Mathematical Geology, Vol. 36, No. 2, February 2004 (© 2004)

Using Bayesian Statistics to Capture the Effects of Modelling Errors in Inverse Problems[1]

J. N. Carter[2]

When the parameters of a numerical model are adjusted, so that the predictions of the model match measurements from the real system, we need to take account of two sources of errors. These being measurement errors and modelling errors. Measurement errors are commonly considered, and a number of different approaches are in general usage, the most common being the weighted sum of squares method. In this paper the standard Bayesian equation, used for inverse problems, is reformulated so as to make it more intuitive to use. This allows the inclusion of both a modelling error and correlations between measurements to be carried out easily. The results are tested on a simple one-parameter numerical model and a cross-sectional model of a petroleum reservoir. In the first case the proposed error model appears to work well. In the second case it appears that the objective function is multimodal, leading to multiple acceptable solutions. The results of this paper are important to those whose numerical models are thought to contain significant modelling error. This encompasses many areas of modelling related to earth science and engineering.

KEY WORDS: likelihood, uncertainty, sum of squares, Bayesian analysis.

INTRODUCTION

Any numerical model of a physical system is, in some sense, only an approximation. For example the model will often require simplifying assumptions, in which detailed descriptions are replaced by averaged descriptions. What matters is how well the predictions of the model compare with the behavior of the real system. In other words, what is the uncertainty associated with each prediction?

The uncertainty in a prediction ultimately derives from one of three sources: the simplifying assumptions used in the model; the measurements used to calibrate the free parameters of the model; and the numerical schemes used to solve the equations that constitute the model. The first and last of these are usually associated with modelling error, the second is usually associated with measurement error. Measurement error is generally well understood and quantified. Modelling

[1]Received 18 March 2002; accepted 17 July 2003.

[2]Department of Earth Sciences and Engineering, Imperial College, London SW7 2AZ, United Kingdom; e-mail: j.n.carter@imperial.ac.uk

errors however are almost impossible to quantify. For this reason it has become common practice among many authors to only deal with the measurement errors (e.g. Floris and others, 2001). The classic book that describes the application of Bayesian statistics to inverse problems is by Tarantola (Tarantola, 1987). In which the necessary functional forms are described for a range of measurement and modelling errors. The weakness of the book appears to be that no method is given to obtain an estimate of the parameters that describe the modelling error, or correlations among the measurement data. There have recently been a number of papers that attempt to quantify the modelling error (DeVolder and others, 2002; Glimm and others, 2001), and the correlation among measurements (Aanonsen and others, 2002; Hauge, Arntzen, and Soleng, 2002; Wu, Reynolds, and Oliver, 1999).

In this paper the standard Bayesian formulation of the inverse problem, as described by Tarantola, is reformulated in an iterative form. This then allows us to construct estimates of both the modelling error and correlations among the measurements.

The results of the calculations are tested on two simple models: the first is a simple curve fitting problem involving the determination of one parameter; the second is a cross-sectional model of an oil reservoir, which involves a sequence of alternating good and poor quality sands and a fault. The unknowns are the sand permeabilities and the throw on the fault.

BAYESIAN ANALYSIS

In a Bayesian framework we wish to calculate the following:

$$P(\underline{m} \mid \underline{\mu}_o^n, \underline{\sigma}^n, M) = \int P(\underline{m}, \underline{\mu}_t^n \mid \underline{\mu}_o^n, \underline{\sigma}^n, M) \mathrm{d}\underline{\mu}_t^n \tag{1}$$

Where $P(. \mid .)$ is a probability distribution function (pdf) that describes our beliefs about a set of parameters given some information. The parameters for which we wish to define the pdf are $\underline{m}$, and the model to which they relate is M. This model includes all of the information necessary to take an instance of the parameters, $\underline{m}$, and to generate a set of predictions for the real system behaviour. The combination of an instance of $\underline{m}$ and M will be referred to as a realization of the model. The model includes things like: the algorithms used to construct the numerical model from the parameters; controls that are used on the real system; the equations that need to be solved; and the numerical algorithms that are used. $\underline{\mu}_o^n$ are the observed measurements, and $\underline{\sigma}^n$ are parameters that define the uncertainty in the associated unobserved true measurements, $\underline{\mu}_t^n$. It could be argued that $\underline{\sigma}^n$ is not needed until a model for the measurement error is introduced. We prefer to include $\underline{\sigma}^n$ separately as this enables us to be clear as to what information is available when defining a pdf,

it is also assumed throughout that the statistical model describing the uncertainty generated by the measuring device is known. The notation used is defined in the nomenclature.

Using a combination of the product rule, Bayes rule, and the following simplifications:

- $P(\underline{m} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M) = P(\underline{m} \mid \underline{\mu}_{\mathrm{t}}^{n}, M)$, $\underline{\mu}_{\mathrm{o}}^{n}$ is an estimate of $\underline{\mu}_{\mathrm{t}}^{n}$ and so can be disregarded when $\underline{\mu}_{\mathrm{t}}^{n}$ is available.
- $P(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\sigma}^{n}, M) = P(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\sigma}^{n})$, the measurement error model does not depend on the system model.
- $P(\underline{\mu}_{\mathrm{t}}^{n} \mid \underline{\sigma}^{n}, M) = P(\underline{\mu}_{\mathrm{t}}^{n} \mid M)$, our prediction of $\underline{\mu}_{\mathrm{t}}^{n}$ does not depend on the measurement device.

We can write, as shown in Appendix A,

$$P(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M) = P(\underline{m} \mid M) \int \frac{P\left(\underline{\mu}_{\mathrm{t}}^{n} \mid \underline{m}, M\right) P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\sigma}^{n}\right)}{P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\sigma}^{n}, M\right)} \mathrm{d}\underline{\mu}_{\mathrm{t}}^{n} \qquad (2)$$

This equation is equivalent to that given by Tarantola (Tarantola, 1987, Eq. 6, page 61). The problem in using this equation can be the difficulty in defining the correlations within the pdf's, and carrying out the multidimensional integrations.

If we wish to include an additional measurement, and its uncertainty, $\mu_{\mathrm{t},(n+1)}$ and $\sigma_{(n+1)}$, then

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M\right) = \int P\left(\underline{m}, \mu_{\mathrm{t},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{o},(n+1)}, \sigma_{(n+1)}, M\right) \mathrm{d}\mu_{\mathrm{t},(n+1)} \qquad (3)$$

using a combination of the product rule, Bayes rule, and the following simplifications:

- Since $\mu_{\mathrm{o},(n+1)}$ is an estimate of $\mu_{\mathrm{t},(n+1)}$, then it can be disregarded when $\mu_{\mathrm{t},(n+1)}$ is available, hence

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{t},(n+1)}, \mu_{\mathrm{o},(n+1)}, \sigma_{(n+1)}, M\right) = P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{t},(n+1)}, M\right)$$

- Knowing the measurement device to be used does not help us to estimate a value for $\mu_{\mathrm{t},(n+1)}$, therefore

$$P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \sigma_{(n+1)}, M\right) = P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right)$$

- The measurement error model does not depend on previous measurements or the system model, so

$$P\left(\mu_{o,(n+1)} \mid \underline{\mu}_o^n, \underline{\sigma}^n, \mu_{t,(n+1)}, \sigma_{(n+1)}, M\right) = P\left(\mu_{o,(n+1)} \mid \mu_{t,(n+1)}, \sigma_{(n+1)}\right)$$

We can obtain, as shown in Appendix B,

$$P\left(\underline{m} \mid \underline{\mu}_o^{n+1}, \underline{\sigma}^{n+1}, M\right) = P(\underline{m} \mid M) \prod_{i=1}^{n+1} Q_i \tag{4}$$

where

$$Q_i = \int \frac{P\left(\mu_{t,(i)} \mid \underline{m}, \underline{\mu}_o^{i-1}, \underline{\sigma}^{i-1}, M\right) P\left(\mu_{o,(i)} \mid \mu_{t,(i)}, \sigma_{(i)}\right)}{P\left(\mu_{o,(i)} \mid \underline{\mu}_o^{i-1}, \underline{\sigma}^{i-1}, \sigma_{(i)}, M\right)} \mathrm{d}\mu_{t,(i)} \tag{5}$$

By comparing Equations 2 and 4, we can see that the multidimensional integral involving pdf's, that may be correlated, has been reduced to a product of one-dimensional integrals. This has been achieved using three very limited simplifications, which means that the final result can be considered as equivalent to the original version.

