Abstract

We present and compare three different grid-based inversion methods for estimation of formation parameters and spatial geological feature identification based on pressure transient test (PTT) data from multiple-well locations. The first and second methods employ efficient adjoint schemes to determine the gradient of the objective functions resulting in the most likely set of reservoir parameters and an ensemble of updated realizations of the parameters, respectively. The second method is based on the Langevin equation. The third method uses ensemble Kalman filtering (EnKF) for data assimilation, in which the outcome is an ensemble of updated parameter realizations. These three methods use a grid-independent prior model (in view of the limited prior knowledge of the system expected to be available), described by as few parameters as possible, and consider a non-uniform grid with the highest resolution near the wells. With these methods, the existence of and location of many sub-seismic features such as strong spatial permeability variations, faults, fractures and pinch outs may be determined using exploration and production data. Such features may not be known a priori, particularly in the exploration of heterogeneous carbonate reservoirs.

We examine each method considering the degree of prior information required, the computational overhead and the applicability to the reservoir characterization workflow. Our results indicate that the first method provides a good history match to the observed PTT data and is suited for the early exploration phase of the reservoir. However, the parameters must be convolved with the smaller scale data to produce multiple realizations away from the implausibly smooth most-likely solution. The observed PTT data lies within the ensemble of predicted pressures in the EnKF and Langevin-based methods which are both applicable to probabilistic workflows where uncertainty is treated rigorously. However, EnKF seems to be computationally more efficient than the Langevin approach.

Introduction

Pressure transient testing (PTT) is a long established procedure for determining the productivity of a well and the properties of the formation (reservoir) from downhole and/or surface pressure and flow rate measurements. The main steps for interpretation are:

1. Model identification: a possible set of reservoir models are found that may fit the data,
2. Model parameter estimation: the model parameters are adjusted until the model behavior matches the observed data,
3. Model verification: the consistency of the final model is verified by measuring the mismatch between the real system and the model or by comparison with other data.

Using conventional interpretation methods (semilog methods such as Horner or MDH, and/or type-curve matching of measured pressure and/or derivative), reservoir pressure, an effective ‘average’ permeability of the formation, skin factor, wellbore storage, etc. can be estimated from the PTT data. In such interpretation, some sort of prior modeling is always necessary to constrain the non-linear parameter estimation because a model with many non-physical reservoir parameters may match the observed PTT data. This prior knowledge may be available at small scale from logs and cores and, at a larger scale, from seismic and outcrop analogies. Recently, non-linear least-squares optimization has been applied to pressure transient data using numerical models with a similarly limited number of parameters; often models that are divided into a small number of regions, within which the reservoir parameters are assumed to be constant.

As the need for more spatial resolution of the parameters increases, we turn to “pixel” methods where the physical properties of the reservoir are discretized on a pixel-like grid over the reservoir domain. Pixel or grid based methods have been applied for classical history matching on a field scale with multi-well data (Oliver 1996; Oliver et al. 2001). Based on
Bayesian estimation techniques and inverse problem theory (Tarantola 2005), the literature has focused on using geostatistics as prior knowledge to constrain the estimated parameters. Geostatistics, which essentially describes the spatial and static behavior of the petrophysical or geological property in question with a variogram or covariance matrix, can be used to create multiple realizations of the property on a grid structure. However, typically geostatistical simulation algorithms are applied to populate a regular grid of locations where the cells being simulated are at the same volume scale as the sampled data. To accurately simulate the pressure response of a pressure transient formation or well test, fine scale gridding is required around the wellbore where the information content of the PTT data is highest. To implement these methods we use a prior model which is grid-independent and (in view of the limited prior knowledge that we expect to be available) can be described by as few parameters as possible.

**Problem formulation**

All three methods for inversion presented in this paper are based on Bayes’ rule in its simplest form. The prior is the existing understanding of the reservoir before observation data are assimilated and the posterior update is the match between the observation data and the simulated response to a realization of the parameters. Data-bias is included and is generally treated as a constant for each measurement device. By updating the probability of the prior with the probability of the likelihood function, the posterior relative probabilities of parameters may be found. The method by which the update is made and how we describe the posterior probability distribution represents the difference between the three methods presented below.

In all three cases, the prior model is a probability density function \( \pi_0[u] \), which describes the likelihood of a particular set of parameters (reservoir or formation properties) where \( u \) is a vector that contains the unknown parameters (e.g. the spatially-variable porosity and components of permeability) representing the true state of the reservoir. The prior model can be written as

\[
\pi_0[u] \propto \exp(-R[u]),
\]

where \( R[u] \) is a quadratic function as required for Gaussian priors. The function \( R[u] \) is typically 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.” Given a particular set of parameters, the likelihood of a particular set of pressure measurements \( M \) being made (i.e. posterior update) is given by

\[
\pi[M|u] \propto \exp(-I[u]),
\]

where

\[
I[u] = \exp \left( \frac{-1}{2} \sum_{m=1}^{N_m} \sum_{r=1}^{N_{tr,m}} \frac{(p_{wm}(u,t_r) - \bar{p}_{wm}(t_r))^2}{\sigma_{wm}^2(t_r)} \right);
\]

where \( N_m \) and \( N_{tr,m} \) respectively the number of wells and pressure measurements for each well \( p_{wm}(u,t_r) \) and \( \bar{p}_{wm}(t_r) \) are respectively the simulated pressures (of the model described by parameters \( u \)) and the pressure measurements at time \( t_r \), and \( \sigma_{wm}^2(t_r) \) is the error variance of the measurements.

A forward model is required to determine the likelihood function given by Eq. 2. The forward or direct model is considered to be pressure diffusion for a slightly-compressible, single-phase fluid flow in a three-dimensional porous medium, where it is assumed that the viscosity is constant. The numerical grid does not capture the geometry of the well-bore, and therefore we apply Peaceman’s model (Peaceman 1983) that treats the well as a line source within the grid blocks penetrated by the well.

