



Quantifying Prediction Uncertainty in Reservoir Modelling Using Streamline Simulation

Mike Christie¹,
Sam Subbey²,
Malcolm Sambridge³,
and Marco Thiele⁴

ABSTRACT

The performance of oil reservoirs is inherently uncertain: data constraining the rock and rock-fluid properties is available at only a small number of spatial locations, and other measurements are integrated responses providing limited constraints on model properties. Calibrating a reservoir model to observed data is time consuming, and it is rare for multiple models to be 'history matched'. Uncertainty quantification usually consists of identifying high-side and low-side adjustments to the base case.

The Neighbourhood Algorithm is a stochastic sampling algorithm developed for earthquake seismology. It works by adaptively sampling in parameter space using geometrical properties of Voronoi cells to bias the sampling to regions of good fit to data. The algorithm evaluates the high dimensional integrals needed for quantifying the posterior probability distribution using Markov Chain Monte Carlo run on the misfit surface defined on the Voronoi cells.

This paper describes the use of the Neighbourhood Algorithm for obtaining multiple history matched reservoir models and using the ensemble of models to quantify uncertainty in reservoir performance forecasting. We describe the changes needed to generate multiple history matched models, and to sample from the posterior distribution to quantify uncertainty in forward predictions.

Effective quantification of uncertainty can require thousands of reservoir model runs, each of which can take several minutes for a relatively coarse grid to several hours for a fine grid. As part of this paper, we describe the use of approximate streamline simulations to rapidly explore parameter space. This allows us to switch to slower conventional simulation in regions of good fit to the data.

We demonstrate the performance of the algorithm on the SPE 10th Comparative Solution Project dataset. This is a benchmark dataset for which a fine grid reservoir description is known. We take this as "truth" and use a coarser model to match the history data for a limited period of time. We then predict both the maximum likelihood performance and the uncertainty envelope for the remaining time. The maximum likelihood solution is close to the truth case for much of the time, and the true solution always lies within the uncertainty bounds predicted by the algorithm.

Keywords: stochastic algorithm, history matching, uncertainty, streamlines.

¹Dept of Petroleum Engineering, Heriot-Watt University, Edinburgh, EH14 4AS, UK. E-mail: mike.christie@pet.hw.ac.uk

²Dept of Petroleum Engineering, Heriot-Watt University, Edinburgh, EH14 4AS, UK. E-mail: sam.subbey@pet.hw.ac.uk

³Research School of Earth Sciences, Institute of Advanced Studies, Australian National University, Canberra, ACT 0200, Australia. E-mail: malcolm@rses.anu.edu.au

⁴Streamsim Technologies, Inc., 2430 Lake Street, Suite 4, San Francisco, CA 94121, USA. E-mail: thiele@streamsim.com

INTRODUCTION

Mathematical modeling of complex systems is common in many branches of science and engineering. In petroleum reservoir modeling the system of interest consists of the petroleum reservoir rock and the various fluids (oil/gas/water), flowing through the rock. The properties of interest usually include those related to the transmission and storage capacity of the reservoir rock.

The principal aim is to accurately predict the reservoir performance and facilitate the making of economic and management decisions. The mathematical models use physical properties of the reservoir such as its geometry and fluid flow properties, and the predictive accuracy of the models depends on the accuracy with which these physical properties are known.

Direct measurements provide data about porosity and permeability fields. These data are obtained from core samples taken from exploration wells. Typically, the cores are of the order of $10^{-17} - 10^{-16}$ of the reservoir volume.

The other sources of data are those inferred from logs, well tests or from reservoir production data. Since these inferred data involve averaged responses over larger length scales, there is a large degree of uncertainty associated with them.

Since there is a large degree of uncertainty inherent in the reservoir description, the reservoir model is uncertain. During the production stage of the reservoir, one obtains dynamic data in the form of oil, water and gas production rates, as well as pressure profiles. In contrast to the static data (e.g., geometry and geology), these data are a direct measure of the reservoir response to the recovery process in application. Hence the only way to reduce the uncertainty in the reservoir model is to constrain the numerical model to the dynamic data. The process of incorporating such dynamic data in the generation of reservoir models is known as history matching. Thus in history matching, by adjusting model parameters, e.g., permeability and other flow properties, one aims at obtaining a model output which is as close to the history (dynamical) data as possible.

History matching involves determining the model parameters on the basis of the observation data. History matching is therefore an *inverse problem*. Specifically, the history matching problem is a non-linear, inverse and ill-posed problem. As is typical of inverse problems, rather than obtaining a unique set of history matching model parameters, there exists a family of sets of parameters, all of which will give good fits to the history data.