This change is important for practical problems where the number of measurements may be in the 100's to 1000's. Carrying out these high-dimensional integrals, as required by Equation 2, is difficult, as the pdf's may be correlated. In the reformulated version this problem has been circumvented. The consequence of this modification will be considered later.

DEFINITION OF THE PDF'S

Each of the integrals, Q_i, defined by Equation 5 contain three pdf's whose functional form needs to be selected before we can make use of Equation 4. In this section we examine each of the pdf's in turn and consider possible functional forms. This will then allow us to test the proposed methodology in a later section. Observation a prior pdf:

$$P\left(\mu_{o,(i)} \mid \mu_{t,(i)}, \sigma_{(i)}\right)$$

This is the measurement error. It is commonly assumed to be Gaussian

$$P\left(\mu_{o,(i)} \mid \mu_{t,(i)}, \sigma_{(i)}\right) = \frac{1}{\sqrt{2\pi}\sigma_{(i)}} \exp\left(-\frac{\left(\mu_{o,(i)} - \mu_{t,(i)}\right)^2}{2\sigma_{(i)}^2}\right) \tag{6}$$

$$P\left(\mu_{o,(i)} \mid \underline{\mu}_o^{i-1}, \underline{\sigma}^{i-1}, \sigma_{(i)}, M\right)$$

We are allowed to use: all the previous measurements; we know the measurement device and its statistics; we also know the system model M, but not the system

model parameters $\underline{m}$. In setting this pdf we can use knowledge about the system, such as when wells are shut-in, as well as more general knowledge.

Since this term is independent of both $\underline{m}$ and $\mu_{\mathrm{t},(i)}$, it will only appear as a constant in any calculations. Provided that the measurement is not considered completely implausible, then it will not have any effect on the analysis. The simplest assumption that we can make is the pdf is a uniform distribution over some range. Modelling errors pdf:

$$P\left(\mu_{\mathrm{o},(i)} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, \sigma_{(i)}, M\right) = \begin{cases} a_i & \mu_{i\,\min} < \mu_{\mathrm{o},(i)} < \mu_{i\,\max} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$P\left(\mu_{\mathrm{t},(i)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, M\right)$$

This is the pdf that should cause us problems. Since this is where all the issues related to the correlations in Equation 2, and the modelling errors get placed. We are asked to define a pdf for the true measurement given: the previous measurements; the measurement uncertainty parameters; the prediction of the system model for all previous measurements; and the prediction of the required measurement by the system model. There are a number of error models that might be considered. Below are those that we consider most useful.

No Modelling Error

The assumption here is that there is no modelling error and that the result of using the system model (M and $\underline{m}$) is exact. If we define $\omega_{(i)}$ to be the prediction by the system model of the measurement $\mu_{\mathrm{t},(i)}$, then

$$P\left(\mu_{\mathrm{t},(i)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, M\right) = \delta\left(\omega_{(i)} - \mu_{t,(i)}\right) \tag{8}$$

The integral term Q_i therefore becomes

$$Q_i = \frac{1}{a_i} \frac{1}{\sqrt{2\pi}\sigma_{(i)}} \exp\left(-\frac{\left(\mu_{\mathrm{o},(i)} - \omega_{(i)}\right)^2}{2\sigma_{(i)}^2}\right) \tag{9}$$

It follows that Equation 4 can be written as

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M\right) = P(\underline{m} \mid M) \prod_{i=1}^{n+1} \frac{1}{a_i} \frac{1}{\sqrt{2\pi}\sigma_{(i)}} \exp\left(-\frac{\left(\mu_{\mathrm{o},(i)} - \omega_{(i)}\right)^2}{2\sigma_{(i)}^2}\right) \tag{10}$$

This implies that when the model is exact, we do not need to take into account any correlation between the measurements. If $P(\underline{m} \mid M)$ is a uniform pdf and the $\sigma_{(i)}$ are independent of the measured values, then by taking the logarithm of $P(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M)$ one obtains a weighted sum of squares as the objective function. This is the approach taken by many people when trying to fit a function to some data, even if the assumption of the model being exact is unlikely to be appropriate. (This assumption is considered by Tarantola (Tarantola, 1987, p. 57).

Uniform Distribution

The assumption here is that the true measurement $\mu_{\mathrm{t},(i)}$, lies somewhere between a lower and an upper bound with uniform probability. Hence

$$P\left(\mu_{\mathrm{t},(i)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, M\right) = \begin{cases} b_i & \mu_{i\,\min} < \mu_{\mathrm{t},(i)} < \mu_{i\,\max} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

and Q_i is hence given by

$$Q_i = \frac{b_i}{a_i} \frac{1}{\sqrt{2\pi}\sigma_{(i)}} \int_{\mu_{i\,\min}}^{\mu_{i\,\max}} \exp\left(-\frac{\left(\mu_{\mathrm{o},(i)} - \mu_{\mathrm{t},(i)}\right)^2}{2\sigma_{(i)}^2}\right) \mathrm{d}\mu_{\mathrm{t},(i)} \tag{12}$$

Correlations between measurements is expressed in the choice of bounds. The more correlated the measurements, then the narrower the range of integration.

Normal Distribution

The assumption here is that the modelling error is normally distributed around the predicted value,

$$P\left(\mu_{\mathrm{t},(i)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, M\right) = \frac{1}{\sqrt{2\pi}e_{1,(i)}} \exp\left(-\frac{\left(\omega_{(i)} - \mu_{\mathrm{t},(i)}\right)^2}{2e_{1,(i)}^2}\right) \tag{13}$$

and Q_i is hence given by

$$Q_i = \frac{1}{a_i} \frac{1}{2\pi\sigma_{(i)}e_{1,(i)}} \int_{-\infty}^{\infty} \exp\left(-\frac{\left(\mu_{\mathrm{o},(i)} - \mu_{t,(i)}\right)^2}{2\sigma_{(i)}^2}\right) \times \exp\left(-\frac{\left(\omega_{(i)} - \mu_{\mathrm{t},(i)}\right)^2}{2e_{1,(i)}^2}\right) \mathrm{d}\mu_{\mathrm{t},(i)} \tag{14}$$

Where $e_{1,(i)}$ is a parameter of the modelling error model. For more correlated measurements then a smaller value of $e_{1,(i)}$ should be selected. (This assumption is considered by Tarantola (Tarantola, 1987, p. 58).

Constant Error

The assumption here is that there is a constant offset between model and true measurement.

$$P\left(\mu_{\mathrm{t},(i)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, M\right) = \delta\left(\omega_{(i)} + e_{2,(i)} - \mu_{t,(i)}\right) \tag{15}$$

hence Q_i is given by

$$Q_i = \frac{1}{a_i} \frac{1}{\sqrt{2\pi}\sigma_{(i)}} \exp\left(-\frac{\left(\mu_{\mathrm{o},(i)} - e_{2,(i)} - \omega_{(i)}\right)^2}{2\sigma_{(i)}^2}\right) \tag{16}$$

Where $e_{2,(i)}$ is a parameter of the modelling error model. This equation implies that any correlations between measurements will be reflected by our ability to choose the value of $e_{2,(i)}$.