Using Bayes’ theorem we are able to deduce that the posterior distribution of the parameters as

\[
\pi[u|M] \propto \exp(-H[u]), \text{ where } H[u] = I[u] + R[u].
\]
The Gaussian prior $R[u]$ has been formulated several ways in the literature, most notably by using geostatistical methods (for example, see Oliver et al. 2008) where each parameter is described by a stationary Gaussian random function whose spatial variability is described by a variogram or covariance matrix, $C_M$, as shown below

$$R[u] = \left( -\frac{1}{2}(u - u_0)^T C_M^{-1}(u - u_0) \right). \quad (5)$$

The correlation coefficients of the matrix $C_M$ are assumed known, but may be zero. This formulation works particularly well where the spatial discretization is regular. However, for this formulation we require that the reservoir parameters are discretized using the same grid as that used for numerical simulation of the PTT. We therefore require fine gridding around the well with logarithmically or geometrically expanding cell sizes. Thus, the gridblocks used for the simulation have different cell volumes, and the coefficients of $C_M$ need to be volume averaged. In this case, we choose a prior that is independent of the grid. Equally, at the time of a PTT, very limited geostatistical information may be available and so ideally the prior is described by very few hyperparameters. Local Gaussian random fields can satisfy all of these requirements (Farmer 2007).

At this point, the three Methods diverge. The first and second methods employ efficient adjoint schemes to determine the gradient of the objective functions, and local Gaussian random fields for priors, resulting in the most likely set of reservoir parameters. Using the second method, that is based on the Langevin equation, an ensemble of realizations of the parameters may be generated as the most likely set of reservoir parameters is reached. The third method uses ensemble Kalman filtering (EnKF) for data assimilation with the outcome being an ensemble of updated parameter realizations that describe the posterior probability distribution. In the following sections, we first describe the prior for Methods 1 and 2, and then the three inversion Methods will be introduced.

**Prior Model for Methods 1 and 2.**

In this section we describe the prior model with local Gaussian random fields. In general, $u(\mathbf{x})$ would refer to a vector which may contain functions of spatially varying parameter fields e.g. permeability and porosity, and additional parameters e.g. skin, wellbore storage. The full derivation of the prior model for the three permeability components and a correlated porosity field is given in (Booth et al. 2010). In this paper, we will only consider a single spatially variable parameter (log-permeability field) which we will denote $u(\mathbf{x})$. We assume the following parameters are known: porosity, initial pressure, wellbore-storage coefficients and skin coefficients. We now give a brief outline of the prior for this case. For a parameter $u(\mathbf{x})$, where $\mathbf{x}$ is a spatial vector, a local Gaussian random field prior is described by

$$R[u] = \frac{1}{2} \int u L u d\mathbf{x}, \quad (6)$$

where $L$ is a linear differential operator. We choose the operator so that it provides Tikhonov regularization of the second, first and zeroth order. The associated second, first, and zeroth order hyperparameters (or coefficients) define a simple correlation function with correlation length given by $\rho$, the local variance of the parameter given by $\sigma^2$ and a structural parameter denoted by $\chi$.

$$L u = \frac{1}{4\pi \sqrt{2(\chi + 1)}} \rho^3 \sigma^2 \left( \rho^4 \nabla^4 u - 2\chi \rho^2 \nabla^2 u + u \right). \quad (7)$$

If we take $u$ to be the components of permeability, each component can be described as log-normally distributed so that

$$u_x = \log k_x - \log \overline{k_x}, \quad (8)$$

which can be modeled by a local Gaussian prior. With the parameters defined, the prior model may be constructed using Eq. 6 that describes the spatial correlation of the parameters and the local correlation between distinct parameters, i.e., permeability and porosity. In this case, we can describe the prior probability distribution as

$$\pi_0[u] \propto \exp \left( -\frac{1}{2} \int \frac{1}{4\pi \sqrt{2(\chi + 1)}} \rho^3 \sigma^2 u(\rho^4 \nabla^4 u - 2\chi \rho^2 \nabla^2 u + u) d\mathbf{x} \right). \quad (9)$$

**Method 1: Steepest Descent Method Using Adjoint State**

Equation 4 defines the posterior distribution function that can be determined by combining the prior model, Eq. 9 with the likelihood function, Eq. 2. It follows that the mode of the posterior distribution may be found by maximizing $\pi[u|M]$ or equivalently minimizing the objective function $H[u]$. 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. The gradient of \( R[u] \) can be easily determined yielding a linear function of \( u \); however, determining the gradient of \( I[u] \) is much more difficult. The most direct method of determining the 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. Furthermore this approach is highly sensitive to the size of the small variations which are made (Oliver 2008). Fortunately, we are able to obtain the gradient much more effectively by employing an adjoint method. This allows the gradient to be calculated with only a single evaluation of the forward model and a single evaluation of an adjoint model. The computation overhead of solving the adjoint problem is comparable to that of the forward problem.

The complete derivation of the adjoint model with respect to the adjoint variables to the pressure in the grid blocks and the wells, \( \lambda_i \) and \( \lambda_{in} \) may be found in (Booth et al. 2010). The adjoint model is almost identical to the forward model for the pressure; however, unlike the forward problem, it must be solved backwards in time starting at \( t = T \), after the last of the pressure measurements. Also, the flow rate from the well must be replaced by the mismatch between the measured values of the pressure in the wells and the values obtained by the forward model. When the pressure, \( p_i \) and its adjoint, \( \lambda_i \), are solved for, we can express the derivative of the likelihood function \( I[u] \) with respect to the isotropic permeability in grid block \( i \) (alternatively represented by \( (I,J,K) \) on a Cartesian grid) to be

\[
\frac{\partial I}{\partial \lambda_i} = \frac{1}{2} V_i \frac{T_{x+1/2,j,k}}{k_i} \int_0^T \left( p_{i+1,j,k} - p_{i,j,k} \right) \left( \lambda_{i+1,j,k} - \lambda_{i,j,k} \right) dt + \cdots, \tag{10}
\]

where \( V_i \) is the volume of the grid cell, \( T_{x+1/2,j,k} \) is the transmissibility between cells \( (i,j,k) \) and \( (i+1,j,k) \), and the other terms are similar expressions for each of the other faces of grid block \( i \). The sensitivity of the likelihood function to changes in the permeability in grid block \( i \) is affected by changes to the size of the grid block, and so a more appropriate property for irregular grids is the value in grid block \( i \) of the sensitivity of the likelihood function to the permeability field, given by

\[
\left( \frac{\partial I}{\partial \lambda_i} \right)_i = \frac{1}{V_i} \frac{\partial I}{\partial \lambda_i}. \tag{11}
\]