Since history matching is characterized by non-uniqueness, an obvious method of estimating the uncertainty would be to generate multiple reservoir realizations, i.e., an ensemble of models which all honor the history data, and forecast future production using this ensemble. The spread in the predictions of the multiple reservoir realizations will in general, be indicative of the degree of uncertainty in the forecasts. Recent approaches to history matching have recognized that quantifying uncertainty requires generating multiple reservoir realizations. In practice however, a major challenge is how to generate these realizations by an exhaustive exploration of the parameter space, if the generated models are to correctly quantify the uncertainty. Further, for problems with high input-output dimensions, each run of the simulator may be very expensive in CPU time. For most conventional simulators, a single run can take several minutes for a relatively coarse grid to several hours for a fine grid. Hence part of the challenge is finding a means of rapidly explore parameter space.

This paper present a stochastic sampling algorithm, the Neighbourhood Algorithm, for generating multiple reservoir realizations conditioned on history data. The new search algorithm makes use of geometrical constructs known as Voronoi cells to derive the search in parameter space. The generated ensemble is used to quantify uncertainty in a Bayesian framework.

THE NEIGHBOURHOOD APPROXIMATION ALGORITHM

The Neighbourhood Approximation algorithm (NA) is a stochastic sampling algorithm developed originally for earthquake seismology. For a full description of the algorithm, see (Sambridge 1999a) and (Sambridge 1999b). The algorithm explores parameter space using geometrical constructs known as Voronoi cells, see (Okabe et al. 1992), (Sambridge 1999a). Here, we adapt it to the problem of generating multiple history matching models. The NA algorithm generates multiple history matching models in parameter space according to the following rule.

At each iterative stage, the algorithm generates n_s models and calculates their misfits values. Then all the models, including those previously generated, are ranked to determine the best n_r cells. New n_s models are then generated in the best n_r cells, i.e., by placing n_s/n_r models in each cell.

The philosophy behind the the algorithm is that the misfit of each of the previous models is representative of the region of its neighbourhood, defined by its Voronoi cell. Therefore at each iteration step, new samples are concentrated in the neighbourhoods surrounding the better data-fitting models. Thus the objective of the algorithm is to bias the sampling to good history-matching regions of the parameter space, (Sambridge 1999a). By its nature, the algorithm exploits information in all previous models to selectively sample parameter space.

From the above, it is intuitive that the two parameters that control the algorithm are n_s and n_r . Indeed, these are the only tuning parameters that control the performance of the algorithm. The amount of exploration and exploitation performed by the algorithm is dependent on these parameters. For instance, increasing n_r implies more Voronoi cells are explored and the algorithm as a whole becomes less local in searching parameter space. As pointed out in (Sambridge 1999a), what is decisive in the performance of the algorithm is the ratio n_s/n_r , rather than the individual values of either n_s or n_r .

ENSEMBLE APPRAISAL – QUANTIFYING UNCERTAINTY

The NA algorithm generates an ensemble of history matching models with their corresponding measure of fit to the observed data. The aim at the appraisal stage is to infer information from this finite set of models in quantifying the uncertainty associated with predicting future reservoir performance.

Methods for sampling from the posterior pdf have been reported in the literature, see (Oliver et al. 1997), (Cunha et al. 1998) and (Omre et al. 1999).

In this application, Voronoi cells are used in constructing an approximate posteriori probability distribution (ppd) from a fixed ensemble, using a Bayesian approach.

Any *prior* assumptions we have of model m_j can be expressed through a (prior) probability density function $p(m_j)$. Given observation or history data \mathcal{O} , this data serves to refine previous knowledge about the model inputs by narrowing their posterior probability distribution. We note that the observation have uncertainty associated with them. Hence comparison of observations with solutions derived from the models will have additional uncertainty associated with uncertainties in the solution process (Glimm and Sharp 1998).

