

GRID-BASED INVERSION OF PRESSURE TRANSIENT TEST DATA WITH STOCHASTIC GRADIENT TECHNIQUES

Richard Booth,^{1,*} Kirsty Morton,¹ Mustafa Onur,² & Fikri Kuchuk³

¹Schlumberger BRGC, Rio de Janeiro, Brazil

²Technical University of Istanbul, Istanbul, Turkey

³Schlumberger Ribould Product Centre, Clamart, France

Original Manuscript Submitted: 15/05/2011; Final Draft Received: 20/10/2011

In any subsurface exploration and development, indirect information and measurements, such as detailed geological description, outcrop studies, and direct measurements (such as seismic, cores, logs, and fluid samples), provide useful data and information for static reservoir characterization, simulation, and forecasting. However, core and log data delineate rock properties only in the vicinity of the wellbore, while geological and seismic data are usually not directly related to formation permeability. Pressure transient test (PTT) data provide dynamic information about the reservoir and can be used to estimate rock properties, fluid samples for well productivity, and dynamic reservoir description. Therefore, PTT data are essential in the industry for the general purposes of production and reservoir engineering as well as commonly used for exploration environments. With the need for improved spatial resolution of the reservoir parameters, grid-based techniques have been developed in which the reservoir properties are discretized over a fine grid and characterization of the probable state of the reservoir is sought using the Bayesian framework. Unfortunately, for the exploration of hydrocarbon-bearing formations, the available prior information is often limited: in particular, unexpected geological features, such as fracture and faults, may be present. There are two groups of recent methods for dynamic characterization of the reservoir: (i) data assimilation techniques, e.g., ensemble Kalman filter (EnKF) and (ii) maximum-likelihood techniques, such as gradient-based methods. The EnKF is designed to produce a set of realizations of the reservoir properties that fit the PTT data; however, the method often fails to honor the data when unexpected features are not captured by the prior model. The alternative gradient-based methods do provide a good fit to the PTT data. They can also be made efficient for high-dimensional problems by using an adjoint scheme for determining the gradient of the log-likelihood function. However, as a maximum likelihood technique, this method only yields a single realization of the reservoir. It is important to maintain a model of the uncertainty of the reservoir characterization after PTT data assimilation, so that the risk associated with future decisions is understood. We therefore present and investigate a stochastic, gradient-based method that allows for proper sampling of realizations of the reservoir parameters that preserve the fit with the PTT data. The results indicate that our proposed method is quite encouraging for efficiently generating realizations of rock property distributions conditioned to PTT data sets and a given prior geostatistical model.

KEY WORDS: Bayesian inference, geological applications, transient well test interpretation, reservoir characterization

1. INTRODUCTION

Pressure transient testing is a long established procedure [1–3] for determining the productivity of a well and the properties of the formation (reservoir) from downhole and/or surface pressure and flow-rate measurements. In a pressure

*Correspond to Richard Booth, E-mail: RBooth2@slb.com

transient test (PTT), a well is produced (or injected) for a known period of time (drawdown test) and subsequently may be shut in for a buildup test, or for a sequence of production and buildup periods. During the PTT, the flow rate at the production well (surface or downhole) and transient pressure in the wellbore and/or in an observation well are measured. Using conventional interpretation methods [Horner or Semilog (MDH) and/or type-curve matching with the aid of derivative plots], reservoir pressure, an effective ‘average’ permeability of the reservoir, the skin factor, the wellbore storage, etc., are obtained from the PTT data.

The permeability obtained from a single drawdown or buildup test is an average permeability over a radius of investigation [4]. In other words, the mild permeability heterogeneity around the wellbore within a radius of investigation is averaged by the transient pressure diffusion during a transient well test. However, if in addition to the well test data acquired at the production well, the transient pressure data acquired at other spatial locations are available (interference tests performed in several wells [5]), it is possible to establish a more complete picture of the spatial distribution of the reservoir parameters. For instance, vertical interference tests conducted with multiprobe wireline formation testers can be used to determine the spatial distribution of the permeability vertically around the wellbore [6]. However, whether using a single transient test or multiwell (multiple spatial location) interference tests, it is well known that the sensitivity of PTT to small-scale spatial variations of the permeability or other formation parameters is fairly low. On the other hand, significant geological features, such as a region of low permeability (sealing faults) or high permeability (conductive fractures and faults), might still be identifiable from the pressure transient data.

At the exploration phase during the early years of the well, when PTTs are used for reservoir characterization and determination of well productivity, the prior reservoir knowledge is usually minimal at the well test scale. Larger scale reservoir properties might be available from seismic and geology and/or outcrop analogies, and at a small scale, the prior formation knowledge might be available from logs or cores. Therefore, for exploration and development pressure transient testing, the existence of and locations of many subseismic features, such as strong spatial permeability variations, faults, fractures, pinch outs, etc., are not known a priori. Furthermore, we are unlikely to have a detailed geostatistical description of the reservoir.

The method of Backus and Gilbert as also described in [4, 7], provides one method for determining the spatial variation of reservoir properties. However, the assumption of small variations in the properties makes the method of Backus and Gilbert inappropriate for geological feature identification. Moreover, its lack of robustness when it is used with inaccurate data suggests the need for some sort of prior modeling.

We therefore need to follow a Bayesian framework to use PTT data for spatial resolution of reservoir parameters, with a probabilistic prior description of the reservoir updated by the measurements to give a (probabilistic) posterior description of the reservoir. To capture the spatial variability of the reservoir parameters, we turn to “pixel” methods, with the reservoir parameters discretized over a grid. In previous works, these grids have typically been regular [8, 9], principally because the prior geostatistical model required a regular grid. Irregular grids with a finer resolution close to the wellbore are useful for the numerical simulation of the well test, and moreover, we expect that the resolution of the inversion should be highest closest to the wellbore. For this reason, we use the same irregular grid that is also used for numerical simulation of the PTT data to discretize the reservoir.

In this paper we first describe a prior model that can efficiently describe the geostatistics of the reservoir properties on an irregular grid. We then obtain an expression for the posterior probability distribution of the reservoir parameters. The probability density function is expensive to evaluate, and therefore, a more simple description is required. One of the most useful pieces of information that can be extracted from the posterior probability density is the maximum likelihood parameters. However, this information alone does not tell us how likely these parameters are and, thus, some measure of the uncertainty is also usually desirable, typically either random samples from the distribution or a Gaussian approximation to the posterior probability distribution.

One method that has received much recent attention [10, 11] is the ensemble Kalman filter (EnKF), in which the probability distribution is represented by an ensemble of samples that are updated as each new measurement becomes available. The ensemble Kalman filter assumes that at each step the probability distribution is Gaussian, and failure of either the Gaussian model to capture the true probability distribution or the ensemble to accurately capture the Gaussian model can lead to failure of EnKF. This is potentially more of a problem for inversion of PTT data than for other applications of EnKF, because in well testing, the early time measurements give important information about the near-wellbore permeability distribution, which cannot easily be recovered from measurements at later times. In [12],

we found that, although the EnKF method can give excellent results, its success is dependent on how accurately the prior described the true model and, therefore, is not always appropriate for identifying and characterizing unexpected reservoir features. The EnKF method does not always accurately match the PTT data.

In contrast, gradient-based methods for determining the maximum a posteriori estimation of parameters are able to accurately match with PTT data and are applicable even when the probability distribution is not adequately approximated by a Gaussian model. In this paper, we review how to efficiently calculate gradients using adjoint variables and then show how the priors that we have introduced are useful in improving the progress of gradient-based descent methods.

Because these methods only determine the most likely parameters, we then investigate methods to produce samples from the posterior distribution. Existing methods once again rely on the assumption of a Gaussian model for the posterior probability distribution, but in this paper, we detail a method for proper sampling from the posterior distribution by means of the Langevin equation.

