Chen, Z., Huan, G., MA, Y.: Computational methods for multiphase flows in Porous Media. Society for Industrial and Applied Mathematics, Philadelphia (2006)

9. Cotter, S.L., Roberts, G.O., Stuart, A.M., White, D.: MCMC methods for functions: modifying old algorithms to make them faster. To Appear. Stat. Sci. (2013, in press). arXiv:1202.0709

10. Dashti, M., Law, K.J.H., Stuart, A.M., Voss, J.: Map estimators and posterior consistency in Bayesian nonparametric inverse problems. (2013, submitted). arXiv:1303.4795

11. Deutsch, C.V.: Geostatistical Reservoir Modeling. Oxford University Press, Oxford (2002)

12. Dovera, L., Della Rossa, E.: Multimodal ensemble Kalman filtering using Gaussian mixture models. Comput. Geosci. **15**, 307–323 (2011)

13. Efendiev, Y., Datta-Gupta, A., Ma, X., Mallic, B.: Efficient sampling techniques for uncertainty quantification in history matching using nonlinear error models and ensemble level upscaling techniques. Water Resour. Res. **45**(W11414), 11 (2009)

14. Emerick, A., Reynolds, A.: EnKF-MCMC. In: Proceedings of the SPE EUROPEC/EAGE Annual Conference and Exhibition, Barcelona, SPE 131375 (2010)

15. Emerick, A., Reynolds, A.: Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. Comput. Geosci. **15**, 251–269 (2011). doi:10.1007/s10596-010-9198-y

16. Emerick, A., Reynolds, A.: Combining the ensemble Kalman filter with Markov chain Monte Carlo for improved history matching and uncertainty characterization. SPE J. **17**(2), 418–440 (2012)

17. Emerick, A.A., Reynolds, A.C.: Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. Comput. Geosci. **17**(2), 325–350 (2013)

18. Gao, G., Zafari, M., Reynolds, A.: Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF. SPE J. **11**, 506–515 (2006)

19. Gaspari, G., Cohn, S.E.: Construction of correlation functions in two and three dimensions. Q. J. R. Meteorol. Soc. **125**(554), 723–757 (1999)

20. Kaipio, J.P., Somersalo, E.: Statistical and Computational Inverse Problems. Springer, New York (2005)

21. Law, K.J.H., Stuart, A.M.: Evaluating data assimilation algorithms. Mon. Weather Rev. **140**, 3757–3782 (2012)

22. Li, G., Reynolds, A.: Iterative ensemble Kalman filters for data assimilation. In: Proceedings of the SPE Annual Technical Conference and Exhibition, Anaheim, California, SPE 109808 (2007)

23. Liu, N., Oliver, D.S.: Evaluation of Monte Carlo methods for assessing uncertainty. SPE J. **8**, 188–195 (2003)

24. Iglesias, K., Law, M., Stuart, A.M.: Ensemble Kalman methods for inverse problems. Inverse Problems **29**, 045001 (2013). arXiv:1202.0709

25. Ma, X., Al-Harbi, A., Datta-Gupta, A., Efendiev, Y.: An efficient two-stage sampling method for uncertainty quantification in history matching geological models. SPE J. **13**(1), 7787 (2008)

26. Reynolds, A.C., Oliver, D.S., Liu, N.: Inverse Theory for Petroleum Reservoir Characterization and History Matching, 1st edn. ISBN:9780521881517. Cambridge University Press, Cambridge (2008)

27. Oliver, D., Chen, Y.: Recent progress on reservoir history matching: a review. Comput. Geosci. **15**, 185–221 (2011). doi:10.1007/s10596-010-9194-2

28. Oliver, D.S., Cunha, L.B., Reynolds, A.C.: Markov chain Monte Carlo methods for conditioning a permeability field to pressure data. Math. Geol. **29**, 61–91 (1997)

29. Russell, T.F., Wheeler, M.F.: Finite element and finite difference methods for continuous flows in porous media. In: Ewing, R.E. (ed.) Mathematics of Reservoir Simulation. SIAM, Philadelphia

30. Sakov, P., Oke, P.R.: Implications of the form of the ensemble transformation in the ensemble square root filters. Mon. Wea. Rev. **136**, 1042–1053 (2008)

31. Stordal, A.S., Karlsen, H.A., Nvdal, G., Skaug, H.J., Valls, B.: Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. Comput. Geosci. **15**, 293–305 (2011)

32. Stordal, A.S., Valestrand, R., Karlsen, H.A., Nvdal, G., Skaug, H.J.: Comparing the adaptive Gaussian mixture filter with the ensemble Kalman filter on synthetic reservoir models. Comput. Geosci. **16**, 467–482 (2012)

33. Stuart, A.M.: Inverse problems: a Bayesian perspective. In Acta Numerica. **19** (2010)

34. Tavakoli, R., Reynolds, A.: Monte Carlo simulation of permeability fields and reservoir performance predictions with SVD parameterization in RML compared with EnKF. Comput. Geosci. **15**, 99–116 (2011). doi:10.1007/s10596-010-9200-8

35. Naevdal, G., Skaug, H.J., Thulin, K., Aanonsen, S.I.: Quantifying Monte Carlo uncertainty in the ensemble Kalman filter. SPE J. **16**(1), 172–182 (2011)

36. Li, G., Wang, Y., Reynolds, A.C.: Estimation of depths of fluid contacts by history matching using iterative ensemble-Kalman smoothers. SPE J. **15**(2), 509–525 (2010)