The usual form of steepest descent can be viewed as an explicit numerical approximation to the differential equation

\[
\frac{\partial u}{\partial \tau} = -\frac{\partial H}{\partial u}, \tag{12}
\]

which will converge to a local minima of \( H \) for large values of the convergence parameter \( \tau \). We implement a partially implicit numerical scheme for the steepest descent (Eq. 13) as

\[
u_{n+1} = u_n - \alpha_n \frac{\partial I}{\partial u} u_n - \alpha_n \theta \frac{\partial R}{\partial u} u_{n+1} - \alpha_n (1 - \theta) \frac{\partial R}{\partial u} u_n, \tag{13}\]

where \( \alpha_n \) is the step size and \( u_n \) is the estimate of the minimum after \( n \) iterations of steepest descent. The parameter of the numerical scheme, \( \theta \), lies between 0 and 1; we typically take the value 0.5. For sufficiently small values of \( \alpha_n \), 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 \( \alpha_n \) as large as possible. Although the scheme is not guaranteed to be unconditionally stable, typically much larger steps may be taken with an implicit scheme compared to the standard explicit steepest descent. To determine an optimal value of \( \alpha_n \), a line search is required. We ensure that the line search satisfies the Wolfe conditions (Nocedal and Wright 2006) which require that we calculate the derivative of \( H[u_{n+1}] \) with respect to \( \alpha_n \).

The algorithm requires a convergence criterion for termination. In this methodology, we apply a convergence criterion based on the value of the objective function, where \( a \) is a small number (e.g., \( 10^{-3} \)) expressing the degree of precision required, as

\[
\frac{H[u_{n+1}] - H[u_n]}{H[u_{n+1}]} < a. \tag{14}
\]

Steepest descent schemes such as this only provide information about the most likely realization of the reservoir parameters. This results in geological feature identification within the reservoir but we have no measure of the remaining uncertainty of the parameters or the shape of the posterior distribution function. Indeed, we should expect that our result will be subject to a great deal of uncertainty as there is insufficient data in the pressure transient response to fully characterize the formation or sector of a reservoir, where well tests are conducted.
One method of resolving this uncertainty is to approximate the second derivative of \( H[u] \) as part of the minimization of \( H \), giving a description of the likelihood of parameters which are close to the most likely set of parameters. With a large number of parameters to be determined, one of the most attractive ways of approximating this second derivative is via quasi-Newton methods that only require knowledge of the gradients of \( H \). These methods also improve the rate of convergence of steepest descent. A particularly attractive quasi-Newton method is L-BFGS (Nocedal and Wright 2006; Oliver et al. 2008) which only requires that a limited number of recent iterates and gradients are stored for the approximation of the second derivative.

With an approximation of the second derivative available, we can obtain realizations of a Gaussian approximation for the posterior probability distribution of the parameters. The actual construction of realizations from this distribution can be achieved by various means: a Cholesky decomposition of the Hessian matrix (which will be numerically expensive when the size of the matrix is large) or by principal component analysis. When the posterior probability distribution deviates significantly from the Gaussian ideal, this approach may not be appropriate. We therefore also look at a method for direct simulation of samples based on the Langevin equation as suggested in Farmer (2007).

**Method 2: Langevin Method**

The steepest descent method allows us to locate the most-likely case, even when the probability distribution described in Eq. 4 is highly nonlinear. However, it does not directly generate samples of the distribution. For highly nonlinear probability distributions, we therefore also consider generating samples via the Langevin equation. The Langevin equation is defined as,

\[
\frac{du}{d\tau} = -\frac{\partial H}{\partial u} + \sqrt{2}\eta(x, \tau),
\]

where \( \eta(x, \tau) \) represents white noise. For large values of \( \tau \), realizations of Eq. 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 \( \tau \)). In principle, the Langevin equation can be used to sample from arbitrarily complicated distributions, including multimodal distributions. The Langevin equation is very similar to Eq. 12, the continuous form of steepest descent, but with additional white noise added. The equation is discretized in a similar way to the implicit form of steepest descent shown in Eq. 13 as

\[
u_{n+1} = u_n - \alpha_n \frac{\partial H}{\partial u}[u_n] - \alpha_n \theta \Delta u_{n+1} - \alpha_n (1 - \theta) \Delta u_n + \sqrt{2\alpha_n} \eta(x),
\]

The introduction of random noise to the scheme prevents a line search being used to optimally choose \( \alpha_n \). Instead, a Metropolis-Hastings step is included to ensure that the \( r \)-step is small enough, and eliminate bias.

Convergence of the scheme will occur over two phases: we first have to run the scheme until feasible samples are produced, and then we should continue to run the scheme until these samples appear to give an adequate description of the posterior. The number of iterations required before the realizations of properties produced by the Langevin method are reliable samples from the posterior depends on: 1) how large \( \tau \) must be to ensure convergence; and 2) how small the steps \( \alpha_n \) must be to ensure stability of the numerical scheme (see Eq. 16). This will be explored with an example.

The previous two methods have relied upon describing the posterior probability function by finding its gradient and then using this to reach either the most-likely set of parameters (or the mode of the posterior distribution function) and then possibly also the shape of the function around the mode, or by using the gradient to guide a random sampling process. Ensemble Kalman filtering (EnKF), first proposed by Evensen (1994), provides an alternative methodology for data assimilation and is the focus of the third method that we introduce next.