For each of the models generated, we can calculate the degree of likelihood that the observed data can be explained by the model for which the likelihood holds. This depends on the model misfit value, and is expressed as $p(\mathcal{O}|m)$, i.e., the probability of the observations \mathcal{O} , given the model m . Having obtained this likelihood, we update our prior knowledge to yield a posterior probability of the model, given the observed data. From Bayes rule the updated

posterior density for m is:

$$p(m|\mathcal{O}) = \frac{p(\mathcal{O}|m)p(m)}{\int p(\mathcal{O}|m)p(m)dm}. \quad (1)$$

The denominator in Equation 1 is a normalising constant which needs to be calculated in order to evaluate Equation 1 above. However, even for low dimensional problems this could be intractable. Markov chain Monte Carlo (MCMC) provides a method of sampling from the posterior distribution without the need of calculating the normalization constant. For detailed literature on MCMC methods see e.g., (Neal 1993), and for their application in petroleum engineering, see (Cunha et al. 1998), (Glimm and Sharp 1998), (Oliver et al. 1997) and (Omre et al. 1999).

Since the only information available about $p(m|\mathcal{O})$ is the posterior probability distribution of the input ensemble, a neighbourhood approximation of $p(m|\mathcal{O})$ is used. This is the backbone of the NA-Bayesian algorithm. Thus by sampling from the posterior distribution one generates new points in parameter space whose distribution $\approx p_{NA}(m|\mathcal{O})$, where $p_{NA}(m|\mathcal{O})$ is the neighbourhood approximation of $p(m|\mathcal{O})$, (Sambridge 1999b).

PROBLEM DESCRIPTION

The problem we intend to solve is an ideal one, and can be summarised as follows. Firstly, we run a fine grid model to obtain fine grid average pressure $P(t)$, produced oil and water rates, $Q_o(t)$ and $Q_w(t)$, respectively. We take these fine grid data as stand in for the field data. Using a coarse grid model, the aim is to history match the fine grid data, by determining relative permeability curves, $K_{ro}(S_w)$ and $K_{rw}(S_w)$. We then predict future fine grid behaviour for time longer than the history matching period, and then quantify the uncertainty with our predictions.

Using an ideal problem is essential in the testing phase of the algorithm since the truth is known and we can compare our results with the true data.

The fine grid is from the 10th SPE Comparative Solution Project, (Christie and Blunt 2001). The consists of $[60 \times 220 \times 85]$ cells. There are 4 producers in the four corners of the model, each producing at 4000psi bottom hole pressure and a central injector with an injection rate of 5000barrels/day. Figures 1.a., 1.b. and 1.c. show plots of the fine grid data $Q_o(t)$, $P(t)$ and oil/water relative permeabilities.

The coarse grid is an upscaling of the coarse grid and consists of $[5 \times 11 \times 10]$ cells. Here, single phase upscaling is used, (Christie 1996). The well positions and completions are the same as in the fine scale model.

PARAMETERIZATION

To model oil and water relative permeabilities, relative permeability functions in (Chierici 1981) were adopted. The model consists of two formations, and for each, the relative permeability curves for oil and water, as functions of water saturation, S_w , are defined respectively by

$$K_{ro}(S_w) = K_{ro}(S_{iw})e^{-\alpha R^\zeta}, \quad (2)$$

$$K_{rw}(S_w) = K_{rw}(S_{or})e^{-\beta R^{-\eta}}, \quad (3)$$

where $R = (S_w - S_{iw})/(1 - S_{or} - S_w)$, is a scaled saturation and S_{or} is the residual oil saturation. This implies that a total of 12 parameters, 6 for each formation, i.e., four empirical parameters ($\alpha, \beta, \zeta, \eta$) and two rock/fluid parameters ($K_{ro}(S_{iw}), K_{rw}(S_{or})$), had to be determined.

QUANTIFYING MISFIT

Two types of errors need to be accounted for namely, model and data errors. Data errors encapsulates the lack of knowledge of the subsurface geology and errors inherent in the dynamic data. The model errors on the other hand depend on the choice of numerical simulator we choose, i.e., mass balance or streamline. These errors arise from the approximation of the originally continuous conservation and flow equations with discrete analogues, and the inability to capture sub-grid details.

In the least square sense, the measure of misfit, E_T , can be expressed as

$$E_T = \sum_{n_w} \langle \Delta \vec{P} | C_p^{-1} | \Delta \vec{P} \rangle + \sum_{n_w} \sum_{j=o,w} \langle \Delta \vec{Q}_j | C_j^{-1} | \Delta \vec{Q}_j \rangle, \quad (4)$$

where $\Delta \vec{P}$ and $\Delta \vec{Q}_j$ represent the difference between the history data and that from the coarse grid, for the average pressure and the fluid rates, respectively. The covariance matrices C_p and C_j contain information about the model and observation errors. Here, we have assumed simple parameterized forms for the solution covariance functions. Data errors have been neglected.