Fractional Error

The assumption here is that the error is a constant fraction of the prediction

$$P\left(\mu_{\mathrm{t},(i)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, M\right) = \delta\left(\omega_{(i)}\left(1 + e_{2,(i)}\right) - \mu_{\mathrm{t},(i)}\right) \tag{17}$$

therefore

$$Q_i = \frac{1}{a_i} \frac{1}{\sqrt{2\pi}\sigma_{(i)}} \exp\left(-\frac{\left(\mu_{\mathrm{o},(i)} - \omega_{(i)}\left(1 + e_{2,(i)}\right)\right)^2}{2\sigma_{(i)}^2}\right) \tag{18}$$

Where $e_{2,(i)}$ is a parameter of the modelling error model. Again this implies that any correlations between measurements will be reflected by our ability to choose the value of $e_{2,(i)}$.

Constant + Normal Error

Finally we consider a combination of a constant offset and a random Gaussian error

$$P\left(\mu_{\mathrm{t},(i)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, M\right) = \frac{1}{\sqrt{2\pi}e_{1,(i)}} \exp\left(-\frac{\left(\omega_{(i)} + e_{2,(i)} - \mu_{\mathrm{t},(i)}\right)^2}{2e_{1,(i)}^2}\right) \tag{19}$$

and Q_i is hence given by

$$Q_i = \frac{1}{a_i}\frac{1}{2\pi\sigma_{(i)}e_{1,(i)}}\int_{-\infty}^{\infty}\exp\left(-\frac{\left(\mu_{\mathrm{o},(i)}-\mu_{\mathrm{t},(i)}\right)^2}{2\sigma_{(i)}^2}\right) \times \exp\left(-\frac{\left(\omega_{(i)}+e_{2,(i)}-\mu_{t,(i)}\right)^2}{2e_{1,(i)}^2}\right)\mathrm{d}\mu_{\mathrm{t},(i)} \tag{20}$$

Where $e_{1,(i)}$ and $e_{2,(i)}$ are parameters of the modelling error model. This error model is essentially the same as that described in equation 14, except the mean of the distribution is now nonzero.

IDENTIFYING THE ERROR MODEL

In a strictly Bayesian framework the constants, $e_{1,(i)}$ and $e_{2,(i)}$, in the model of the "modelling error," would be handled as uncertain parameters. Whose effects would be dealt with through the normal "marginalization" framework. The narrower the range of significantly nonzero pdf values, the more correlated the measurements are. For the purposes of this paper we will seek a maximum likelihood estimate for the constants, we are implicitly assuming that the measurements are highly correlated, i.e. we seek $\underline{e}^*_{(i)} = (e^*_{1,(i)}, e^*_{2,(i)})$ such that

$$P\left(\underline{e}^*_{(i)} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, E\right) = \max P\left(\underline{e}_{(i)} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, E\right) \tag{21}$$

where

$$E = \text{error model} + \text{system realization}(M + \underline{m})$$

The pdf is given by

$$P\left(\underline{e}_{(i)} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, E\right) = \int P\left(\underline{e}_{(i)}, \underline{\mu}_{\mathrm{t}}^{i-1} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, E\right)\mathrm{d}\underline{\mu}_{t}^{i-1} \tag{22}$$

which is equivalent to Equation 2, but with different parameters and model. We can therefore write

$$P\left(\underline{e}_{(i)} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, E\right) = P\left(\underline{e}_{(i)} \mid E\right)\prod_{j=1}^{i-1} R_j \tag{23}$$

where

$$R_j = \int \frac{P\left(\mu_{t,(j)} \mid \underline{e}_{(i)}, \underline{\mu}_o^{j-1}, \underline{\sigma}^{j-1}, E\right) P\left(\mu_{o,(j)} \mid \mu_{t,(j)}, \sigma_{(j)}\right)}{P\left(\mu_{o,(j)} \mid \underline{\mu}_o^{j-1}, \underline{\sigma}^{j-1}, \sigma_{(j)}, E\right)} \mathrm{d}\mu_{t,(i)} \qquad (24)$$

This time we would argue that the error model is exact, therefore

$$P\left(\mu_{t,(j)} \mid \underline{e}_{(i)}, \underline{\mu}_o^{j-1}, \underline{\sigma}^{j-1}, E\right) = \delta\left(\Omega_{(j)} - \mu_{t,(j)}\right) \qquad (25)$$

where $\Omega_{(j)}$ is the prediction of $\mu_{t,(j)}$ after applying both the system model and the error model. Therefore

$$R_j = \frac{1}{c_j} \frac{1}{\sqrt{2\pi}\sigma_{(j)}} \exp\left(-\frac{\left(\mu_{o,(j)} - \Omega_{(j)}\right)^2}{2\sigma_{(j)}^2}\right) \qquad (26)$$

Given that $\underline{e}_{(i)}$ is of low order (one or two in our examples), it should be relatively simple to obtain the maximum likelihood estimate.

NUMERICAL EXPERIMENT ON A ONE-DIMENSIONAL PROBLEM

To illustrate the value of accounting for the unknown modelling errors, a simple one-parameter model is considered. From a known analytical model, $f(x) = (x^2 + 0.1x)^2$, six function values were created, to these a random Gaussian error was subtracted. The mean and the standard deviation of the error both being 5% of the function value, the data used is given in Table 1. It is then assumed that the most appropriate model is of the form x^y, where y is a real number, and one needs to obtain the likelihood function for y.

To use the standard Bayesian approach needs the pdf's mentioned in Equation 2 to be defined. The author, for one, does not feel confident about how

Table 1. Data for the One-Parameter Curve Fitting Problem

x	$f(x)$	Measurement
2	17.64	17.103
3	86.49	78.951
4	268.96	254.320
5	650.25	642.377
6	1339.56	1254.231
7	2470.09	2367.028

to define the correlated pdf's required for this problem. Working with the revised Bayesian approach as defined by Equation 4, seems easier to handle. We use Equations 6 and 7 with the following simplifications:

$$P\left(\mu_{o,(i)} \mid \mu_{t,(i)}, \sigma_{(i)}\right) = \frac{1}{\sqrt{2\pi}(0.05\mu_{t,(i)})} \exp\left(-\frac{\left(\mu_{o,(i)} - \mu_{t,(i)}\right)^2}{2\left(0.05\mu_{t,(i)}\right)^2}\right) \tag{27}$$

which implies that the standard deviation of the error model associated with a measurement is 5% of the unobserved true measurement.

$$P\left(\mu_{o,(i)} \mid \underline{\mu}_o^{i-1}, \underline{\sigma}^{i-1}, \sigma_{(i)}, Y\right) = a_i \tag{28}$$

which is a uniform pdf for the measurement $\mu_{o,(i)}$ given all the previous measurements and a realization, Y, of the proposed model. We consider two possible situations: the first uses the naive approach of assuming an exact model, which will be similar to a weighted sum of squares approach; the second will assume that the error is a fraction of the predicted value, this was chosen since in the author's opinion this is likely to be a good error model to choose for this problem.