**Method 3: Ensemble Kalman Filtering Method**

EnKF is a Monte Carlo method where uncertainty is represented by an ensemble of realizations. The prediction of the uncertainty is performed by ensemble integration using the forward model. A full derivation of the Ensemble Kalman filter methodology is provided in Evensen (2003) while an overview of the applications in the petroleum industry is given in Aanonsen et al. (2009). Our implementation of the EnKF in this work follows those of Zafari and Reynolds (2007) and Li et al. (2010). Here, we provide an overview of the algorithm developed for application to pressure transient tests, and its practical implementation.

The method consists of a forecast step (stepping forward in time) and an assimilation step in which parameters of the system are updated or corrected to honor the measured data. In the pressure transient case, the initial ensemble is an ensemble of realizations of the prior probability distribution function of a particular uncertain parameter (e.g., isotropic permeability) generated by random fields (i.e., we produce an ensemble of model states by adding some kind of perturbations to a best-guess estimate) and the dynamic variables that define the state of the system.

We use the following EnKF update equation for the \( j \)th ensemble member at time \( t_n \), where we have observed data (Li et al. 2010):
\[ y_{f}^{r,p} = y_{f}^{r,p} + C_{D,Dr}^{-1}(C_{D,Dr}^{-1} - d_{j}^{r,p}). \]  

for \( j = 1, 2, \ldots, N_{e} \), where \( N_{e} \) is the number of ensemble members. In Eq. 17, the superscripts \( u \) and \( p \) refer to updated and predicted quantities, respectively. The quantity \( y_{f}^{r} \) is an \( N_{y} \)-dimensional state vector for ensemble \( j \) at time \( t \), defined by

\[ y_{f}^{r} = \begin{bmatrix} u \\ p \end{bmatrix}. \]  

where \( u \) represents the \( N_{u} \)-dimensional vector of model parameters (e.g., natural logarithm of reservoir grid block permeabilities in our applications), \( N_{u} \) represents the total number of unknown model parameters, and \( p \) represent \( N_{p} \)-dimensional vector of gridblock and well pressures generated by our numerical simulator, \( N_{p} \) is the total number of gridblocks plus the number of wells. Note that \( N_{r} = N_{p} + N_{u} \). In Eq. 17, \( y_{f}^{r,p} \) refers to the forecast of the \( y_{f}^{r} \) i.e. the model steps forward to time \( t \), changing the value of \( p^{r} \) but not \( u \). At assimilation time \( t_{a} \), the number of observed data to be assimilated is \( N_{y} \), and the \( N_{y} \)-dimensional vector of perturbed observed data, \( d_{obs}^{y} \), is generated by adding Gaussian noise of mean zero and covariance \( C_{D} \) to the observed data. The \( N_{y} \)-dimensional vector \( d_{obs}^{y} \), contains the forecasted observed data for ensemble member \( j \) at the data assimilation time \( t_{a} \). In our formulation, we express the observed data to be \( p_{obs}(t_{a}) \) measured at time \( t_{a} \) with a known, uncorrelated Gaussian, error variance of \( \sigma_{wm}^{2}(t_{a}) \); therefore, the matrix \( C_{D} \) is a diagonal matrix. In Eq. 17, the covariance matrices \( C_{D,Dr} \) and \( C_{D,Dr}^{-1} \) are computed from the ensemble predictions by:

\[ C_{D,Dr}^{-1} = \frac{1}{(N_{e} - 1)} \sum_{j=1}^{N_{e}} (y_{f}^{r,p} - \bar{y}^{r,p})(d_{j}^{r,p} - \bar{d}^{r,p})^{T}, \]  

and

\[ C_{D,Dr} = \frac{1}{(N_{e} - 1)} \sum_{j=1}^{N_{e}} (\bar{d}^{r,p} - \bar{d}^{r,p})(d_{j}^{r,p} - \bar{d}^{r,p})^{T}, \]  

where \( \bar{d}^{r,p} \) and \( \bar{y}^{r,p} \) represent the ensemble averages of the predicted data and state vectors, computed from

\[ \bar{d}^{r,p} = \frac{1}{N_{e}} \sum_{j=1}^{N_{e}} d_{j}^{r,p}, \]  

and

\[ \bar{y}^{r,p} = \frac{1}{N_{e}} \sum_{j=1}^{N_{e}} y_{f}^{r,p}. \]  

Note that \( C_{D,Dr} \) is an \( N_{y} \times N_{y} \) matrix, and \( C_{D,Dr}^{-1} \) is an \( N_{y} \times N_{y} \) matrix; these are the same for all ensemble members. The nice feature of this EnKF implementation is that the covariance matrices do not need to be explicitly computed or stored. Further details of the EnKF method can be found in Evensen (2003), Aanonsen et al. (2009), and Li et al. (2010).

**Feature identification**

Adjoint gradient methods and EnKF data assimilation have been applied traditionally in clastic reservoirs (Haugan et al. 2008) to update realizations of the geostatistical model with the production and pressure data. However, features of carbonate reservoirs may not be well described by geostatistics (Corbett 2009) and discrete features such as faults and fractures may not be known a priori. Such features must be identified and placed within a geological or dynamic reservoir model for accurate prediction of future performance. In the following example, we explore the applicability of the three methods outlined in this paper for reservoir feature identification based on pressure transient data. The true model is intentionally made difficult to describe geostatistically.