SPEEDUP

The attraction with streamline simulators is the decoupling of the pressure and flow solvers. Thus by opting for fewer pressure solves, one achieves a speedup in the simulation at the expense of accuracy. Our principal aim is to rapidly explore the parameter space to identify the good history matching regions. We can then switch to a conventional finite-difference simulator and selectively explore the identified parameter space regions. Hence the aim is maximum speed at a slightly compromised accuracy.

In this paper, we shall use a streamline simulator, 3DSL, to obtain a speedup in exploring parameter space. Basically, 3DSL solves a 3D problem by decoupling it into a series of 1D problems. Each of the 1D problems is then solved along a streamline, and then mapped back onto the underlying Cartesian grid to obtain a full 3D solution at a new time level. In contrast to conventional simulators, the streamline simulation is based on fluid transport along a dynamically changing streamline-based flow grid. Thus large time steps can be taken without numerical instabilities and giving the method a near linear scaling in terms of CPU efficiency versus model size, (Thiele et al. 1997), (Baker et al. 2001).

RESULTS AND DISCUSSIONS

Using 3DSL, we generated 984 history matching models. We simulated using a time-step of 25 days, a total history matching period of 300 days and a single pressure solve update. The total CPU time for the ensemble was 2.84 hours. This gives an average of 0.17min per model. If we update the pressure solve more frequently, e.g., every 10days, this factor increases to 0.35min per model. Thus using a single pressure solve gives a speed up of about 48% in our particular case. In the sequel we see that this massive speed up helps in the rapid exploration of the parameter space without remarkably compromising the accuracy of our predictions. Before switching to a conventional finite-difference simulator, one needs to determine the regions in parameter space to explore further, i.e., the ranges of the parameter. To do this, we plot the 1-dimensional posterior probability distribution of each of the parameters. These are projections of the multidimensional posterior distribution onto each of the 12 parameter axes. Fig. 2 shows the 1D marginal probability plots for all the parameters, generated using a single long chain of the MCMC walk.

As mentioned earlier, our prediction of the fine grid behaviour is uncertain. Hence we quantify the uncertainty with predicting the fine grid behaviour from 300 to 2000 days. We approach this in two ways.

Firstly, we register the nearest neighbour Voronoi cells visited during the MCMC random walk. Then we run a full simulation on each of the nearest neighbour cells and plot the profiles. Fig. 3 shows the distribution for oil and water. Also plotted in Fig. 3 are the maximum likelihood model as well as the fine grid (true) data. These plots show that models visited during the MCMC walk form an envelope about the true solution. Our coarse grid predictions are very close to the fine grid data. This envelope however, may be broader than the true uncertainty envelope since the influence of each model in defining the uncertainty depends on its posterior probability. We note that during this walk, new history matching models are also created, which were not in the original ensemble.

In the second instance, we assume Gaussian statistics and, using the models generated during the MCMC walk, perform a Bayes update of the probabilities. Thus this approach takes into account the relative influence of each of the models visited during the MCMC walk. We then calculate the first and second moments (the mean and standard deviation) of the profiles. Using these parameters, we determined the P_{10} and P_{90} , i.e., the 10th and 90th percentile cut-offs for each of the profiles. These cut-offs are plotted in fig. 4, together with our mean and maximum likelihood predictions. The P_{10} and P_{90} cut-offs rightly envelope our predicted fine grid solution.

CONCLUSIONS

We have demonstrated the use of the Neighbourhood Approximation algorithm in conjunction with streamline simulation to generate multiple history matching models. We have investigated using synthetic data from the SPE 10th Comparative Solution Project and shown that the maximum likelihood model is an accurate predictor of the fine grid behaviour.

By using the complete ensemble of models generated and a MCMC approach, we have demonstrated two methods of quantifying the uncertainty in future oil and water production. We have shown that the true solution lies within the uncertainty bounds predicted by the algorithm.

We investigated the use of streamline simulation to achieve speedup in exploring the parameter space and rapidly identify regions of parameter space for further exploration using a finite-difference simulator. Our results show that we achieved a useful speedup in exploring parameter space by the use of streamline simulation, without significantly compromising the accuracy of our results.

Using a MCMC method, we are able to identify regions of good fit to the data and quantify this by plotting the 1D marginal probability distribution of the parameters.

ACKNOWLEDGEMENT

We thank BP for funding this work as part of the ‘Prediction Under Uncertainty’ project at Heriot-Watt University.