1. The first being the assumption of there being no modelling error. Then the pdf is given by

$$P\left(y \mid \underline{\mu}_o^n, \underline{\sigma}^n, Y\right) = P(y \mid Y) \times \prod_{i=1}^{n} \left\{ \frac{1}{a_i} \frac{1}{\sqrt{2\pi}\left(0.05\omega_{(i)}\right)} \exp\left(-\frac{\left(\mu_{o,(i)} - \omega_{(i)}\right)^2}{2\left(0.05\omega_{(i)}\right)^2}\right)\right\} \tag{29}$$

   The resulting pdf for y is given in Figure 1.
2. The second to be considered, is that for the error to be a constant factor of the prediction. Therefore

$$P\left(\mu_{t,(i)} \mid y, \underline{\mu}_o^{i-1}, \underline{\sigma}^{i-1}, Y\right) = \delta\left(\omega_{(i)}\left(1 + e^*_{2,(i)}\right) - \mu_{t,(i)}\right) \tag{30}$$

   where $e^*_{2,(i)}$ is a maximum likelihood estimate based on the previous $(i-1)$ measurements. Therefore

$$Q_i = \frac{1}{a_i} \int \delta\left(\omega_{(i)}\left(1 + e^*_{2,(i)}\right) - \mu_{t,(i)}\right) \times \frac{1}{\sqrt{2\pi}\left(0.05\left(1 + e^*_{2,(i)}\right)\omega_{(i)}\right)} \exp\left(-\frac{\left(\mu_{o,(i)} - \mu_{t,(i)}\right)^2}{2\left(0.05\mu_{t,(i)}\right)^2}\right) \mathrm{d}\mu_{t,(i)} \tag{31}$$

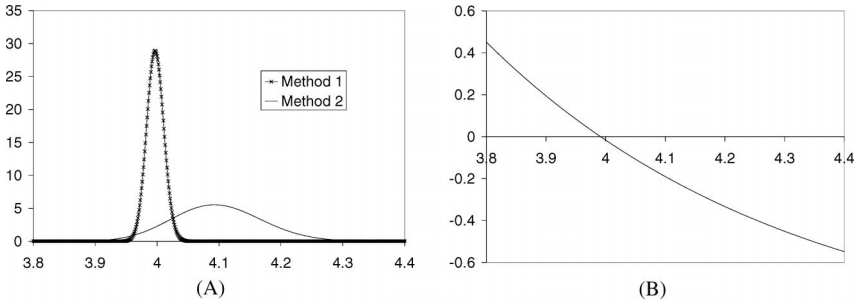

Figure 1. (A) The posterior pdf's obtained for the one-parameter model, (B) the maximum likelihood estimate of the error model parameter, y.

it follows that

$$P\left(y \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, Y\right) = P(y \mid Y) \times \prod_{i=1}^{n} \left\{ \frac{1}{a_i} \frac{1}{\sqrt{2\pi}\left(0.05\left(1 + e_{2,(i)}^{*}\right)\omega_{(i)}\right)} \times \exp\left(- \frac{\left(\mu_{\mathrm{o},(i)} - \left(1 + e_{2,(i)}^{*}\right)\omega_{(i)}\right)^{2}}{2\left(0.05\left(1 + e_{2,(i)}^{*}\right)\omega_{(i)}\right)^{2}} \right) \right\} \tag{32}$$

Where $e_{2,(i)}^{*}$ is chosen such that

$$P\left(e_{2,(i)}^{*} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, E\right) = \max P\left(e_{2,(i)} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, E\right) \tag{33}$$

and $E =$ error model + function realization $(y + Y)$. Using Equation 4 we can write this as

$$P\left(e_{2,(i)} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, E\right) = P\left(e_{2,(i)} \mid E\right) \prod_{j=1}^{i-1} R_j \tag{34}$$

where

$$R_j = \int \frac{P\left(\mu_{\mathrm{t},(j)} \mid e_{2,(i)}, \underline{\mu}_{\mathrm{o}}^{j-1}, \underline{\sigma}^{j-1}, E\right) P\left(\mu_{\mathrm{o},(j)} \mid \mu_{\mathrm{t},(j)}, \sigma_{(j)}\right)}{P\left(\mu_{\mathrm{o},(j)} \mid \underline{\mu}_{\mathrm{o}}^{j-1}, \underline{\sigma}^{j-1}, \sigma_{(j)}, E\right)} \mathrm{d}\mu_{\mathrm{t},(j)} \tag{35}$$

This time we assume that the error model is exact, hence

$$P\left(\mu_{\mathrm{t},(j)} \mid e_{2,(i)}, \underline{\mu}_{\mathrm{o}}^{j-1}, \underline{\sigma}^{j-1}, E\right) = \delta\left(\Omega_{(j)} - \mu_{\mathrm{t},(j)}\right) \tag{36}$$

and hence

$$R_j = \frac{1}{a_j} \frac{1}{\sqrt{2\pi}\left(0.05\left(1+e_{2,(j)}\right)\omega_{(j)}\right)} \exp\left(-\frac{\left(\mu_{\mathrm{o},(j)} - \left(1+e_{2,(j)}\right)\omega_{(j)}\right)^2}{2\left(0.05\left(1+e_{2,(j)}\right)\omega_{(j)}\right)^2}\right) \tag{37}$$

The pdf obtained is shown in Figure 1.

From Figure 1(A) one can see that when no account is taken of the modelling error we obtain a very localized pdf. The wider pdf comes from the second method of analysis. The maximum likelihood estimate obtained from method 1 is $y = 3.996$, using this value we obtain a prediction of $f(10) \approx 9908$. The true value being $f(10) = 10201$. The maximum likelihood estimate gained from method 2 is $y = 4.092$, with the modelling error parameter (obtained from the maximum likelihood estimate using all but one of the measurements) having a value of $e^* = -0.17769$, the maximum likelihood estimate, e^*, for a range of model parameter, y, is given in Figure 1(B). The prediction, using this value of e^* is $f(10) \approx 10163$.

On the left hand side of Figure 2 is a plot of the cumulative probability function plotted against the prediction for $f(10)$. One can see that method 2 is predicting a much narrower range for $f(10)$ compared to method 1, despite method 2 having a much wider range of nonzero pdf values. On the right hand side of the figure we see how the prediction for $f(10)$ varies over the same range for y. The horizontal line is the true value (10201), the highly sloping line is the method 1 estimate, and

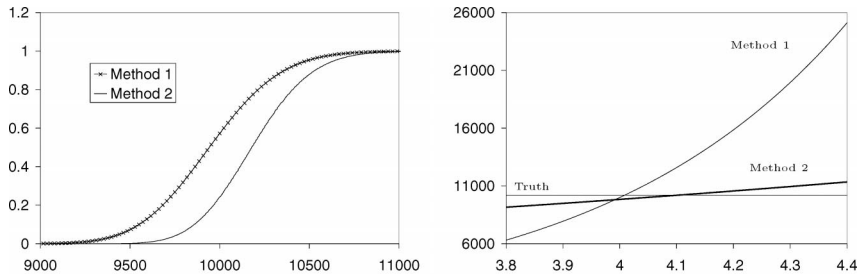

Figure 2. (A) The cumulative probability distributions for the prediction of $f(10)$, (B) prediction of the $f(10)$ for a range of parameter values.

the gently sloping line is the method 2 estimate. We can see that method 2 gives a much narrower range than method 1. The probabilistic expectation for $f(10)$ is: for method 1 $\langle f(10) \rangle = 9939.32$, with a standard deviation $\sigma = 316.50$; for method 2 $\langle f(10) \rangle = 10187.24$, with a standard deviation $\sigma = 263.51$.

NUMERICAL EXPERIMENT ON A CROSS-SECTION MODEL

The second example that we consider is a cross-sectional model of a simple layered reservoir with a simple fault, containing just oil and water, as illustrated in Figure 3. We use the same approaches to estimate the model parameters as were used in the first example. Method 1 is retained as this is the simple approach that would be used by many people when faced with this problem. Method 2 is retained because in the opinion of the author it is the error model, of the six discussed previously, that is most likely to be useful (as we shall see, it turns out not to be the error model that we need).