A single-layer heterogeneous formation with 9801 gridblock \( (N_{x} = 99, \ N_{y} = 99 \ and \ N_{z} = 1) \) is created. A production well and 2 observation wells are placed as shown in Fig. 1 and the pressure response for the production well (named as PROD) flowing at 100 STB/D for 200 hrs followed by a buildup of 200 hrs is generated. The pressure response in the observation wells (named as OBS1 and OBS2) is also generated. Gaussian noise is added to the synthetic pressure data; the observation wells have identical noise levels with \( \sigma_{wm} = 0.00316 \) psi, and the production well data has noise level with \( \sigma_{wm} = 0.1 \) psi during drawdown and \( \sigma_{wm} = 0.0316 \) psi during build up. The number of observed (noisy) data points used for each well is identical and equal to 400. The total number of data points used in the analysis is 1200. Other pertinent information relating to the input data to the simulator is given below in Table 1.
Using the methodologies outlined previously, we seek to identify a low permeability (30 mD) strip lying between the production well, PROD, and the passive observation well, OBS1. The permeability is assumed to be 300 mD everywhere outside of the low permeability strip. The porosity distribution is assumed known and uniform. Fine gridding is created around the production well to accurately capture the pressure transient behavior at early time, Fig. 2.

<table>
<thead>
<tr>
<th>Table 1- Formation and fluid properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>h, ft</td>
</tr>
<tr>
<td>$c_{e}$, psi$^{-1}$</td>
</tr>
<tr>
<td>$t_{r}$, cp</td>
</tr>
<tr>
<td>$r_{w}$, ft</td>
</tr>
<tr>
<td>Skin, non-dimensional</td>
</tr>
<tr>
<td>$c$, RB/psi</td>
</tr>
<tr>
<td>$p_{min}$, psi</td>
</tr>
<tr>
<td>$p_{orig}$, psi</td>
</tr>
<tr>
<td>$B_o$, RB/STB</td>
</tr>
<tr>
<td>$\phi$, non-dimensional</td>
</tr>
</tbody>
</table>

**Feature Identification with Method 1: Steepest Descent.**

The permeability distribution used as a starting point for the steepest descent method is homogeneous with a permeability identical to the permeability away from the zone of low permeability. This permeability distribution is also used as the mode in the prior model.

Fig. 3 shows the most-likely reservoir parameters derived from the steepest descent method (Method 1) after 200 iterations. The most likely set of parameters is considerably smoother than the true model, while areas of the reservoir that are not sufficiently sensitive to the pressure response remain unchanged from the initial guess/prior mode. As shown in Fig. 4, the history match between the observed data with noise and the simulated pressure response for the most likely permeability set is excellent. It should be noted that the results shown in Figs. 3 and 4 are for the prior model of CASE 1 in Table 2 ($\sigma^2 = 0.0001$, $\rho = 300$, $\chi = 0$).
Method 1 was formulated to provide spatial feature identification where prior knowledge of the formation is limited, i.e. the prior model is defined by a local Gaussian field with four parameters. In exploration well testing, the prior knowledge may be limited with only very basic conceptual ideas about the length scales expected in the reservoir. To explore the impact of the prior on the most-likely permeability distribution and the rate of convergence, we run cases as outlined in Table 2 until convergence or 200 iterations. The value of the mismatch between the observed data and simulated case at iteration 200 is shown in the table below. After 200 iterations, the pressure mismatch becomes significantly smaller than the log-likelihood of the prior, as expected.

Table 2 - Prior model and value of objective function at 200 iterations

<table>
<thead>
<tr>
<th>CASE</th>
<th>$\sigma^2$ (local variance)</th>
<th>$\rho$, ft</th>
<th>$\chi$</th>
<th>Well pressure mismatch</th>
<th>Prior log-likelihood</th>
<th>Objective function</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASE 1</td>
<td>0.0001</td>
<td>300</td>
<td>0</td>
<td>2,250.23</td>
<td>7,936.16</td>
<td>10,186.39</td>
</tr>
<tr>
<td>CASE 2</td>
<td>0.000002</td>
<td>200</td>
<td>0</td>
<td>2,638.43</td>
<td>1,8615.30</td>
<td>21,253.73</td>
</tr>
<tr>
<td>CASE 3</td>
<td>0.000002</td>
<td>400</td>
<td>0</td>
<td>3,295.62</td>
<td>2,4620.90</td>
<td>27,916.52</td>
</tr>
<tr>
<td>CASE 4</td>
<td>0.00005</td>
<td>200</td>
<td>0</td>
<td>1,790.73</td>
<td>2,604.17</td>
<td>4,394.90</td>
</tr>
<tr>
<td>CASE 5</td>
<td>0.00005</td>
<td>400</td>
<td>0</td>
<td>1,907.42</td>
<td>3,161.97</td>
<td>5,069.39</td>
</tr>
<tr>
<td>CASE 6</td>
<td>0.00002</td>
<td>300</td>
<td>-0.9</td>
<td>2,816.55</td>
<td>20,493.20</td>
<td>23,309.75</td>
</tr>
<tr>
<td>CASE 7</td>
<td>0.00002</td>
<td>300</td>
<td>-0.9</td>
<td>1,784.03</td>
<td>28,452.00</td>
<td>5,235.87</td>
</tr>
<tr>
<td>CASE 8</td>
<td>0.00005</td>
<td>300</td>
<td>-0.9</td>
<td>1,814.54</td>
<td>34,51.84</td>
<td>4,659.06</td>
</tr>
<tr>
<td>CASE 9</td>
<td>0.00005</td>
<td>300</td>
<td>-0.9</td>
<td>2,374.52</td>
<td>7,449.16</td>
<td>9,823.68</td>
</tr>
<tr>
<td>CASE 10</td>
<td>0.00001</td>
<td>200</td>
<td>-0.9</td>
<td>2,374.52</td>
<td>7,449.16</td>
<td>9,823.68</td>
</tr>
<tr>
<td>CASE 11</td>
<td>0.00001</td>
<td>200</td>
<td>-0.9</td>
<td>1,955.25</td>
<td>6,511.70</td>
<td>8,466.95</td>
</tr>
<tr>
<td>CASE 12</td>
<td>0.00001</td>
<td>400</td>
<td>-0.9</td>
<td>2,547.69</td>
<td>13,772.00</td>
<td>16,319.69</td>
</tr>
<tr>
<td>CASE 13</td>
<td>0.00001</td>
<td>400</td>
<td>-0.9</td>
<td>1,989.19</td>
<td>6,617.50</td>
<td>8,606.69</td>
</tr>
</tbody>
</table>