2. PRIOR MODEL

The prior model is a probability density functional $\pi_0[\mathbf{u}]$, which describes the likelihood of a particular set of parameters (reservoir or formation properties), \mathbf{u} (e.g., the spatially variable porosity and components of permeability), representing the true state of the reservoir. The prior model can always be written in the form

$$\pi_0[\mathbf{u}] \propto \exp(-R[\mathbf{u}]),$$

where for commonly used Gaussian priors the function $R[\mathbf{u}]$ will be quadratic. The function $R[\mathbf{u}]$ will typically itself be described by parameters that characterize our prior knowledge of the reservoir, such as a typical correlation length or the local variance of the permeability. These parameters that describe the prior are known as “hyperparameters,” to distinguish them from the parameters of the model under analysis. Given a particular set of parameters, we need to model the likelihood of a particular set of pressure measurements \mathcal{M} being made, or the “posterior update.” The pressure measurements will never be exact due to various sources of noise, particularly when oil is being produced from the well. When the errors of each measurement are considered as independent and Gaussian, the likelihood of a particular set of measurements \mathcal{M} is

$$\pi[\mathcal{M}|\mathbf{u}] \propto \exp(-I[\mathbf{u}]), \quad \text{where} \quad I[\mathbf{u}] = \frac{1}{2} \sum_{m=1}^{N_w} \sum_{r=1}^{N_t} \frac{[p_{wm}(\mathbf{u}, t_r) - \tilde{p}_{wm}(t_r)]^2}{\sigma_{wm}^2(t_r)},$$

where N_w , N_t are respectively the number of wells and pressure measurements; $p_{wm}(\mathbf{u}, t_r)$ and $\tilde{p}_{wm}(t_r)$ are respectively the simulated pressures in the well (of the model described by parameters \mathbf{u}) and the measured pressures in the well at times t_r ; and $\sigma_{wm}^2(t_r)$ is the error variance of the measurement. Using Bayes’ theorem, we are able to deduce that the posterior distribution of the parameters as

$$\pi[\mathbf{u}|\mathcal{M}] \propto \exp(-H[\mathbf{u}]), \quad \text{where} \quad H[\mathbf{u}] = I[\mathbf{u}] + R[\mathbf{u}].$$

For this application, it is important that the prior is independent of the grid and that it provides a sufficient description of the small-scale spatial correlation of the parameters. At the same time, only very limited geostatistical information may be available and, thus, ideally the prior should be described by a small number of hyperparameters. Local Gaussian random fields can satisfy all of these requirements. For a parameter $u(\mathbf{x})$, where \mathbf{x} is a spatial vector, a local Gaussian random field prior is described by

$$R[u(\mathbf{x})] = \frac{1}{2} \int u \mathcal{L} u \, d\mathbf{x},$$

where \mathcal{L} is a linear differential operator, such as

$$\mathcal{L} u = h_2 \nabla^4 u - h_1 \nabla^2 u + h_0 u,$$

for arbitrary constants h_0, h_1, h_2 . This random field has a useful variogram, as shown in [13], and is best described by rewriting the linear operator as

$$\mathcal{L}u = \frac{1}{\sigma^{*2}} (\rho^4 \nabla^4 u - 2S\rho^2 \nabla^2 u + u), \quad (1)$$

where ρ is a correlation length, S (which can vary between -1 and 1) controls the shape of the variogram, and σ^{*2} is related to the local covariance σ^2 by

$$\sigma^{*2} = \frac{4\pi\rho^2\sqrt{1-S^2}}{\cos^{-1}S}\sigma^2,$$

in two dimensions, and for three-dimensional random fields by

$$\sigma^{*2} = 4\pi\sqrt{2(S+1)}\rho^3\sigma^2.$$

The key advantage of employing this type of random field is that, once the problem is discretized, the inverse covariance matrix is readily obtained—it is simply the finite difference matrix for the differential operator described in Eq. (1)—and it will be sparse.

Each component of the permeability is approximately log-normally distributed, and thus,

$$u_x = \log k_x - \overline{\log k_x},$$

can be modeled by a local Gaussian prior as above. The remaining components of the permeability distribution may be treated similarly, defining u_y and u_z . It is possible to treat the porosity in the same way,

$$u_\phi = \log \phi - \overline{\log \phi}.$$

However, such a model allows the possibility of $\phi > 1$ and therefore can only be applied when either it is clear beforehand that the porosity is small (i.e., $\phi \ll 1$) or that the variation of the porosity is small.

The prior model should not only describe the spatial correlation of the parameters, but also the local correlation between distinct parameters (e.g., permeability and porosity). Then, the complete prior model can be written as

$$\pi_0[\mathbf{u}] \propto \exp\left(-\frac{1}{2} \int \mathbf{u}^T \mathbf{A} \mathbf{u} + u_x \mathcal{L}_x u_x + u_y \mathcal{L}_y u_y + u_z \mathcal{L}_z u_z + u_\phi \mathcal{L}_\phi u_\phi \, d\mathbf{x}\right), \quad (2)$$

where $\mathbf{u} = (u_x, u_y, u_z, u_\phi)^T$ and \mathbf{A} is a 4×4 matrix, the inverse of the covariance matrix between the parameters. The operators \mathcal{L}_x , etc., are of the form (1), although the h_0 term can be included within \mathbf{A} . This model is described by at most 22 hyperparameters: four average values; four correlation lengths; four structure hyperparameters; and ten entries in \mathbf{A} . However, it is easy to significantly reduce this number of hyperparameters by, for example, assuming that the correlation lengths and structure hyperparameters are all equal, and/or by assuming that the permeability is either fully isotropic or instead that there are only horizontal and vertical components of permeability, with the permeability identical in the x and y directions. If the permeability is known to be isotropic with $u = \log k - \overline{\log k}$, then the prior model is simply

$$\pi_0[u] \propto \exp\left[-\frac{1}{2} \int u (h_2 \nabla^4 u - h_1 \nabla^2 u + h_0 u) \, dx\right]. \quad (3)$$

3. FORWARD AND ADJOINT MODELS

3.1 Forward Model

In this paper, we consider pressure diffusion for a single-phase, constant-viscosity, and slightly compressible (the rate of change of density with pressure is linear) fluid in a three-dimensional porous media. In the absence of any sources or sinks, the pressure field in the reservoir is described by the diffusivity equation as

$$c_t \phi \frac{\partial p}{\partial t} = \nabla \cdot \left(\frac{\mathbf{K}}{\mu} \nabla p \right), \quad (4)$$

where c_t is the total compressibility, \mathbf{K} is the diagonal permeability tensor, and μ is the fluid viscosity. Note that gravity is neglected throughout, because for a weakly compressible fluid, the only effect of gravity is to introduce a hydrostatic term to the pressure, which is normally termed as potential.

We do not intend for our numerical grid to capture the geometry of the wellbore. Therefore, we apply Peaceman's model [14] and treat the well as a line source within the grid blocks through which it penetrates. The wellbore pressure in the m th well then evolves according to

$$C_{wm} \frac{dp_{wm}}{dt} = \sum_{j \in \mathcal{B}_m} \text{WI}_j (p_j - p_{wm}) - q_m(t), \tag{5}$$

where C_{wm} is a property of the well and the fluid, q_m is the flow rate of fluid produced from the well, \mathcal{B}_m is the set of grid blocks that well m penetrates, and in the j th grid block WI_j is the Peaceman well index and p_j is the grid block pressure. The discretized model for the evolution of the pressure in grid block i is

$$c_t V_i \phi_i \frac{\partial p_i}{\partial t} - A_{ij} p_j + \sum_{m=1}^{N_w} \sum_{j \in \mathcal{B}_m} \delta_{ij} \text{WI}_j (p_j - p_{wm}) = 0, \tag{6}$$

where V_i is the volume of the i th grid block, $A_{ij} p_j$ is the (symmetric) finite-volume discretization of the right-hand side of (4), and δ_{ij} is the Kronecker δ function. The initial pressure in both wells and reservoir is known (although it would also be possible to include this initial pressure as an additional unknown parameter to be determined from the pressure data). The reservoir permeability enters the forward model [Eqs. (5) and (6)] via A_{ij} and WI_j .