The geological model consists of six layers of alternating good and poor quality sands. The three good quality layers have identical properties, and the three poor quality layers have a different set of identical properties (details can be found in Appendix C). The thickness of the layers has an arithmetic progression, with the top layer having a thickness of 12.5 ft, the bottom layer a thickness of

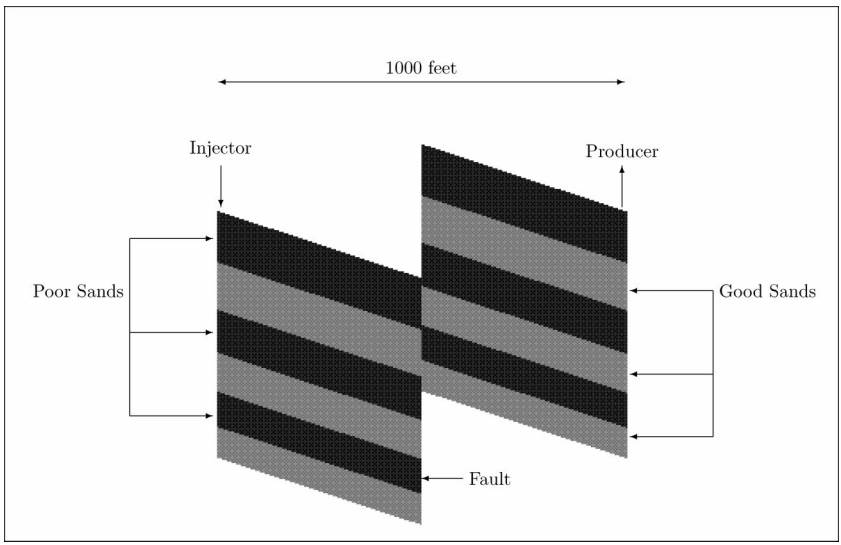

Figure 3. Schematic of a realization of the IC Fault Model, showing the alternating good and poor sands and the fault.

7.5 ft, and a total thickness of 60 feet. The width of the model is 1000 feet, with a simple fault at the midpoint, which offsets the layers. There is a water injector well at the left-hand edge, and a producer well on the right-hand edge. Both wells are completed on all layers, and operated at fixed bottom hole pressures (as described in Appendix C).

The simulation model is 100×12 grid blocks, with each geological layer divided into two simulation layers with equal thicknesses, each grid block is 10 ft wide. The model is constructed such that the vertical positions of the wells are kept constant and equal, even when different fault throws are considered. The well depth is 8325 ft to 8385 ft.

To generate historical data a slightly heterogeneous reservoir realization was created, with a throw of 10.3 ft. The porosity and permeabilities in each grid block were randomly drawn from uniform distributions with no correlations. The range for the porosities was $\pm 10\%$ of the mean value, while the range for the permeabilities was $\pm 1\%$ of the mean value. The means for the porosities were: good quality sand, 0.30; poor quality sand, 0.15. The means of the permeabilities were: good quality sand, 158.6 mD; poor quality sand, 2.0 mD.

The simulation was then run, using the ECLIPSE simulator, for 36 months, with three quantities (water production rate, Q_{wp}, oil production rate, Q_{op}, and water injection rate, Q_{wi}) being recorded at the end of each month. To each truth value, random Gaussian noise, with mean zero and standard deviation of 3% of the truth value (a minimum of 0.01 was allowed), was added. The resulting measurement was then truncated after one decimal place, the values obtained and used in the analysis are in Table 2.

The model used to try and match the results was identical, except it was assumed that the good/poor quality sands were homogeneous. The porosities of the two sands were set to 0.30 or 0.15 as appropriate. It was assumed that the two permeabilities and the fault throw were the unknowns that needed to be found.

A previous study using this model (Bush and Carter, 1996) had demonstrated that it is difficult to obtain a history match using simple optimization methods. Therefore the approach used for this study was to generate a large number of realizations, 159645 were produced, which could be searched to find the best match according to the criteria chosen. The values of the three parameters were drawn independently from uniform distributions with ranges: fault throw, $h \in (0, 60)$; good quality permeability, $K_{\mathrm{good}} \in (100, 200)$; poor quality permeability, $K_{\mathrm{poor}} \in (0, 50)$. Whilst this approach does not guarantee finding the global optimum, one should be able to get sufficiently close. One should also be able to identify local optima of the objective function should they exist.

In Table 3 are the results of applying method 1, we can see that there are at least two optima among the best six realizations. If the $h = 34.6$ realization is assumed to be the maximum likelihood estimate, then the $h = 0.1$ realization has a pdf

Table 2. Observations of Production for the IC Fault Model

Date			Q_{po}	Q_{wp}	Q_{wi}	Date			Q_{op}	Q_{wp}	Q_{wi}
31	JAN	1	464.9	0.0	767.1	31	JUL	2	530.0	0.0	903.9
28	FEB	1	511.4	0.0	860.0	31	AUG	2	573.6	0.0	861.0
31	MAR	1	525.0	0.0	827.5	30	SEP	2	558.8	0.0	919.7
30	APR	1	544.4	0.0	908.5	31	OCT	2	571.9	0.0	856.2
31	MAY	1	550.1	0.0	871.8	30	NOV	2	539.6	0.0	916.0
30	JUN	1	542.6	0.0	910.5	31	DEC	2	563.6	0.0	929.3
31	JUL	1	568.3	0.0	903.4	31	JAN	3	572.4	0.0	885.7
31	AUG	1	529.2	0.0	920.4	28	FEB	3	559.5	0.0	886.7
30	SEP	1	561.2	0.0	932.0	31	MAR	3	553.3	0.0	908.0
31	OCT	1	568.8	0.0	921.1	30	APR	3	565.5	0.4	928.1
30	NOV	1	561.9	0.0	884.1	31	MAY	3	517.0	40.8	907.6
31	DEC	1	543.4	0.0	900.7	30	JUN	3	527.3	112.0	943.4
31	JAN	2	578.6	0.0	908.4	31	JUL	3	492.4	149.1	852.1
28	FEB	2	566.0	0.0	916.7	31	AUG	3	454.9	177.0	943.8
31	MAR	2	533.9	0.0	919.1	30	SEP	3	457.5	206.8	922.3
30	APR	2	574.8	0.0	901.6	31	OCT	3	460.3	205.0	898.9
31	MAY	2	578.7	0.0	910.9	30	NOV	3	414.0	227.6	944.2
30	JUN	2	553.3	0.0	877.1	31	DEC	3	387.5	298.3	917.2

value which is significantly smaller. In Figure 4 we plot the realization estimates against the measured data. The two realizations clearly match the measurements quite well. The only place where they substantially disagree is in the estimate for the last two months of production. The a priori realization uses the true mean values as the parameters ($h = 10.30$, $K_{good} = 158.6$, $K_{poor} = 2.0$), it is quite unable to estimate the measurements. The actual values of production rates at 48 months are: $Q_{op}(48) = 171.4$, $Q_{wp}(48) = 681.4$ and $Q_{wi}(48) = 962.0$. As we can see from the value in Table 3, the $h = 34.6$ realization is a poor predictor. The $h = 0.1$ realization is much better at predicting, but we have little reason to choose it in preference to the $h = 34.6$ realization.