To provide a visual indicator of how the local Gaussian field prior model relates to a spatial permeability field, we show samples of the prior from: the control CASE 1, the least variable prior CASE 3 and the most variable prior CASE 4 in Fig. 5 below. We also show samples of the prior for CASE 10, CASE 11, and CASE 13 to illustrate the impact of the structural parameter, $\chi$. Notice the altered color scheme to highlight the variability of the parameter.
The rate of convergence for the cases (percentage change) is shown in Fig. 6. Most cases converge smoothly after 20 iterations. However, where the prior model is very strong, i.e. $\sigma^2 = 0.00005$, and particularly if the local correlation parameter, $\rho$, is low as in CASE 4, we see that the steepest descent method struggles to converge smoothly.

The most likely distribution for isotropic permeability for each case is shown in Fig. 7. CASE 1 represents the mode case. In general, the local variance parameter, $\sigma^2$, has the most impact on convergence and goodness of fit of the end result (compare CASE 2 and CASE 4; CASE 6 and CASE 8; and, CASE 7 and CASE 9). In each pair, the lower variance results in a smoother permeability distribution. Local variance reflects the degree of certainty in the prior mode. The higher certainty placed on the prior mode value results in less variability in most-likely parameters and less overshoot or zones of high permeability created away from the well (compare CASE 6 and CASE 8). Longer correlation lengths allow for the identification of features (strip) away from the well but result in increasingly smeared picture of the feature (compare CASE 2 and CASE 3; CASE 4 and CASE 5; CASE 10 and CASE 12; and, CASE 11 and CASE 13). However, the effect appears less noticeable than the impact of the local variation (e.g. CASE 4 and CASE 5 are visually similar and have a similar prior mismatch).

The impact of the structural parameter is subtle. The structural parameter $\chi$ describes the shape of the variogram: $\chi = +0.9$ approaches the spatial variation shown in an exponential variogram, while $\chi = -0.9$ is closer to a Gaussian variogram with an almost constant covariance for small separations but with a distance beyond which the covariance is no longer positive, and $\chi = 0$ represents an intermediate state with a decaying covariance and a cutoff separation. Negative values of $\chi$ appear to lead to smoother permeability distributions than positive values (compare CASE 6 and CASE 7; CASE 8 and CASE 9; CASE 10 and CASE 11; and, CASE 12 and CASE 13). The orientation of the strip is also better captured by those priors with negative values of $\chi$, particularly in CASE 10 and CASE 12.

Those priors with $\chi = -0.9$ exhibit a degree of oscillation in the variogram which can induce equivalent oscillations in the permeability field. The wavelength of the oscillation depends upon $\rho$. Although oscillation in the most-likely distribution is apparent in CASE 4, CASE 5, CASE 8, and CASE 9, this is as a result of the larger local variance.

The history matches obtained between the observed and simulated pressure data for the most likely permeability set for each case, Fig. 8, are generally good. Every case shows a bias for observation well OBS1 as with the hyperparameter mode case (Fig. 4). This bias indicates that the prior is not entirely sufficient to model the true reservoir parameters, i.e. the true feature is too sharp to lie within the variability of the necessarily smooth prior. Indeed CASE 6, which has the most highly
weighted prior and a smooth value of the structural parameter, has the poorest prior model match and a high remaining pressure mismatch (see Table 2). CASE 3 is also similar.
Fig. 7 – The “most-likely” permeability distribution determined using hyperparameters given in Table 2.

The method is successful for identifying reservoir features in a qualitative sense. A simple feature can be identified without overreliance on a complex prior model. It is impossible to define a feature with sharp edges using the technique applied but the simple sensitivity study suggests that there is no need to be dogmatic in the selection of the prior.

The computational overhead for Method 1 is related to the number of forward/adjoint simulations that have to be run. In these examples, around 800 forward/adjoint simulations were required for 200 iterations (the extra simulations were required for the line search).

Feature Identification with Method 2: Langevin Method.

The permeability distribution used as a starting point for the Langevin method is homogeneous with permeability identical to the permeability away from the zone of low permeability. This permeability distribution is also used as the mode in our prior model and we apply the same hyperparameters as in CASE 1 from Table 2. The Langevin method took far more iterations to converge with a reasonable fit to the pressure data than steepest descent, Method 1. This is primarily because a simple line-search scheme is not available due to the presence of random noise. It is known that the Metropolis-Hastings method, which we use in place of the line search is extremely sensitive to scaling, i.e. one must have a good estimate of the search distance beforehand. Further investigation of methods to determine appropriate scaling is required.

Fig. 9 shows a sample from the permeability distribution, which appears similar to the results from steepest descent. As can be seen in Fig. 10, there is also a similar fit with the observed (synthetic) pressure measurements. By close inspection of the background permeability field away from the wells, where the pressure measurements have not substantially altered our prior permeability field, we can see in Fig. 11 that the Langevin method has preserved the properties of the prior permeability field.

Once a reasonable fit with the observed pressure data is found, further samples are generated that all honor the pressure transient data. However, the samples remain strongly correlated for several iterations ("slow mixing"). Simultaneous approximation of the second derivative, as suggested for the steepest descent method, might allow for a higher order approximation of Eq. 15 and provide faster mixing.
The Langevin method, Method 2, is as appropriate for identifying the strip as in Method 1. The method is slower to converge and the resulting realizations can be overly similar to each other (Fig. 11). The slow mixing rate is reinforced by the small $\tau$-steps that must be made to identify the feature. With a stronger prior, mixing starts at an earlier value of $\tau$ and therefore the produced realizations are better mixed after fewer iterations. This methodology benefits from a more confident prior. The Langevin method provides an exciting technique for combining a feature with a background of properties that can be describe in a similar method to standard geostatistics.