To determine the sensitivity of the pressure transient data to changes in the reservoir parameters, we must determine the gradient of $H[\mathbf{u}]$. The gradient of $R[\mathbf{u}]$ can be easily determined yielding a linear function of \mathbf{u} ; however, determining the gradient of $I[\mathbf{u}]$ is more difficult. The most direct method of determining this gradient is simply to use finite differences with small variations made to the parameter values. This approach is impractical because we would need to vary the parameters in each grid cell, leading to an extremely large number of model evaluations. Fortunately, we are able to obtain the gradient much more effectively by employing an adjoint method, which will require only a single evaluation of the forward model [Eqs. (5) and (6)] and a single evaluation of an adjoint model.

3.2 Adjoint Model

The difficulty in finding the gradient of I with respect to \mathbf{u} arises because we are unable to find a closed form for $p_{wm}[\mathbf{u}]$; however, by introducing adjoint variables to the pressure in the grid blocks and the wells, λ_i, λ_{wm} , we need never explicitly calculate the gradient of p_{wm} with respect to \mathbf{u} . We introduce

$$\begin{aligned} J[\mathbf{u}, p_i, p_{wm}, \lambda_i, \lambda_{wm}] &= \frac{1}{2} \sum_{m=1}^{N_w} \sum_{r=1}^{N_t} \frac{[p_{wm}(t_r) - \tilde{p}_{wm}(t_r)]^2}{\sigma_{wm}^2(t_r)} \\ &+ \sum_{m=1}^{N_w} \int_0^T \lambda_{wm}(t) \left[C_{wm} \frac{dp_{wm}}{dt} + q_m(t) - \sum_{j \in \mathcal{B}_m} \text{WI}_j (p_j - p_{wm}) \right] dt \\ &+ \sum_i \int_0^T \lambda_i(t) \left[c_t V_i \phi_i \frac{\partial p_i}{\partial t} - A_{ij} p_j + \sum_{m=1}^{N_w} \sum_{j \in \mathcal{B}_m} \delta_{ij} \text{WI}_j (p_j - p_{wm}) \right] dt, \end{aligned}$$

and do not strictly enforce that p_{wm} and p_i satisfy (5) and (6), but instead note that when these equations are satisfied, $I[\mathbf{u}] = J[\mathbf{u}, p_i, p_{wm}, \lambda_i, \lambda_{wm}]$, $\partial J / \partial \lambda_{wm} = 0$, and $\partial J / \partial \lambda_i = 0$. We can also seek the necessary conditions which λ_i and λ_{wm} must satisfy if we are to ensure that $\partial J / \partial p_{wm} = 0$ and $\partial J / \partial p_i = 0$, and find that we require

$$-C_{wm} \frac{d\lambda_{wm}}{dt} = \sum_{j \in \mathcal{B}_m} \text{WI}_j (\lambda_j - \lambda_{wm}) - \sum_{r=1}^{N_t} \frac{p_{wm}(t_r) - \tilde{p}_{wm}(t_r)}{\sigma_{wm}^2(t_r)} \delta(t - t_r), \tag{7}$$

and

$$-c_t V_i \phi_i \frac{\partial \lambda_i}{\partial t} - A_{ij} \lambda_j + \sum_{m=1}^{N_w} \sum_{j \in \mathcal{B}_m} \delta_{ij} W I_j (\lambda_j - \lambda_{wm}) = 0, \quad (8)$$

with the final conditions $\lambda_i(T) = 0$ and $\lambda_{wm}(T) = 0$. The adjoint problem [Eqs. (7) and (8)] is almost identical to the forward model for the pressure; however, unlike the forward problem, it must be solved backward in time, starting at a time $t = T$, after the last of the pressure measurements. Also in solving the adjoint problem, the rate of fluid production from the well must be replaced by the mismatch between the measured values of pressure in the wells and the values obtained by the forward model.

When the pressure and its adjoint solve (5)–(8), we have (writing U_j for the value of a generic parameter in grid block j),

$$\frac{dJ}{dU_j} = \frac{\partial J}{\partial U_j} + \sum_{m=1}^{N_w} \left(\frac{\partial J}{\partial p_{wm}} \frac{dp_{wm}}{dU_j} + \frac{\partial J}{\partial \lambda_{wm}} \frac{d\lambda_{wm}}{dU_j} \right) + \sum_i \left(\frac{\partial J}{\partial p_i} \frac{dp_i}{dU_j} + \frac{\partial J}{\partial \lambda_i} \frac{d\lambda_i}{dU_j} \right) = \frac{\partial J}{\partial U_j}$$

and moreover, with (5) and (6) satisfied, $I[\mathbf{u}] = J[\mathbf{u}, p_i, p_{wm}, \lambda_i, \lambda_{wm}]$ for all \mathbf{u} , and so the derivative of $I[\mathbf{u}]$ is

$$\frac{dI}{dU_j} = \frac{\partial J}{\partial U_j}.$$

For example, the derivative of I with respect to the porosity in grid block i is

$$\frac{\partial I}{\partial \phi_i} = V_i c_t \int_0^T \lambda_i(t) \frac{\partial p_i}{\partial t} dt. \quad (9)$$

It should be noted that this is *not* the same as the value in grid block i of the derivative of I with respect to the parameter, which is instead

$$\left(\frac{\partial I}{\partial \phi} \right)_i = \frac{1}{V_i} \frac{\partial I}{\partial \phi_i} = c_t \int_0^T \lambda_i(t) \frac{\partial p_i}{\partial t} dt. \quad (10)$$

This second representation of the gradient is more useful because it is not affected by the size of the grid block. The derivative with respect to u_x, u_y, u_z may be obtained in a similar way, although the matrix A_{ij} and the well indexes must be differentiated.

The approach that we have outlined thus far is applicable to determining a fully anisotropic permeability distribution and a porosity distribution. Henceforth, to aid the exposition of the problem, we restrict ourselves to a reservoir with an isotropic permeability distribution, with log-permeability denoted by u , and with known porosity. The more general methodology follows along similar lines to the simplified approach that we now describe.

4. GRADIENT DESCENT

Our aim in studying this inverse problem is primarily to find the true reservoir parameters; however, we must recognize that the information that we provide will almost never be sufficient to conclusively answer this question. A complete understanding of the posterior probability distribution would be the most complete answer that we could hope to give, but even when the reservoir is described by a modest number of parameters, a complete description of the posterior probability distribution is often numerically intractable.

A starting point for describing the posterior probability distribution is to find the most likely set of parameters. The simplest method for finding the most likely set of parameters is to use the steepest descent method. Crucially, the steepest descent method only uses the value of $H[u]$ and its gradient, which, as we have described in the previous section, may be evaluated inexpensively. The steepest descent method is usually given by

$$u_{n+1} = u_n - \alpha_n \frac{\partial H}{\partial u} [u_n],$$

where α_n is a parameter to be determined and u_n is the estimate of the minimum after n iterations of the method. For sufficiently small values of α_n , it is guaranteed that $H[u_{n+1}]$ will be lower than $H[u_n]$; however, to make good progress with as few iterations, and therefore model simulations, as possible we need to choose α_n as large as possible. A line search scheme is therefore used to determine an optimal value of α_n . As noted in the previous section, when working with an irregular grid it is appropriate to use the discretization of the gradient as given by (10) rather than (9), as this expression is not affected by the size of the grid blocks.

One interpretation of the steepest descent scheme is as the explicit Euler method discretization of

$$\frac{\partial u}{\partial \tau} = -\frac{\partial H}{\partial u}. \tag{11}$$

A minimizer of $H[u]$ is a stable, steady-state solution to the above equation, and thus for large values of τ , we expect this equation to find a local minimizer of H . The explicit Euler discretization of Eq. (11) is only conditionally stable. In practice, the line search ensures that the τ -step is sufficiently small to ensure stability. The maximum permissible τ -step for which the explicit Euler discretization is stable will be small whenever the problem is stiff. Because from (3)

$$\frac{\partial R}{\partial u} = h_2 \nabla^4 u - h_1 \nabla^2 u + h_0 u,$$

the local Gaussian regularization that we are proposing will yield a stiff problem at this stage, becoming stiffer as the grid is refined. This stiff behavior is likely to be a generic property of all scale-independent prior models. If we are forced to take small steps, then we will have slow convergence and be required to perform many simulations of the model and its adjoint.