Table 3. Results for Method 1 Applied to the Cross-Sectional Model

h	K_{good}	K_{poor}	$Q_{op}(48)$	$Q_{wp}(48)$	$Q_{wi}(48)$	$\ln(P(\underline{m}))$
34.6	140.8	2.6	246.4	603.3	995.0	-7.14×10^2
31.0	137.3	1.8	244.1	543.3	949.2	-7.23×10^2
38.1	139.0	2.5	243.4	567.9	970.4	-7.28×10^2
0.1	130.2	1.3	145.2	708.9	960.7	-7.29×10^2
34.6	139.4	2.2	244.9	571.4	971.2	-7.31×10^2
1.0	129.7	3.7	193.0	646.1	979.1	-7.40×10^2

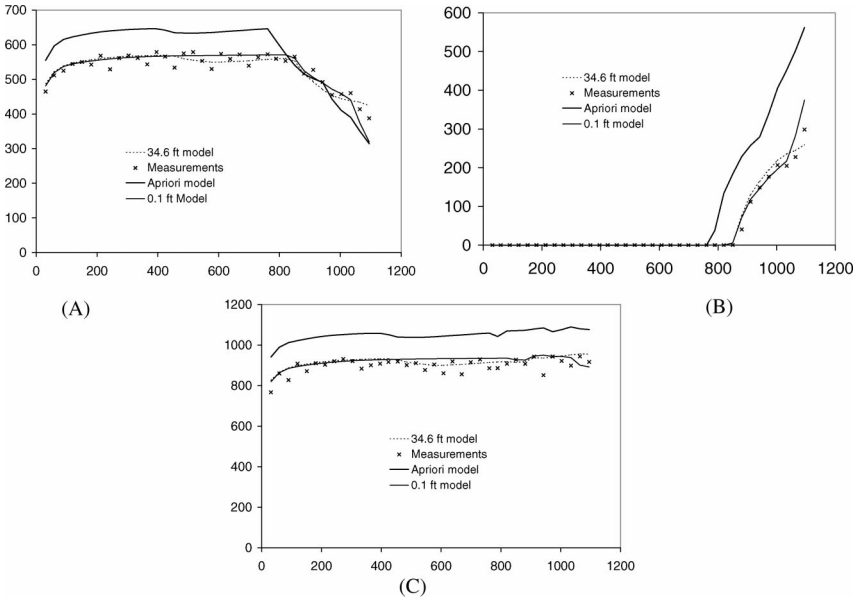

Figure 4. (A) Oil production rate Q_{op}, (B) water production rate Q_{wp}, (C) water injection rate Q_{wi}.

We can make the following observations:

- Within the Bayesian framework one of our realizations is significantly more likely than any of the others. However, a visual comparison of the two realizations does not give the same impression.
- Neither of the two realizations appear to be related to the truth case. It therefore seems likely that we are not able to recover a realization that is a good representation of the truth.
- If we were to have good quality data which provide a strong prior pdf, then it may not be possible to match the measurements.

Table 4. Results for Method 2 Applied to the Cross-Sectional Model

h	K_{good}	K_{poor}	$Q_{op}(48)$	$Q_{wp}(48)$	$Q_{wi}(48)$	e^*_{po}	e^*_{pw}	e^*_{iw}
28.8	136.6	0.5	297.4	444.4	941.8	0.24	−0.04	0.11
44.8	138.2	1.3	236.1	685.3	948.2	0.01	0.44	0.09
44.6	138.7	1.6	234.8	682.8	948.3	0.003	0.37	0.08
27.1	143.3	0.4	292.2	471.1	954.4	0.17	0.02	0.10
41.2	135.4	1.4	233.7	679.7	945.0	−0.004	0.36	0.07
45.3	131.7	0.6	261.7	666.0	953.0	0.15	0.72	0.26

In Table 4 we give the same results using method 2. We also give the maximum likelihood estimate of the modelling errors for each of the three measured quantities. Again we have identified at least two local optima among the best six realizations. But these are different from those obtained by method 1. One of the two optima appears to match the water production quite well (implied by the small value of e^*_{pw}) but has a poor match to the oil production. The second appears to match the oil production but not the water production. The predictions are again poor, so we conclude that our error model estimates have not been able to compensate for the bias introduced by the choice of model for the history matching process.

Examining Figure 4 suggests that the most appropriate error model might be a combination of a linear offset and fractional error (sections and), i.e.

$$P\left(\mu_{\mathrm{t},(i)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, M\right) = \delta\left(\omega_{(i)}\left(1 + e_{2,(i)}\right) + e_{1,(i)} - \mu_{\mathrm{t},(i)}\right) \qquad (38)$$

with only the previous 6 months used to define the error model parameters. A consequence of this would be that our model would have limited predictive capability.

CONCLUSIONS

In this paper we have rewritten the standard Bayesian equations, used to solve inverse problems, in an alternative form. The standard form requires the user to define the correlations between measurements and to carry out difficult multidimensional integrations. In the author's experience these requirements have largely limited the application of the method to idealized test models. When the method is applied to practical problems simplifying assumptions are used, even if these assumptions are of doubtful validity. The alternative form presented in this paper has the equation written as the product of many one-dimensional integrals. The advantage of this form is that the integrations can be carried out reasonably easily, and that the inclusion of information about the error model and the correlations is done in a more intuitive way.

Two examples have been considered. The first was a simple curve fitting problem involving the determination of one parameter, using six data points. Using the proposed form of the Bayesian equations we were able to introduce an appropriate error model and estimate the correlation between measurements. The result of this was a better predictive capability. The second example was an inverse problem based on a petroleum reservoir. It was shown that if one assumes an exact reservoir model, then we can obtain at least two matches to the production data. Neither of the matches were consistent with the a priori data. When the second error model was applied, the situation did not improve. It has been suggested that

a simple deterministic error model matched to the measurements over a limited window might provide the best approach.

NOMENCLATURE

$P(. \mid .)$ a probability distribution function (pdf) that describes our beliefs about a set of parameters given some information.
M our model of the real system, it includes things like: the algorithms used to construct the numerical model from the parameters; controls that are used on the real system; the equations that need to be solved; and the numerical algorithms that are used.
$\underline{m}$ the parameters of our system model.
$\mu_{\mathrm{t},(i)}$ the ith true, but unobserved, measurement.
$\underline{\mu}_{\mathrm{t}}^{n}$ a vector of n true measurements.
$\mu_{\mathrm{o},(i)}$ the ith observed measurement.
$\underline{\mu}_{\mathrm{o}}^{n}$ a vector of n observed measurements.
$\sigma_{(i)}$ the uncertainty parameter associated with the ith measurement.
$\underline{\sigma}^{n}$ a vector of n uncertainty parameters associated with the n observed measurements.
Q_i an integral that defines a pdf associated with the ith measurement.
$e_{1,(i)}$ a parameter used to define the modelling error model associated with the ith measurement.
$e_{2,(i)}$ a parameter used to define the modelling error model associated with the ith measurement.
E the modelling error model.
$e_{1,(i)}^{*}$ the maximum likelihood estimate of $e_{1,(i)}$.
$e_{2,(i)}^{*}$ the maximum likelihood estimate of $e_{2,(i)}$.
$\underline{e}_{(i)}$ a vector of parameters used to define the modelling error model associated with the ith measurement.
R_i an integral that defines a pdf associated with the modelling error at the ith measurement.
x the ordinate used in the first example.
$f(x)$ the function used in the first example.
Ω the unknown parameter used in the first example.
y the model parameter used in the first example.
Y a realization in the first example.
h a parameter in the second example.
K_{good} a parameter in the second example.
K_{poor} a parameter in the second example.
Q_{op} oil production rate.
Q_{wp} water production rate.
Q_{wi} water injection rate.

e^*_{po} maximum likelihood estimate for the oil production modelling error parameter.

e^*_{pw} maximum likelihood estimate for the water production modelling error parameter.

e^*_{iw} maximum likelihood estimate for the water injection modelling error parameter.

REFERENCES

Aanonsen, S. I., Cominelli, A., Gosselin, O., Aavatsmark, I., and Barkve, T., 2002, Integration of 4D data in the history match loop by investigating scale dependent correlations in the acoustic impedance cube, *in* Proceedings of the 8th European Conference on Maths of Oil Recovery, p. E59. European Association of Geoscientists and Engineers, Houten, The Netherlands.

Bush, M. D., and Carter, J. N., 1996, Applications of a modified genetic algorithm to parameter estimation in the petroleum industry: Intelligent engineering systems through artificial neural networks, v. 6, p. 397. ASME Press, New York.