**Feature Identification with Method 3: Ensemble Kalman Filtering Method.**

Finally, we try to identify the permeability heterogeneity using the EnKF method, Method 3. Our initial numerical experiments used the low variance prior model of CASE 1 that we used for Method 1 and Method 2. The results from this experiment failed to identify the feature, and the history matches obtained for all wells were consequently poor. To make a fairer test, we
provide a prior ensemble which contains sufficient variability to capture the true model. This would be more appropriate for feature location rather than identification, i.e. the feature is expected but the exact location is unknown.

We start with a log-normal prior permeability distribution with mean = 300 mD, std-dev = 1 (ln-space). The spatial variability of the field is isotropic with a Gaussian variogram, nugget = 0.01 and range = 500 ft. This is qualitatively similar and has a similar correlation length to the Method 1 and 2 control case (CASE 1). One hundred realizations are made of the permeability field, three of the ensemble realizations are shown in Fig. 12. The pressure response of the wells for each of the realizations before updating is shown in Fig. 13 with the observed pressure response of the well from the true case. The prior realizations do not provide an adequate match to the pressure history, particularly for observation well OBS1.

Fig. 14 shows the match of the pressure data during sequential assimilation with EnKF. We note that the match with the observed data during assimilation is very good.

Fig. 12 – Samples of 100 ensemble members for an isotropic, spatially correlated permeability field with a Gaussian variogram.

Fig. 13 – Computed well pressure responses (grey lines) vs. time for the 100 realizations similar to those shown in Fig. 12 and observed well pressure data (circular points). The mean of the initial ensembles is also shown (pink line).

Fig. 14 – Observed (circular data points) pressures and predicted (grey lines) pressures during sequential data assimilation vs. time for the 100 realizations of the isotropic permeability field with a Gaussian variogram.

The ensemble mean of the updated realizations is shown in Fig. 15. The low permeability strip is identified in the ensemble mean lying between PROD and OBS1. As with the steepest descent method, the size of the strip is determined by the correlation length of the prior model. EnKF, unlike steepest descent, continues to honor the geostatistical model in areas that
are not updated by the pressure data. Samples of the updated ensemble members are shown in Fig. 16. Here we see that there is some degree of correlation between the outputs. We also observe that the posterior ensemble mean and the samples of the posterior ensemble members are more variable than the prior ensemble. This may be a result of noise in the observed data (and thus the random perturbation applied to the data) and/or insufficient number of ensembles used.

Fig. 17 shows the predictions of the wellbore pressure obtained by rerunning the forward model from time zero using the final ensembles of predicted (or updated) permeabilities obtained after all the test data have been assimilated. Comparison of the pressure response during assimilation and after the assimilated models have been rerun (Fig. 14 and Fig. 17) shows the degree to which previously assimilated data is down-weighted as new data points are assimilated. The true, observed data lies within the spread of the rerun posterior ensemble and lies close to the mean of the ensemble pressure (pink line in Fig. 17). The level of noise in the ensemble reruns is much larger than in the true model and much larger than in the assimilated pressure responses; therefore, while on average the ensemble gives a good match to the truth, the individual responses can be poor. The mismatch between the rerun ensemble pressures and the true observed data of Fig. 17 is perhaps not surprising as a good match is only guaranteed if the problem is linear, the ensemble size is sufficient, and the prior variogram used to create initial ensembles of the permeability is consistent with the true one.
In Method 1 we also investigated the effect of changing the shape of the prior variogram. We repeat this experiment with Method 3 by using an exponential variogram for the prior. The prior permeability distribution has an exponential variogram with nugget $= 0.0001$ and isotropic range $= 500$ ft. As before, mean $= 300$ mD, and std-dev $= 1$ (ln-space). This is a limiting case of the prior used in Method 1, when $\chi$ tends to 1. One hundred realizations are made of the permeability field, three of the ensemble realizations are shown in Fig. 22. As in the previous case, the prior realizations do not provide an adequate match to the pressure history.

During assimilation with EnKF, the match of the pressure data to the observed pressure was, once again, very good confirming that the EnKF algorithm was working correctly. The ensemble mean of the updated realizations is shown in Fig. 19. The strip is identified. Samples of the updated ensemble members are shown in Fig. 20. Compared to the previous Gaussian prior example we see much stronger correlation between the samples. Yet again, the posterior ensemble mean and the posterior ensemble members are more variable than those of the prior ensemble.

Fig. 21 shows the predictions of the wellbore pressure obtained by rerunning the forward model from time zero using the final ensembles of predicted (or updated) permeabilities obtained after all the test data have been assimilated. In our case, almost all of the predicted pressures for the production well and observation well 2 lie to one side of the observed data; the ensemble is underestimating the degree of connectivity between the wells. Perhaps the relatively small ensemble size required to be computationally attractive may lead to an inaccurate representation of the covariance matrix of the ensemble prior which could result in a biased history match result. Our results suggest that a larger ensemble size is required to adequately describe a prior with an exponential variogram than one with a Gaussian variogram.
Fig. 20 – Three samples of 100 updated ensemble members of those shown in Fig. 18.

Fig. 21 – Comparison of the observed (synthetic) pressure data with those generated by the updated ensembles (grey lines) running the simulator from time zero and the ensemble mean of the updated ensembles (pink line).

The previous tests highlighted a number of problems: 1) realistic measurement errors may introduce additional variability into a highly variable permeability distribution and 2) the strip is not well identified when the initial ensemble of permeability is inconsistent with the true one. We test the impact of data noise and prior selection by running a second EnKF example.

In this test, our prior model is an anisotropic, spatially correlation log-normal permeability field. As before, the permeability mean = 300 mD, std-dev = 1 (ln-space). The spatial variability of the field is anisotropic with an exponential variogram, nugget = 0.0001, minor range = 500ft (y-direction) and major range (x-direction) = 1500 ft (roughly 3/4 of the strip length). The ensemble is formed of 100 realization, samples of these realizations are shown in Fig. 22. We reduce the applied Gaussian noise in the synthetic well pressures to $\sigma_{wm} = 0.000316$ psi for the observation wells, and $\sigma_{wm} = 0.01$ psi for the production well data.