To avoid the restrictions of the stability criterion, we should consider implicit schemes for the discretization of (11), e.g. ,

$$u_{n+1} = u_n - \alpha_n \theta \frac{\partial H}{\partial u}[u_{n+1}] - \alpha_n (1 - \theta) \frac{\partial H}{\partial u}[u_n],$$

where θ lies between 0 and 1 (this scheme is unconditionally stable provided $\theta \geq 1/2$). However, because $\partial H/\partial u$ is both nonlinear and nonlocal, it is not easy to solve for u_{n+1} , unless $\theta = 0$. We shall therefore treat the term in H arriving from the prior model differently from the term arriving from the update and consider the scheme

$$u_{n+1} = u_n - \alpha_n \frac{\partial I}{\partial u}[u_n] - \alpha_n \theta \mathcal{L}u_{n+1} - \alpha_n (1 - \theta) \mathcal{L}u_n. \tag{12}$$

The linear part of the scheme shown in Eq. (12) is still unconditionally stable for $\theta \geq 1/2$. Thus, this scheme avoids the stiffness problem caused by the prior model while still allowing us to easily solve for u_{n+1} , because the update is treated explicitly. When the nonlinear and nonlocal term provides the dominant contribution toward stiffness, it may be appropriate to use a predictor-corrector scheme; however, the fast convergence of such schemes is still dependent on the use of a reasonably small step size.

The semi-implicit scheme (12) permits larger steps to be taken, although the scheme is still not guaranteed to be unconditionally stable. To determine an optimal value of α_n , a line search is still required. We ensure that the line search satisfies the Wolfe conditions [15] given as

$$H[u_n + 1] \leq H[u_n] + c_1 \int \left[\frac{\partial H}{\partial u}(u_n) \right]^2 dx, \tag{13}$$

and

$$\left| \frac{\partial H[u_{n+1}]}{\partial \alpha_n} \right| \leq c_2 \left| \frac{\partial H[u_n]}{\partial \alpha_n} \right| \tag{14}$$

where $0 < c_1 < c_2 < 1$ are chosen to control the accuracy of the line search. The second Wolfe condition (14) requires the derivative of the likelihood function with respect to the line-search parameter. Because we are using the

scheme (12), rather than the standard steepest descent scheme, this is no longer simply given by the scalar product of the current gradient and the descent direction. Nevertheless, it is easily evaluated and given by

$$\frac{dH[u_{n+1}]}{d\alpha_n} = \int \frac{\partial H}{\partial u}[u_{n+1}] \frac{du_{n+1}}{d\alpha_n} dx,$$

with

$$(1 + \alpha_n \theta \mathcal{L}) \frac{du_{n+1}}{d\alpha_n} = -\frac{\partial I}{\partial u}[u_n] - \theta \mathcal{L}u_{n+1} - (1 - \theta) \mathcal{L}u_n,$$

found by differentiating (12) with respect to α .

The steepest descent method will converge toward the most likely reservoir parameters. The adaptation to the steepest descent method (12) allows us to work with appropriately stiff priors on fine grids.

4.1 Application to an example

We seek to identify a zone of low-permeability across a single-layered reservoir, as shown in Fig. 1, with PTT data taken from an active production well and two observation wells. The true permeability distribution contains a zone of low permeability, separating one of the observation wells from the producer. The permeability distribution used as a starting point for steepest descent is homogeneous, with a permeability identical to the permeability away from the zone of low permeability. This permeability field is also used as the mode in our prior model. A drawdown test, in which fluid is produced from a well, is followed by a buildup test, where the well is shut in and no fluid is produced. Observation wells are present on either side of the band of low permeability as shown in Fig. 1. All the wells are vertical and fully completed within the reservoir. The observed pressure measurements are shown in Fig. 2, along with the pressure response from both the initial homogeneous model of the reservoir and the final model suggested from the gradient descent scheme.

5. UNCERTAINTY

The steepest descent scheme can only provide information about the most likely realization of the reservoir parameters; however, the information provided from well testing will never be sufficient to completely determine the reservoir properties and, thus, we should expect our result to be subject to a great deal of uncertainty. There are two broad approaches that can be applied to dealing with this uncertainty. One possibility is to describe the likelihood of realizations that are close to the most likely realization and thereby attempt to quantify the level of uncertainty in the realization.

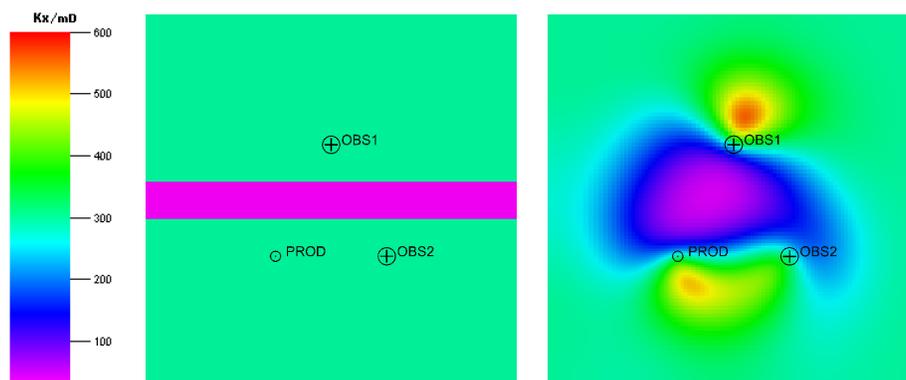


FIG. 1: Plots of the permeability distribution. The first image represents the true permeability distribution, from which synthetic pressure measurements were produced. The second image is an estimate of the most likely permeability distribution consistent with the synthetic pressure measurements, as generated by the gradient descent scheme. The location of the producer well is marked with PROD, and the observation well locations are marked with OBS1/OBS2.