DeVolder, B., Glimm, J., Grove, J. W., Kamg, Y., Lee, Y., Poa, K., Sharp, D. H., and Ye, K., 2002, Uncertainty quantification for multiscale simulations: J. Fluids Eng., v. 124, p. 1–13.

Floris, F. J. T., Bush, M. D., Cuypers, M., Roggero, F., and Syversveen, A. R., 2001, Methods for quantifying the uncertainty of production forecasts—a comparative study: Pet. Geosci., v. 7, p. S87–S96.

Glimm, J., Hou, S., Kim, H., Lee, Y., Sharp, D. H., and Zou, Q., 2001, Risk management for petroleum reservoir production: A simulation-based study of prediction: Comp. Geosci., v. 5, p. 173–197.

Hauge, R., Arntzen, O. J., and Soleng, H., 2002, Choosing objective function for conditioning on production data, *in* Proceedings of the 8th European Conference on Maths of Oil Recovery, p. E47.

Sivia, D. S., 1996, Data analysis: A Bayesian tutorial: Oxford University Press, Oxford.

Tarantola, A., 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier Science, Amsterdam.

Wu, Z., Reynolds, A. C., and Oliver, D. S., 1999, Conditioning geostatistical models to two-phase production data: SPE J., v. 4, p. 142–155.

APPENDIX A: STANDARD ANALYSIS

For those unfamiliar with Bayesian statistics, a good introduction is provided by Sivia (Sivia, 1996). Starting from the standard Bayesian statement

$$P\left(\underline{m} \mid \underline{\mu}_o^n, \underline{\sigma}^n, M\right) = \int P\left(\underline{m}, \underline{\mu}_t^n \mid \underline{\mu}_o^n, \underline{\sigma}^n, M\right) \mathrm{d}\underline{\mu}_t^n \tag{A1}$$

we can use the product rule to obtain

$$P\left(\underline{m} \mid \underline{\mu}_o^n, \underline{\sigma}^n, M\right) = \int P\left(\underline{m} \mid \underline{\mu}_t^n, \underline{\mu}_o^n, \underline{\sigma}^n, M\right) P\left(\underline{\mu}_t^n \mid \underline{\mu}_o^n, \underline{\sigma}^n, M\right) \mathrm{d}\underline{\mu}_t^n \tag{A2}$$

Then using Bayes Rule, we get

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) = \int P\left(\underline{m} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) \frac{P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\sigma}^{n}, M\right) P\left(\underline{\mu}_{\mathrm{t}}^{n} \mid \underline{\sigma}^{n}, M\right)}{P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\sigma}^{n}, M\right)} \mathrm{d}\underline{\mu}_{\mathrm{t}}^{n} \quad \text{(A3)}$$

We now introduce the first simplification
Simplification: $P\left(\underline{m} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) = P\left(\underline{m} \mid \underline{\mu}_{\mathrm{t}}^{n}, M\right)$
Where the assumption is that if we were to know the true measurements, then we would not make any use of the observed measurements or the associated uncertainty parameters. Which lets us write

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) = \int P\left(\underline{m} \mid \underline{\mu}_{\mathrm{t}}^{n}, M\right) \frac{P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\sigma}^{n}, M\right) P\left(\underline{\mu}_{\mathrm{t}}^{n} \mid \underline{\sigma}^{n}, M\right)}{P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\sigma}^{n}, M\right)} \mathrm{d}\underline{\mu}_{\mathrm{t}}^{n} \quad \text{(A4)}$$

Using Bayes Rule this then becomes

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) = \int \frac{P\left(\underline{\mu}_{\mathrm{t}}^{n} \mid \underline{m}, M\right) P\left(\underline{m} \mid M\right)}{P\left(\underline{\mu}_{\mathrm{t}}^{n} \mid M\right)} \times \frac{P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\sigma}^{n}, M\right) P\left(\underline{\mu}_{\mathrm{t}}^{n} \mid \underline{\sigma}^{n}, M\right)}{P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\sigma}^{n}, M\right)} d\underline{\mu}_{\mathrm{t}}^{n} \quad \text{(A5)}$$

We now introduce two further simplifications.
Simplification: $P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\sigma}^{n}, M\right) = P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\sigma}^{n}\right)$
where we are assuming that the measurement error model is independent of the system model.
Simplification: $P\left(\underline{\mu}_{\mathrm{t}}^{n} \mid \underline{\sigma}^{n}, M\right) = P\left(\underline{\mu}_{\mathrm{t}}^{n} \mid M\right)$
where we are assuming that in the absence of observed measurements that knowledge of the uncertainty in those measurements is of no value. Hence we obtain the standard result

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) = P\left(\underline{m} \mid M\right) \int \frac{P\left(\underline{\mu}_{\mathrm{t}}^{n} \mid \underline{m}, M\right) P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\mu}_{\mathrm{t}}^{n}, \underline{\sigma}^{n}\right)}{P\left(\underline{\mu}_{\mathrm{o}}^{n} \mid \underline{\sigma}^{n}, M\right)} \mathrm{d}\underline{\mu}_{\mathrm{t}}^{n} \quad \text{(A6)}$$

APPENDIX B: ITERATIVE ANALYSIS

Starting from the standard Bayesian statement

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M\right) = \int P\left(\underline{m}, \mu_{t,(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{o},(n+1)}, \sigma_{(n+1)}, M\right) \mathrm{d}\mu_{\mathrm{t},(n+1)} \tag{A7}$$

Using the product rule we can obtain

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M\right) = \int P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{t,(n+1)}, \mu_{o,(n+1)}, \sigma_{(n+1)}, M\right) \times P(\mu_{\mathrm{t},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{o},(n+1)}, \sigma_{(n+1)}, M) \mathrm{d}\mu_{\mathrm{t},(n+1)} \tag{A8}$$

Introducing the first simplification.
Simplification:

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{t},(n+1)}, \mu_{\mathrm{o},(n+1)}, \sigma_{(n+1)}, M\right) = P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{t},(n+1)}, M\right)$$

where the assumption is that if we know the true value of the $(n+1)$th measurement, we gain no additional information from knowing its observed value and the associated uncertainty. Hence we can obtain

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M\right) = \int P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{t},(n+1)}, M\right) \times P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{o},(n+1)}, \sigma_{(n+1)}, M\right) \mathrm{d}\mu_{\mathrm{t},(n+1)} \tag{A9}$$

Using Bayes Rule on both pdf's we get

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M\right) = \int \frac{P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right)}{P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right)} \times P\left(\mu_{\mathrm{o},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{t},(n+1)}, \sigma_{(n+1)}, M\right) \times \frac{P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \sigma_{(n+1)}, M\right)}{P\left(\mu_{\mathrm{o},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \sigma_{(n+1)}, M\right)} \mathrm{d}\mu_{\mathrm{t},(n+1)} \tag{A10}$$

Introducing two more simplifications.

Simplification:

$$P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \sigma_{(n+1)}, M\right) = P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right)$$

where we are assuming that in the absence of the observed measurement that knowledge of the uncertainty in that measurement is of no value.

Simplification:

$$P\left(\mu_{\mathrm{o},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \mu_{\mathrm{t},(n+1)}, \sigma_{(n+1)}, M\right) = P\left(\mu_{\mathrm{o},(n+1)} \mid \mu_{\mathrm{t},(n+1)}, \sigma_{(n+1)}\right)$$

where we are assuming that the measurement error model is independent of the previous measurements.