REFERENCES

- Baker, R. O., Kuppe, F., Chugh, S., Bora, R., Stojanovic, S., and Batycky, R. (2001). "Full-Field Modeling Using Streamline-Based Simulation: 4 Case Studies." *SPE*, (SPE 66405), 1–11.
- Chierici, G. L. (1981). "Novel Relations for Drainage and Imbibition Relative Permeabilities." *SPE*, (SPE 10165), 1–10.
- Christie, M. A. (November 1996). "Upscaling for Reservoir Simulation." *JPT*, (SPE 37324), 1004–1008.
- Christie, M. A. and Blunt, M. J. (2001). "Tenth SPE Comparative Solution Project: A Comparison of Upscaling Techniques." *SPE*, (SPE 66599), 1–9.
- Cunha, L. B., Oliver, D. S., Rednar, R. A., and Reynolds, A. C. (1998). "A hybrid Markov chain Monte Carlo methods for generating permeability fields conditioned to multiwell pressure data and prior information." *SPE Journal*, 3(3), 261–271.
- Glimm, J. and Sharp, D. (December 1998). "Stochastic Methods for the Prediction of Complex Multiscale Phenomena." *Quarterly of Applied Mathematics*, LVI(4), 741–765.
- Neal, R. M. (25 September 1993). "Probabilistic Inference Using Markov Chain Monte Carlo Methods." *Technical Report CRG-TR-93-1*, University of Toronto, Department of Computer Science.
- Okabe, A., Boots, B., and Sugihara, K. (1992). *Spatial Tessellations - Concepts and Applications of Voronoi diagrams*. JOHN WILEY & SONS, ENGLAND.
- Oliver, D. S., Cunha, L. B., and Reynolds, A. C. (1997). "Markov chain Monte Carlo methods for conditioning a permeability field to pressure data." *Mathematical Geology*, 29(1), 61–91.
- Omre, H., Tjelmeland, H., and Wist, H. T. (June, 1999). "Uncertainty in History Matching - Model Specification and Sampling Algorithm." *Preprint 6/1999*, Norwegian University of Science and Technology, Trondheim, Norway.
- Sambridge, M. (1999a). "Geophysical inversion with a neighbourhood algorithm -I. Searching a parameter space." *Geophysics J. Int.*, (138), 479–94.
- Sambridge, M. (1999b). "Geophysical inversion with a neighbourhood algorithm -II. Appraising the ensemble." *Geophysics J. Int.*, (138), 727–745.
- Thiele, M. R., Batycky, R. P., and Blunt, M. J. (1997). "A Streamline-Based 3D Field-Scale Compositional Reservoir Simulator." *SPE*, (SPE 38889), 1–12.