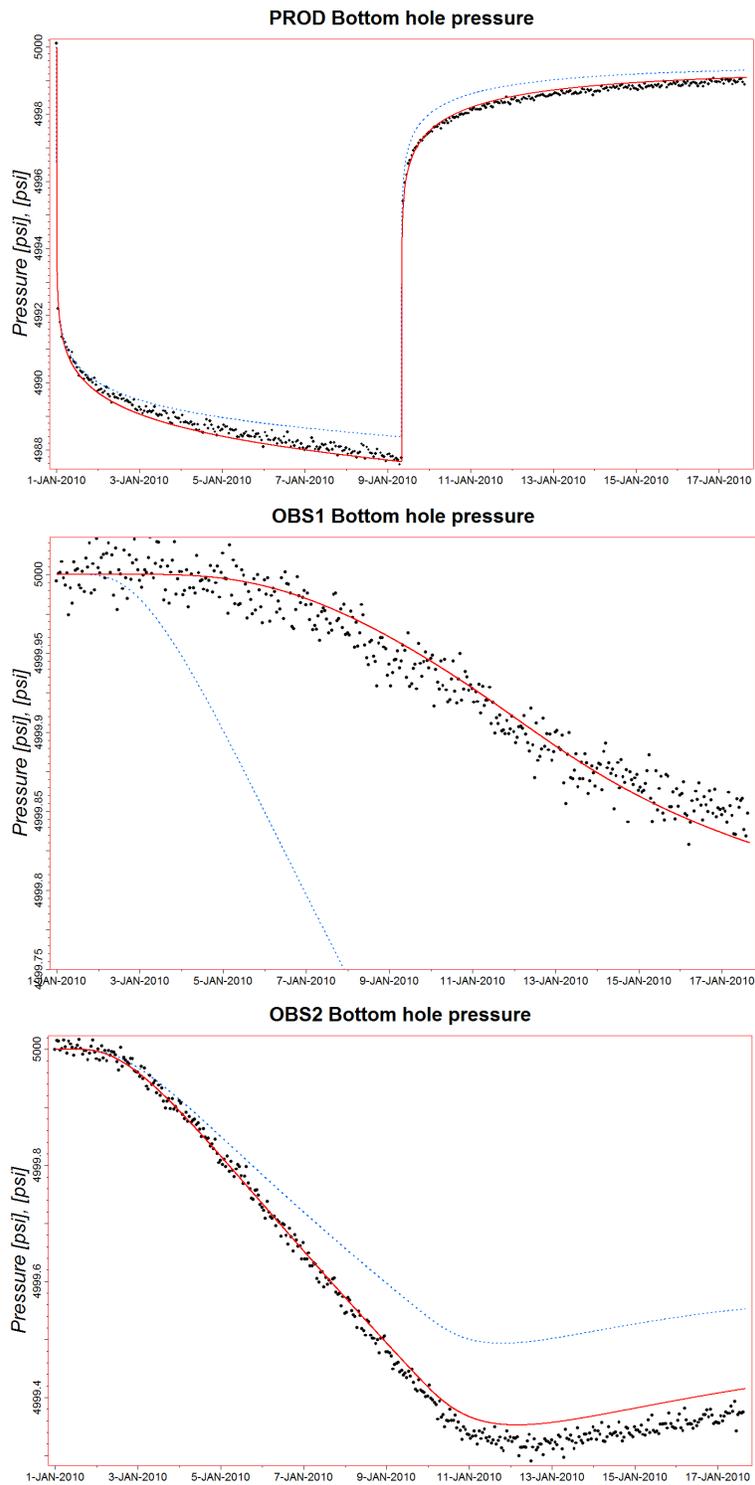


FIG. 2: Comparison of the observed pressure measurements (points) with the pressure response generated by the reservoir model of Fig. 1 (solid line). The pressure response of the initial homogeneous permeability distribution is shown as a dotted line.

This could be done by approximating the second derivative of the log-likelihood $H[u]$ for the most likely realization, which can be done efficiently by making use of the gradients used as part of the minimization procedure. An alternative approach is to seek actual realizations from the posterior, and even with this approach the gradient can be used to help match the pressure transient data. We present, in this section, a method for each of these two approaches.

5.1 Quasi-Newton Methods

Quasi-Newton methods attempt to build an increasingly accurate description of the second derivative of $H[u]$ as part of the minimization of H . This is achieved by using the many gradients of H that are calculated as part of the minimization scheme. The availability of the second derivative also improves the convergence of the scheme, beyond that of steepest descent. One of the most effective quasi-Newton schemes is L-BFGS scheme [15], an adaptation of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) scheme. Both the BFGS and L-BFGS schemes use the values of the parameters and their gradients at previous iterations to approximate the Hessian matrix; however, the L-BFGS scheme only retains a limited number of the previous iterates. This allows the Hessian matrix to be represented by a low-rank update to an initial approximation.

If the updates to the parameters and the gradients are written as

$$s_n = u_{n+1} - u_n \quad \text{and} \quad y_n = \nabla H(u_{n+1}) - \nabla H(u_n).$$

then the Hessian matrix of the BFGS scheme can be represented [15] in the following compact form:

$$B_n = B_0 - \begin{pmatrix} B_0 S_n & Y_n \end{pmatrix} \begin{pmatrix} S_n^T B_0 S_n & \hat{L}_n \\ \hat{L}_n^T & -E_n \end{pmatrix}^{-1} \begin{pmatrix} S_n^T B_0 \\ Y_n^T \end{pmatrix},$$

where B_0 is a simple initial approximation,

$$S_n = [s_0, \dots, s_{n-1}], \quad Y_n = [y_0, \dots, y_{n-1}],$$

and \hat{L}_n and E_n are defined by

$$(\hat{L}_n)_{ij} = \begin{cases} s_{i-1}^T y_{j-1} & \text{for } i > j, \\ 0 & \text{otherwise,} \end{cases}$$

$$E_n = \text{diag}[s_0^T y_0, \dots, s_{n-1}^T y_{n-1}].$$

To obtain the L-BFGS scheme, one simply must discard the earliest entries from S_n and Y_n to ensure that no more than a limited number m of updates are stored. Once an approximation of the Hessian matrix is obtained, the new reservoir parameters are found by solving

$$B_n s_n = -\alpha_n \nabla H(u_n),$$

where α_n should be determined by a line search. As the scheme begins to converge and the approximation to the Hessian improves, we expect that α_n should approach 1. Because the L-BFGS scheme will only make a low-rank update to B_0 , it is sensible to include the prior covariance matrix as a term in B_0 , e.g. ,

$$B_0 = \chi + B_{\text{prior}},$$

where B_{prior} is the discretized covariance matrix for the prior and χ is a parameter that can be scaled to accommodate the new information obtained from the calculation of gradients.

With a local approximation to the posterior covariance, one can try to obtain samples close to the parameters of maximum a posteriori likelihood; however, the samples produced by this approach may not be reliable for nonlinear problems, failing to accurately reproduce the data [16]; which an alternative approach, the randomized maximum likelihood method (RML) is suggested to produce more reliable samples of the posterior distribution. This method

replaces the prior mode with a random sample drawn from the prior, and adds noise to the true data measurements. A sample is then drawn by finding the parameters of maximum a posteriori likelihood for this new distribution. This method seems to be a reasonable approach and will clearly produce samples that appear to honor both the prior model and the data. However, in addition to being computationally intensive to produce samples, it is only guaranteed to draw samples correctly from the posterior if it is Gaussian. We instead propose a method based on the Langevin equation in which noise is added as part of the minimization process, rather than a priori, and which, in the long term, is guaranteed to draw true samples from the posterior, regardless of its distribution.

5.2 Langevin Method

In the context of history matching of geological models, the Langevin method has previously been suggested in [13] and applied in [17, 18]. We now extend the use of adjoint-derived gradients and semi-implicit schemes to the Langevin method. As noted in [13], the Langevin equation given by

$$\frac{\partial u}{\partial \tau} = -\frac{\partial H}{\partial u} + \sqrt{2}\eta(\mathbf{x}, \tau), \tag{15}$$

where $\eta(\mathbf{x}, \tau)$ represents white noise, allows one to generate samples from the posterior distribution. The distribution of u as given by (15) is governed by the Focker-Planck equation [19]. The Focker-Planck equation shows that the equilibrium distribution, achieved for large values of τ , is equal to the posterior distribution. Therefore, for large values of τ , realizations of (15) are all samples of the posterior distribution and it is possible to produce many samples from the posterior (although these will only become independent when separated by large values of τ).

One can note the similarity between the Langevin equation and the continuous form of steepest descent (11), and thus, we approximate (15) using a similar scheme to (12), which can be written as

$$u_{n+1} = u_n - \alpha_n \frac{\partial I}{\partial u}[u_n] - \alpha_n \theta \mathcal{L}u_{n+1} - \alpha_n (1 - \theta) \mathcal{L}u_n + \sqrt{2\alpha_n} \eta(\mathbf{x}), \tag{16}$$

where $\eta(\mathbf{x})$ represents white noise only in the variable \mathbf{x} and the $\sqrt{\alpha_n}$ factor is a consequence of integration of white noise. Using a partially implicit scheme for simulating the Langevin equation has been considered before [20], although applied to a much simpler application than considered here.

The introduction of random noise in (16) prevents a line search being used to optimally choose α_n . Indeed, we must allow the objective function to increase sometimes. However, it is still important that a check is made because the scheme (16) is only partially implicit. We cannot expect to generate samples if the value of α_n has been chosen so large that the underlying deterministic scheme is numerically unstable. We can apply a check based on the Metropolis-Hastings algorithm, in place of a line search, thereby guaranteeing that we correctly draw samples from the posterior distribution.

The Metropolis-Hastings algorithm [21] requires that, to determine the next sample u_{n+1} , we have a method of picking a proposed sample u'_n (the “proposal”) from a probability distribution (the “proposal distribution”) with density function $Q[u'_n; u_n]$, which may depend on the current sample u_n . Once a proposal has been drawn, we calculate

$$p_{\text{accept}} = \frac{\pi[u'_n | \mathcal{M}] Q[u_n; u'_n]}{\pi[u_n | \mathcal{M}] Q[u'_n; u_n]},$$

i.e., the ratio between relative likelihoods of u_n and u'_n in the posterior distribution and the Langevin-derived proposal distribution. If $p_{\text{accept}} > 1$, then the proposal u'_n should always be accepted as the new value for u_{n+1} ; otherwise, it should only be accepted with probability p_{accept} , and on rejection u_{n+1} should take the value u_n . In the long term, the Metropolis-Hastings algorithm will produce samples from the target posterior distribution provided that the proposal distribution allows all possible values of u to be reached. However, in practice, convergence can be very slow, either because many of the proposals are rejected or because there is slow mixing (i.e., u'_n is strongly correlated to u_n). In the ideal case, where the proposal distribution is identical to the posterior distribution, we always find that $p_{\text{accept}} = 1$ and, thus, acceptance is guaranteed and there is perfect mixing because the proposal distribution is independent of u_n .