Therefore

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M\right) = P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) \times \int \frac{P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) P\left(\mu_{\mathrm{o},(n+1)} \mid \mu_{\mathrm{t},(n+1)}, \sigma_{(n+1)}\right)}{P\left(\mu_{\mathrm{o},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \sigma_{(n+1)}, M\right)} \mathrm{d}\mu_{\mathrm{t},(n+1)} \tag{A11}$$

This can be written as

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M\right) = P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) \times Q_{n+1} \tag{A12}$$

where

$$Q_{n+1} = \int \frac{P\left(\mu_{\mathrm{t},(n+1)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, M\right) P\left(\mu_{\mathrm{o},(n+1)} \mid \mu_{\mathrm{t},(n+1)}, \sigma_{(n+1)}\right)}{P\left(\mu_{\mathrm{o},(n+1)} \mid \underline{\mu}_{\mathrm{o}}^{n}, \underline{\sigma}^{n}, \sigma_{(n+1)}, M\right)} \mathrm{d}\mu_{t,(n+1)} \tag{A13}$$

This clearly suggests that we can use an iterative formulation. Hence the required pdf can be written as

$$P\left(\underline{m} \mid \underline{\mu}_{\mathrm{o}}^{n+1}, \underline{\sigma}^{n+1}, M\right) = P\left(\underline{m} \mid M\right) \prod_{i=1}^{n+1} Q_i \tag{A14}$$

where

$$Q_i = \int \frac{P\left(\mu_{\mathrm{t},(i)} \mid \underline{m}, \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, M\right) P\left(\mu_{\mathrm{o},(i)} \mid \mu_{t,(i)}, \sigma_{(i)}\right)}{P\left(\mu_{\mathrm{o},(i)} \mid \underline{\mu}_{\mathrm{o}}^{i-1}, \underline{\sigma}^{i-1}, \sigma_{(i)}, M\right)} \mathrm{d}\mu_{t,(i)} \tag{A15}$$

APPENDIX C: DESCRIPTION OF THE IC FAULT MODEL

When designing this test model the objectives were: the model should be sufficiently complex that all the obvious symmetries should be avoided; it should be possible for the user to introduce and control the level of modelling and measurement errors; the model should be of sufficiently low dimension that it was feasible to understand the complete parameter space. The model chosen was: a simple cross-section model; with two homogeneous sand types in a layer cake formation; a simple vertical fault; an injector/producer pair operated at constant bottom hole pressures; no oil–water or gas–oil contacts; and three unknowns: the fault throw; and the two permeabilities. All other properties are kept constant, an ECLIPSE data file for the model is given below. The user can control the errors introduced in a number of ways: the models used to generate measurement data can be made different from the model used for solving the inverse problem; and measurement errors can be added. The user has complete control over the type and level of errors introduced.

The base case reservoir realization used in this paper has heterogeneous porosity and permeability fields. The variations have been kept small and uncorrelated, the measurement errors have also been kept small. The measurement data used in this paper is given in Table 2.

ECLIPSE Data File

```
RUNSPEC

GRID

INCLUDE

 'PERM.INC' \

\

INCLUDE

 'PORO.INC' \

\

MULTIPLY
```

```
      'PORO'  0.15      1  100  1  1  1  2     /
      'PORO'  0.3       1  100  1  1  3  4     /
      'PORO'  0.15      1  100  1  1  5  6     /
      'PORO'  0.3       1  100  1  1  7  8     /
      'PORO'  0.15      1  100  1  1  9 10     /
      'PORO'  0.3       1  100  1  1 11 12     /
      'PERMX' 1.0      1  100  1  1  1  2     /
      'PERMX' 100      1  100  1  1  3  4     /
      'PERMX' 1.0      1  100  1  1  5  6     /
      'PERMX' 100      1  100  1  1  7  8     /
      'PERMX' 1.0      1  100  1  1  9 10     /
      'PERMX' 100      1  100  1  1 11 12     /
/
COPY
      'PERMX'    'PERMY'   1  100  1  1  1 12  /
      'PERMX'    'PERMZ'   /
/
MULTIPLY
 'PERMX' 2.000000 1 100 1 1  1  2  /
 'PERMX' 2.000000 1 100 1 1  5  6  /
 'PERMX' 2.000000 1 100 1 1  9 10  /
```

```
 'PERMX' 1.586000 1 100 1 1  3  4  /
 'PERMX' 1.586000 1 100 1 1  7  8  /
 'PERMX' 1.586000 1 100 1 1 11 12  /
 /
NEWTRAN
INCLUDE
 'COORD.INC'
/
INCLUDE
 'ZCORN.INC'
/
PROPS
SWFN
    0.1   0.000       0
    0.2   0.012       0
    0.3   0.049       0
    0.4   0.111       0
    0.5   0.197       0
    0.6   0.308       0
    0.7   0.444       0
    0.8   0.605       0
```

```
    0.9   0.790       0

    1.0   1.000       0  /

SOF2

0.2   0.000

0.3   0.016

0.4   0.065

0.5   0.147

0.6   0.261

0.7   0.408

0.8   0.588

0.9   0.800 /

-- PVT PROPERTIES OF WATER

--    REF. PRES. REF. FVF  COMPRESSIBILITY  REF VISCOSITY
      VISCOSIBILITY

PVTW

       4014.7   1.029      3.13D-6         0.31          0 /

-- ROCK COMPRESSIBILITY

--    REF. PRES   COMPRESSIBILITY

ROCK

        14.7          3.0D-6          /

-- SURFACE DENSITIES OF RESERVOIR FLUIDS

--        OIL   WATER   GAS
```

```
DENSITY
         49.1   64.79  0.06054  /
--    POIL  FVFO  VISO
PVDO
     4014.7 1.695  0.51
     5014.7 1.671  0.549
     9014.7 1.579  0.74
/
RSCONST
1.270 4014.7 /
SOLUTION
--    DATUM DATUM  OWC   OWC   GOC   GOC   RSVD  RVVD  SOLN
--    DEPTH PRESS DEPTH  PCOW DEPTH  PCOG TABLE TABLE  METH
EQUIL
       8400  4500  8500   0    8200   0     1     0      0  /
SUMMARY
WOPR
 'PRODUCER' /
WWPR
 'PRODUCER' /
WWIR
```

```
 'INJECTOR' /

SCHEDULE

DRSDT

      0  /

--     WELL   GROUP LOCATION  BHP   PI

--     NAME   NAME    I  J   DEPTH DEFN

WELSPECS

    'PRODUCER' 'G'   100 1    8400 'OIL'  /

    'INJECTOR' 'G'    1  1    8335 'WATER'  /

/

-- COMPLETION SPECIFICATION DATA

--     WELL     -LOCATION- OPEN/ SAT CONN  WELL

--     NAME     I  J K1 K2 SHUT  TAB FACT  DIAM

COMPDAT

    'PRODUCER'  100 1 1 12 'OPEN' 0   -1   0.5  /

    'INJECTOR'   1  1 1 12 'OPEN' 1   -1   0.5  /

  /

-- PRODUCTION WELL CONTROLS

--      WELL    OPEN/ CNTL  OIL WATER  GAS LIQU  RES  BHP

--      NAME    SHUT  MODE RATE  RATE RATE RATE  RATE

WCONPROD
```

```
     'PRODUCER' 'OPEN' 'BHP'   100000  4*              4300 /

  /
-- INJECTION WELL CONTROLS
--      WELL     INJ   OPEN/  CNTL    FLOW
--      NAME    TYPE   SHUT   MODE    RATE
WCONINJ
     'INJECTOR' 'WATER' 'OPEN' 'BHP'  4*    4700  /
  /
DATES
 31 'JAN'  1 /
 28 'FEB'  1 /
 31 'MAR'  1 /
 30 'APR'  1 /
 31 'MAY'  1 /
 30 'JUN'  1 /
 31 'JUL'  1 /
 31 'AUG'  1 /
 30 'SEP'  1 /
 31 'OCT'  1 /
 30 'NOV'  1 /
 31 'DEC'  1 /
```

```
 :
 31 'DEC'  4 /
 /
END
```