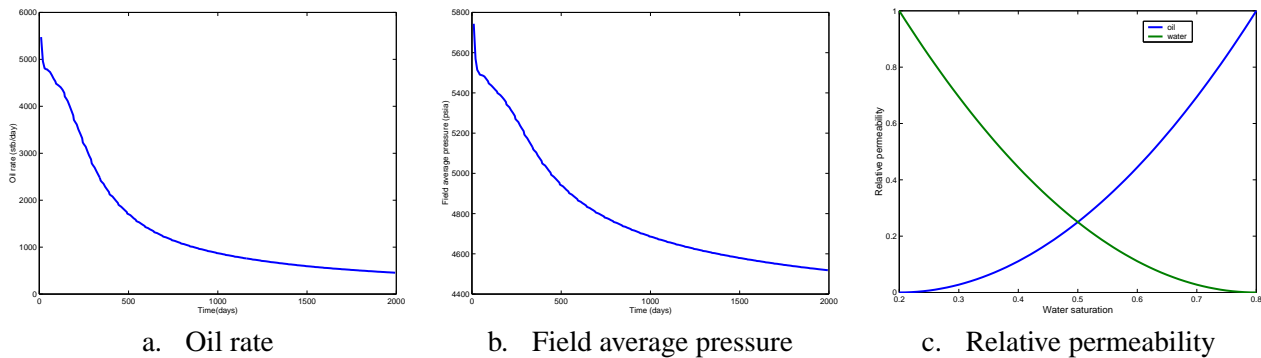


FIG. 1. Fine grid oil rate, pressure and relative permeability data

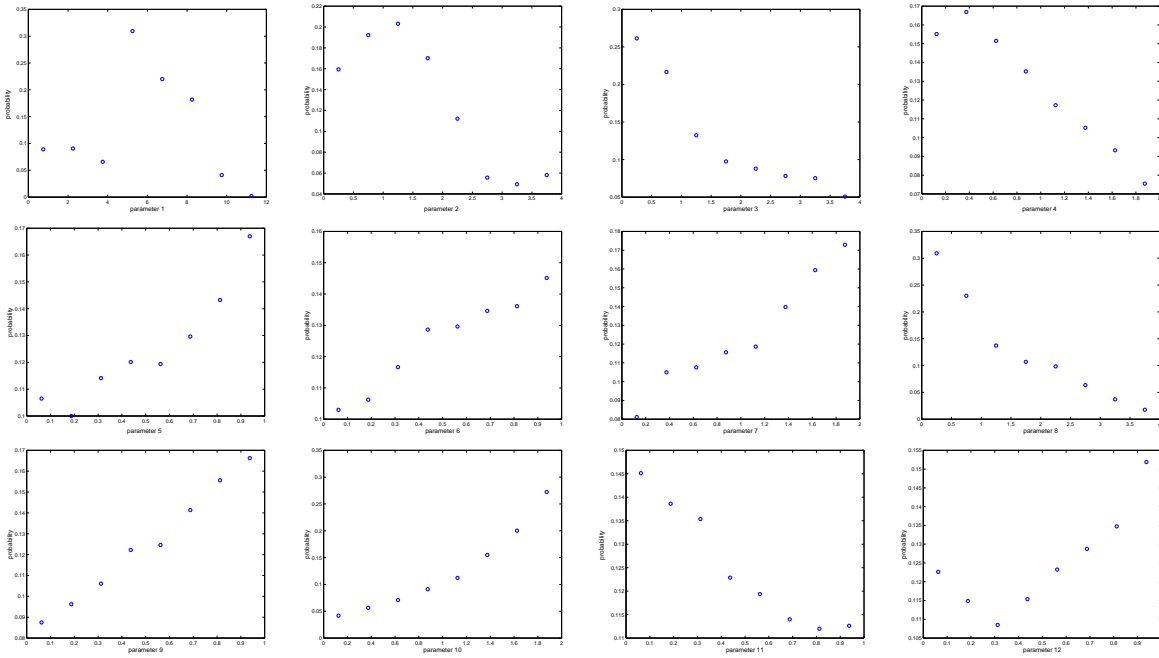


FIG. 2. 1D marginal probabilities (Parameters 1 - 12)

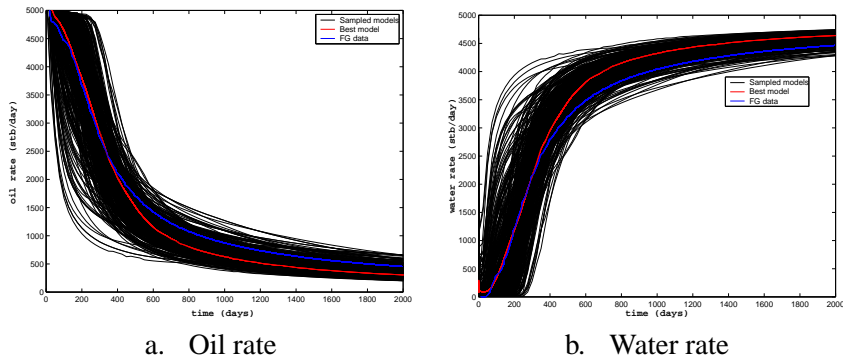


FIG. 3. Quantifying uncertainty - I

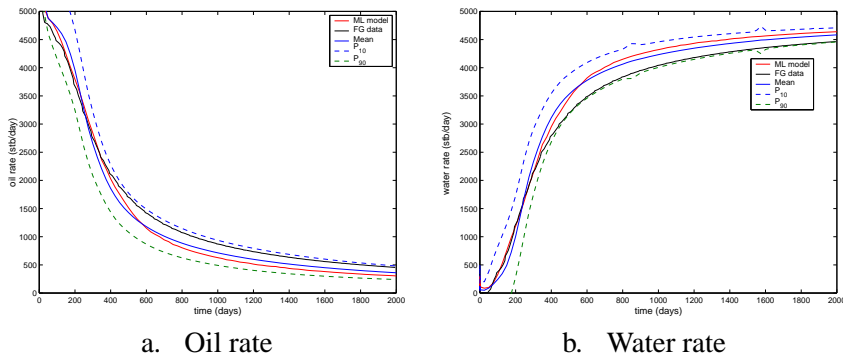


FIG. 4. Quantifying uncertainty - II