The effectiveness of the Metropolis-Hastings scheme lies in choosing a good proposal distribution for which rejections are rare and the level of mixing is high (i.e., proposals are not strongly correlated to the current s). Simple proposal distributions, such as setting u'_n equal to u_n plus some random noise, often fail to simultaneously meet these two requirements.

To use the Metropolis-Hastings algorithm as a check on our results from the Langevin equation, we use (16) to define the proposal distribution as

$$Q[u'_n; u_n] = \pi_\eta \left[\frac{1}{\sqrt{2\alpha_n}} \left(u'_n - u_n + \alpha_n \frac{\partial I}{\partial u}[u_n] + \alpha_n \theta \mathcal{L}u'_n + \alpha_n (1 - \theta) \mathcal{L}u_n \right) \right],$$

where

$$\pi_\eta[\eta] \propto \exp \left[-\frac{1}{2} \int \eta(x)^2 dx \right].$$

It is reasonable to expect that, for the same rate of rejection, the use of the gradient should allow for greater mixing than is possible when the proposal only updates the previous sample with random noise. The parameter α_n controls the rate of mixing, with greater mixing for large values of α_n . However, because the deterministic scheme fails for large α_n we should expect a high probability of rejection if α_n is chosen to be too large. If the proposal is accepted, then the Metropolis-Hastings check does not require any additional expensive evaluations of the forward model or adjoint model.

We can investigate the rate of acceptance by first calculating p_{accept} when there is no update to the posterior, i.e., when $I = 0$, for which

$$\log(p_{\text{accept}}) = \frac{\alpha_n}{4} \int (1 - 2\theta) [(\mathcal{L}u_n)^2 - (\mathcal{L}u'_n)^2] dx. \quad (17)$$

Equation (17) shows that when we are sampling from a Gaussian distribution, the proposal generated by the Langevin method is guaranteed to be accepted provided we take $\theta = 1/2$. This is, therefore, the optimal choice of θ in (15) and shows that this method can quickly sample from Gaussian probability distributions (although other methods will often be even more efficient). For a more abstract nonlinear Brownian bridge sampling problem considered in [20], the choice of $\theta = 1/2$ proved crucial to ensure that the optimal step size did not tend to zero as the number of parameters to be sampled increased. As noted in [20] it is likely that such behavior holds more generically when sampling from a distribution that is a nonlinear update of a Gaussian distribution, and our experience supports using $\theta = 1/2$ for the realization of samples conditioned to pressure transient testing data.

With $\theta = 1/2$, we find that the acceptance probability satisfies

$$\begin{aligned} \log(p_{\text{accept}}) &= I[u_n] - I[u'_n] + \frac{1}{2} \int (u'_n - u_n) \left[\frac{\partial I}{\partial u}(u_n) + \frac{\partial I}{\partial u}(u'_n) \right] dx + \frac{\alpha_n}{4} \int \left[\left[\frac{\partial I}{\partial u}(u_n) \right]^2 - \left[\frac{\partial I}{\partial u}(u'_n) \right]^2 \right. \\ &\quad \left. + 2 \left[\frac{\partial I}{\partial u}(u_n) - \frac{\partial I}{\partial u}(u'_n) \right] (\mathcal{L}u_n + \mathcal{L}u'_n) \right] dx. \end{aligned}$$

For small values of α_n , we have

$$\log(p_{\text{accept}}) \approx I[u_n] - I[u'_n] + \frac{1}{2} \int (u'_n - u_n) \left[\frac{\partial I}{\partial u}(u_n) + \frac{\partial I}{\partial u}(u'_n) \right] dx,$$

and the conditions under which each of these two terms will be positive are analogous to the first and second Wolfe conditions (13) and (14). Moreover, for a descent direction, each of these terms would be positive for sufficiently small α_n in the absence of noise, and so acceptance is likely for small α_n .

The step size, α_n , serves as a mixing parameter with small values of α_n giving slow mixing, but ensuring acceptance, and large values of α_n giving enhanced mixing, but risking rejection. The efficiency of the scheme depends carefully on an optimal choice of α_n to obtain a balance, avoiding both slow-mixing and frequent rejection. To date we have only employed ad hoc methods to determine α_n based on the rejection likelihoods obtained.

5.2.1 Application to the Example

We have applied the Langevin method to the problem considered in Section 4.1. The initial permeability distribution used was identical to the homogeneous permeability distribution used to initialize the steepest descent method. This allows us to demonstrate the ability of the Langevin method to invert for the pressure response, although in practice it might be more efficient to initialize the Langevin method with the final state of a deterministic scheme, such as steepest descent.

The step size required to achieve a reasonable acceptance rate must be small for the first few steps at the start of the inversion procedure, but we are soon able to increase it to $\alpha_n = 100$ while ensuring an acceptance rate which remained consistently high. A further increase of the step size was feasible for many iterates, but the consistent application of a higher step size eventually caused the algorithm to become stuck at a certain set of parameters with all generated proposals having an extremely low acceptance probability. The importance of taking $\theta = 0.5$, and therefore avoiding any bias in simulating the prior, was also demonstrated: significantly smaller step sizes were required to maintain the acceptance rate with $\theta \neq 0.5$

As can be seen in the first plot of Fig. 3, the log-likelihood function approaches its equilibrium fairly rapidly, despite the lack of an efficient method of choosing the search step. However, the rapid convergence of the objective function does not necessarily confirm stationarity of the distribution, and it can be seen from the second plot of Fig. 3 that it takes longer before the initial state is forgotten. The progress toward providing a history match and towards

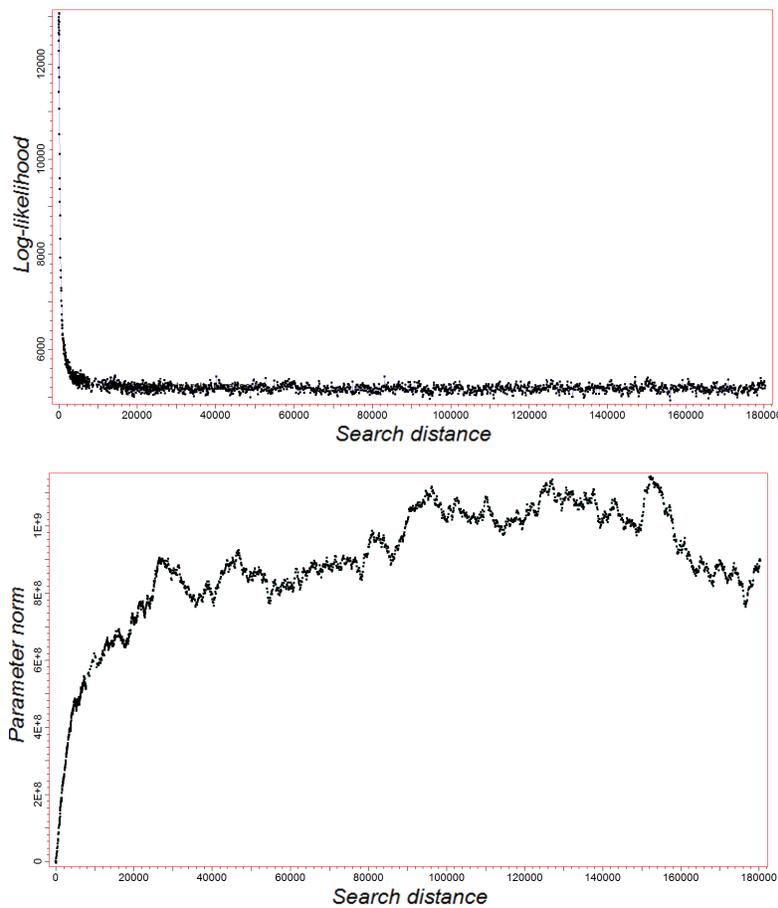


FIG. 3: Plots of the progress of the objective function (log likelihood) and the norm of the vector of parameters against the increase in the Langevin τ parameter.

sampling from the stationary posterior distribution is shown in the permeability samples of Fig. 4, whereas samples of the permeability distribution from later iterates are shown in Fig. 5. The pressure responses corresponding to these samples are shown in Fig. 6.