Fig. 22 – Three samples of 100 ensemble members for an anisotropic, spatially correlated permeability field with an exponential variogram.

The mean of the updated realizations is shown in Fig. 23 where we compare with the true model. The strip is more obviously identified between PROD and OBS1 and is maintained in the conditioned realizations as shown in Fig. 24. However, we notice some similarity between the output samples as we observed in the previous isotropic test. Immediately, by
comparing Figs 22 and 24, we observe that variability in the updated ensemble is similar to the prior ensemble due to the low level of measurement noise in the observed data. As with the isotropic case, we rerun the forward model from time zero using the final ensemble of updated permeabilities obtained after all test data have been assimilated, Fig. 25. For this prior, we see the ensemble history match is particularly poor for observation well 1 (OBS1). We would probably obtain better results using 200 or 300 initial ensemble members.

Fig. 23 – Permeability field of the true model (left) and ensemble mean (right) of updated permeability fields for the anisotropic case with an exponential variogram.

Fig. 24 – Three samples of 100 updated ensemble members of those shown in Fig. 22.

Fig. 25 – Comparison of the observed (synthetic) pressure data with those generated by the updated ensembles (grey lines) running the simulator from time zero and the ensemble mean of the updated ensembles (pink line).

The computational overhead for Method 1 and Method 2 is related to the number of forward/adjoint simulations made. The number of simulations is determined by the number of iterations required to converge and how tight we wish to make the line search (or equivalent in Method 2). In the Method 1 examples, around 800 simulations were required for 200 iterations. For the Langevin method, 4000 simulations were required for 1000 iterations (necessary for sufficient mixing). For EnKF, the computational overhead is related to number of ensemble members. For the EnKF examples, we had 100 ensemble members suggesting an 8-fold improvement in speed over Method 1. However, we have also rerun the 100 realizations from the start to
check the assimilation and the matrix calculations also require significant overhead. Our experience has indicated that EnKF, even with reruns, is twice as fast as the steepest descent method.

From the above applications of the EnKF method, it can be stated that the EnKF method looks promising, particularly due to its superior computational performance compared to Methods 1 and 2, for history matching of spatially distributed pressure transient test data sets, but the number of ensembles to be used, and the prior variogram to be chosen to generate ensembles seem to be critical for successful feature identification and feature location with such data sets. Although we have not considered it in this work, the EnKF methods using iterative and/or covariance localization approaches may also improve the results (see for example, Aanonsen et al. 2009 for such methods).

Conclusions
In this paper we have introduced three methods for grid-based parameter estimation. The focus has been to test the methods for spatial feature identification in a relatively unknown reservoir (e.g. exploration well testing). Method 1 differs from standard gradient-based optimization problems through the use of a local Gaussian prior rather than the full covariance matrix required in many of these schemes. This reduces the requirement for a full geostatistical model and regular gridding. Method 2 introduced a novel technique based on the Langevin equation for sampling the posterior distribution function by adding small random noise to create realizations around the most-likely estimate. This technique is attractive in that it allows the background variability, normally lost during gradient-based optimization, to be reintroduced. Method 3 employed a well established technique, Ensemble Kalman Filtering, for a novel purpose: non-Gaussian feature identification.

Our results indicate that the first method provides a good history match to the observed PTT data and a qualitative indication of features within the reservoir from very sparse prior knowledge. The identification of the feature was essentially prior-independent. The most-likely estimate of reservoir properties is implausibly smooth (as expected for the most likely solution) and the output would not be used in a reservoir modeling workflow except in a qualitative sense.

The Langevin method is essentially an extension to the adjoint gradient technique and identifies features equally well. The methodology allows for multiple realizations of the pressure conditioned posterior pdf. However, the posterior pdf is generated from a non-stationary sequence of samples and further assimilation with geological data at a later stage may be difficult. In its current state, a large number of iterations are required for both feature identification and mixing and this needs to be resolved before it becomes a viable method for field data analysis.

The EnKF method is computationally more efficient than the adjoint gradient based steepest descent and related Langevin approach. The test feature could be identified from the pressure data but the output posterior distribution and simulated pressure was more sensitive to the prior distribution. Not only does the value chosen for the variance of the prior permeability field appear to be crucial for successful feature identification, but also the shape of the prior variogram affected the success of the history match. We suspect that this result may be because a larger number of ensemble members are required to adequately describe a prior with an exponential variogram than one with a Gaussian variogram. One of the most attractive features about the EnKF methodology is the ability for additional information to be assimilated rigorously. Method 3 is recommended for probabilistic workflows and where the prior distribution of parameters is well defined.

Nomenclature

- \( C \) = wellbore storage
- \( C_M \) = covariance matrix
- \( c_t \) = total compressibility
- \( H \) = log-likelihood of posterior
- \( h \) = model thickness
- \( I \) = log-likelihood of parameters given data
- \( N_w \) = number of wells
- \( N_{m,s} \) = number of pressure measurements in well \( m \)
- \( N_e \) = number of ensemble members
- \( p_{\text{true}}(u; t) \) = simulated pressures (of the model described by parameters \( u \)) at time \( t \)
- \( p_{\text{true}}(t) \) = measured pressures at time \( t \)
- \( \phi \) = porosity
- \( p_o \) = initial reservoir pressure
- \( p_{\text{true}} \) = initial wellbore pressure
- \( \rho \) = correlation length for simple correlation function
- \( R \) = log-likelihood of prior
- \( r_w \) = wellbore radius
- \( \chi \) = structural parameter for simple correlation function
- \( \sigma^2 \) = local variance for simple correlation function
- \( \sigma_{\text{true}}^2(t) \) = error variance of the measurement
- \( u \) or \( \mu \) = a vector which may contain spatially variable functions or a single spatially variable parameter
- \( \mu \) = viscosity
References