5.2.2 Improvement of Performance of the Langevin Method

It is critical that the mixing rate is improved and/or that the work required for each proposal is reduced if the Langevin method is to be applicable for real-world problems. We have found that using a semi-implicit method allows one to increase the mixing rate while avoiding frequent rejections, and clearly, the adjoint method significantly reduces the numerical effort per proposal, but these steps alone are unlikely to be sufficient to make the method feasible in practice. In [17] the number of parameters to be determined was reduced by the use of the Karhunen-Loeve expansion to represent the prior model of the permeability distribution. In [17] the gradient was also first calculated using a coarse grid to quickly establish the likely acceptance probability of the true permeability distribution; although the benefit of this strategy is somewhat diminished here by our use of adjoint methods for determining the gradient.

Because the use of a semi-implicit method for the prior helped significantly, we might try to apply a similar idea for the nonlinear nonlocal likelihood update (the $I[u]$ term). Although it is possible to use an explicit predictor-implicit corrector scheme to allow this term to be treated implicitly, even when we choose $\theta = 1/2$, the proposal will not always be accepted because the objective function is not generally quadratic.

Improved methods [22] of optimally choosing the value of the search distance would also improve the mixing rate.

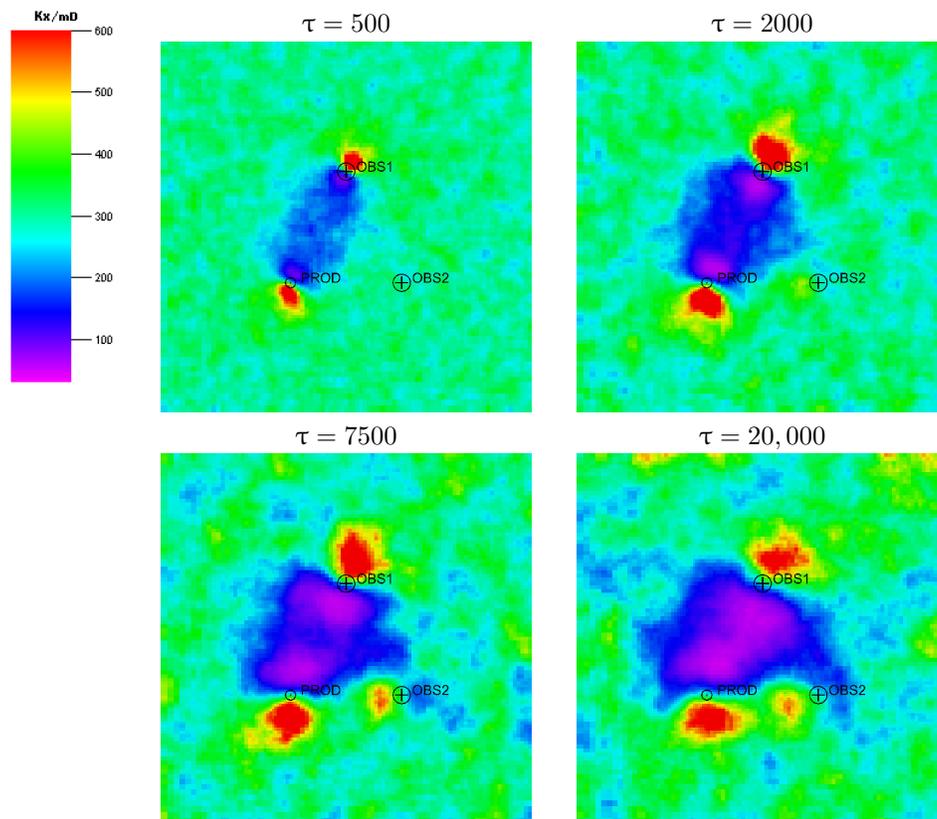


FIG. 4: Samples of the permeability distributions obtained from the initial stages of the Langevin method. In this period an increasingly good match with the measured pressure response is obtained, but it is not clear that independence from the initial state has been achieved.

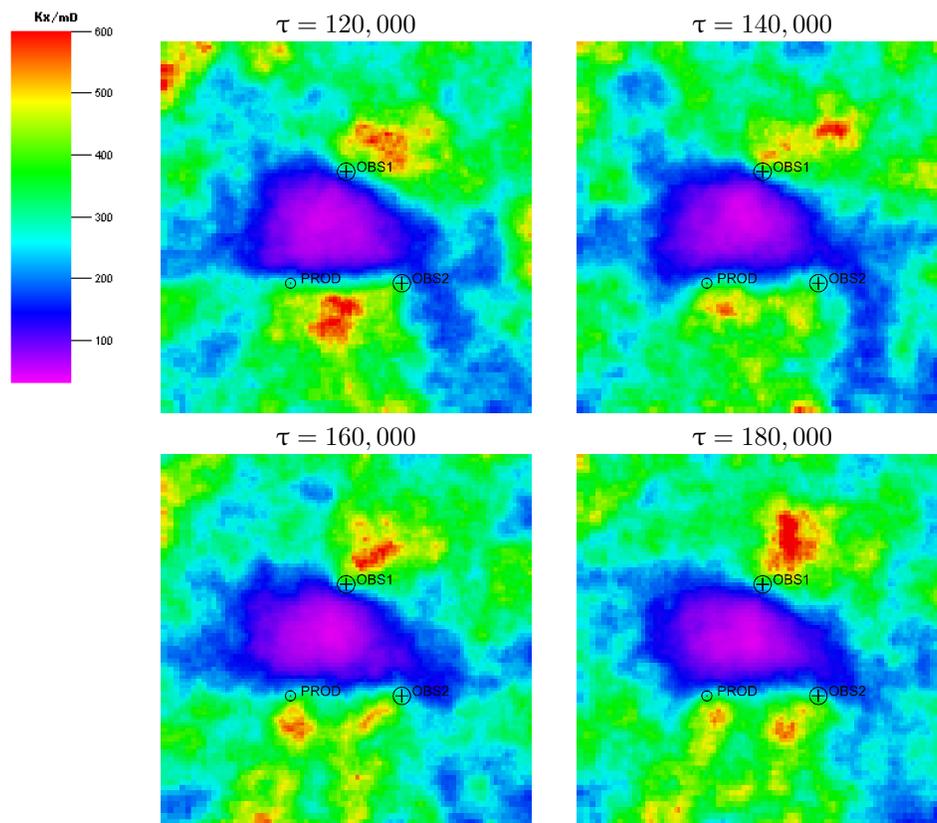


FIG. 5: Samples of the permeability distributions obtained from the later stages of the Langevin method. The match with the measured pressure response is maintained, samples are more likely to be representative of the stationary distribution, although mixing is still slow with strong correlation between successive samples.

6. CONCLUSIONS

In this paper we have developed methodologies for estimating the likely spatial variation of reservoir parameters based on pressure transient testing data from multiple well locations. The resolution of the grid over which reservoir parameters are discretized and, in particular, the resolution of the grid near the wellbore, need not be an obstacle to the matching of these reservoir parameters to pressure transient testing data. Provided that the adjoint method is employed, there is no increase in the numerical effort required to obtain the gradient as the number of parameters increases. Some prior information is always required, and as the resolution of the grid increases, the local correlation implied by the prior model becomes important. Prior models that have a strong local correlation will typically lead to stiff problems when gradient descent schemes or quasi-Newton schemes are used to search for likely posterior parameters. The stiffness can be overcome by using a semi-implicit scheme to integrate the prior gradient.

The information provided by the pressure transient test data is usually not sufficient to eliminate the uncertainty in the estimation of the reservoir parameters. We have shown that the Langevin approach may be used to directly produce realizations of the posterior distribution. The use of a semi-implicit scheme proved to be crucial for the Langevin method to work efficiently with a large number of parameters. However, more work is needed to further improve the rate of mixing of the Langevin method.

ACKNOWLEDGMENTS

We are grateful to Schlumberger and the Technical University of Istanbul for giving permission to publish this paper.

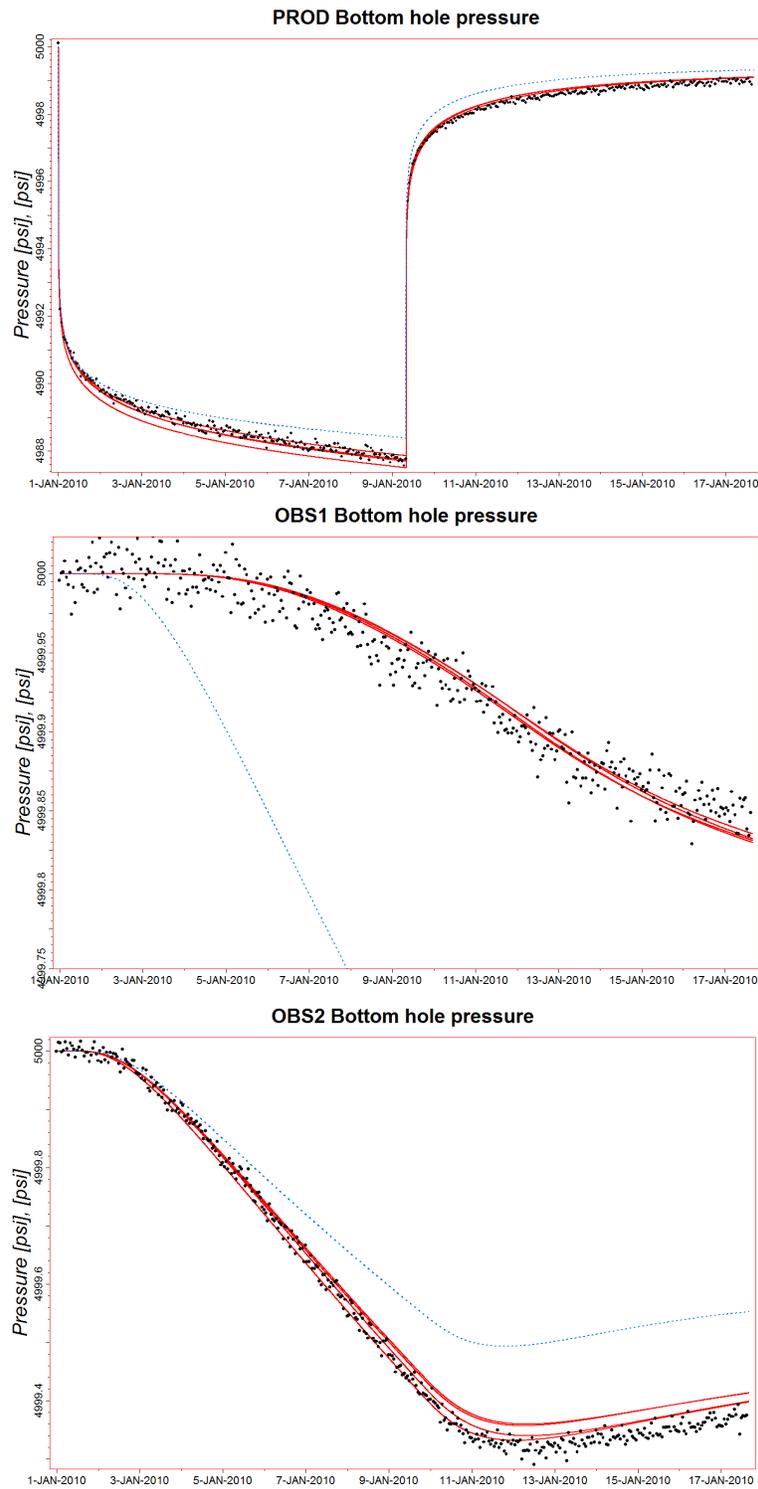


FIG. 6: Comparison of the observed pressure measurements (points) with the pressure responses generated by the reservoir models of Fig. 5 (solid lines). The pressure response of the initial homogeneous permeability distribution is shown as a dotted line.

REFERENCES

1. Earlougher, R., *Advances in Well Test Analysis*, 1st ed., Monograph Series 5, Society of Petroleum Engineers of AIME, Dallas, 1977.
2. Horner, D., Pressure buildup in wells, *Proc. of 3rd Petroleum Congress*, Brill, E. J., ed., Leiden, pp. 503–521, 1951.
3. Matthews, C. and Russell, D., *Pressure Buildup and Flow Tests in Wells*, 1st ed., Monograph Series 1, Society of Petroleum Engineers of AIME, Dallas, 1967.
4. Oliver, D. S., Estimation of radial permeability distribution from well test data, *SPE Form. Eval.*, 7:290–296, 1992.
5. McKinley, P. M., Vela, S., and Carlton, L. A., A field application of pulse-testing for detailed reservoir description, *J. Pet. Technol.*, 20(3):313–321, 1968.
6. Kuchuk, F. J. and Onur, M., Estimating permeability distribution from 3D interval pressure transient tests, *J. Pet. Sci. Eng.*, 39:5–27, 2003.
7. Backus, G. E. and Gilbert, J. F., Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. R. Astron. Soc.*, 13:247–276, 1967.
8. Oliver, D. S., Multiple realizations of the permeability from well test data, *SPE J.*, 1(2):145–154, 1996.
9. Oliver, D. S., Reynolds, A. C., Bi, X., and Abacioglu, Y., Integration of production data into reservoir models, *Pet. Geosci.*, 7:65–73, 2001.
10. Liu, N. and Oliver, D., Critical evaluation of the ensemble Kalman filter on history matching of geological facies, *SPE J.*, 8(6):470–477, 2005.
11. Li, G., Han, M., Banerjee, R., and Reynolds, A., Integration of well-test pressure data into heterogeneous geological reservoir models, *SPE J.*, 13(3):496–508, 2010.
12. Morton, K., Booth, R., Onur, M., and Kuchuk, F., Grid-based inversion methods for spatial feature identification and parameter estimation from pressure transient tests, Vienna, SPE EUROPEC, Paper No. 142996, 2011.
13. Farmer, C. L. Bayesian field theory applied to scattered data interpolation and inverse problems, *Algorithms for Approximation*, Iske, A. and Levesley, J., eds., Springer: Verlag, Berlin, pp. 147–166, 2007.
14. Peaceman, D., Interpretation of well-block pressures in numerical reservoir simulation with nonsquare grid blocks and anisotropic permeability, *SPE J.*, 23(3):531–543, 1983.
15. Nocedal, J. and Wright, S. J., *Numerical Optimization*, 2nd ed., Springer, New York, 2006.
16. Oliver, D. S., Reynolds, A. C., and Liu, N., *Inverse Theory For Petroleum Reservoir Characterization and History Matching*, Cambridge University Press, Cambridge, UK, 2008.
17. Dostert, P., Efendiev, Y., Hou, T. Y., and Luo, W., Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification, *J. Comput. Phys.*, 217:123–142, 2006.
18. Ma, X., Al-Harbi, M., Datta-Gupta, A., and Efendiev, Y., An efficient two-stage sampling method for uncertainty quantification in history matching geological models, *SPE J.*, 13(1):77–87, 2008.
19. Gardiner, C. W., *Handbook of Stochastic Methods*, 3rd ed., Springer, New York, 2004.
20. Beskos, A., Roberts, G., Stuart, A., and Voss, J., MCMC methods for diffusion bridges, *Stochas. Dyn.*, 3:319–350, 2008.
21. Hastings, W., Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57(1):97–109, 1970.
22. Roberts, G. O. and Rosenthal, J. S., Optimal scaling for various metropolis-hastings algorithms, *Stat. Sci.*, 16(4):351–367, 2001.