

How Does Sampling Strategy Affect Uncertainty Estimations?

D. Erbaş and M. Christie

Institute of Petroleum Engineering, Heriot-Watt University, Edinburgh, EH14 4AS, UK
e-mail: demet.eras@pet.hw.ac.uk - mike.christie@pet.hw.ac.uk

Résumé— Comment la stratégie de l'échantillonnage affecte-t-elle les estimations d'incertitude ? — Les techniques d'inférence bayésienne pour déterminer les incertitudes concernant le comportement d'un gisement demandent la création de multiples modèles conditionnés par les données du terrain. L'identification des bons modèles nécessite l'échantillonnage d'un espace de paramètres de grande dimension. Dès lors, la capacité de la technique d'inférence dépend fortement de la qualité des méthodes utilisées pour générer les modèles qui expliquent bien les données. Afin d'améliorer cette capacité, les facteurs affectant les estimations d'incertitude doivent être identifiés. Cette étude vise à examiner l'effet de différentes stratégies d'échantillonnage sur l'estimation des incertitudes de prédiction. Un modèle synthétique de gisement a été étudié pour comparer les estimations d'incertitude aux résultats de différents échantillonnages réalisés en utilisant les algorithmes génétiques (GA) et l'algorithme de voisinage (NA). Les principales différences des résultats d'échantillonnage de GA et NA sont le niveau d'exploration de l'espace de paramètres et le nombre de régions donnant des modèles s'ajustant correctement sur les données. Nous démontrons que différentes stratégies d'échantillonnage peuvent aboutir à des estimations d'incertitude significativement différentes. Nous démontrons également que la capacité prédictive des modèles obtenus par calage d'historique peut être utilisée comme un indicateur de la dispersion de la distribution ultérieure de probabilité a posteriori.

Abstract— How Does Sampling Strategy Affect Uncertainty Estimations?— Bayesian inversion techniques for assessing reservoir performance uncertainties involve generating multiple reservoir models conditioned to the available field data. This process requires sampling in a high-dimensional parameter space to identify good data-fitting models. Therefore, the robustness of the inversion technique strongly depends on the efficiency of the sampling methods used to generate the reservoir models. In order to improve the robustness, the factors affecting the uncertainty estimations must be identified. This paper aims to investigate the effect of sampling strategies on the prediction uncertainty estimations. A synthetic reservoir model has been studied to compare the uncertainty estimations based on different sampling outcomes obtained using Genetic Algorithms (GA) and Neighbourhood Algorithm (NA). The main differences in the sampling outcomes from GA and NA are the degree of exploration and the number of good data-fitting regions identified in the parameter space. We show that different sampling strategies may result in significantly different uncertainty estimates. We also demonstrate that the predictive capability of the history-matched models can be used as an indicator for the spread of the posterior probability distribution.

INTRODUCTION

Quantifying uncertainties in reservoir performance predictions is important, as high-cost investment decisions rely on these predictions. Our knowledge of the subsurface is always limited to small-scale samples taken from a limited number of points in the field. Moreover, the errors in data measurements, experimental results and the mathematical reservoir modelling bring additional uncertainties. Therefore, it is impossible to make accurate predictions. However, we can estimate the probability of the predictions to assess the risks of our decisions.

The main issues with making probabilistic predictions are the difficulty of incorporating all sources of uncertainties and the requirement for a high number of reservoir simulations. Bayes theorem [1] allows incorporation of all sources of uncertainties including data, simulation and interpolation errors.

There are 3 main factors to be investigated for robustness in Bayesian workflows:

- Prior knowledge,
- Available data,
- Sampling quality.

Prior knowledge comes into the workflow in determining the uncertain parameters of the history-matching problem, along with the possible ranges that these parameters fall in. Inadequate parameterization would result in inaccurate uncertainty estimations. Observed data (indirect measurements such as production rate or pressure) or direct measurements (e.g. porosity or permeability measurements from core analysis) can be used to calculate the likelihood term in Bayes' formula. Incorrect likelihood definition or poor error handling could result in biased uncertainty estimates.

The third factor, sampling quality, is related to how effectively we explore the parameter space when generating multiple models conditioned to the available data. Usually an optimization technique is employed to find good history-matched models with a reasonable number of simulation runs. The most common techniques include gradient optimization [2], gradual deformation [3], experimental design [4, 5], stochastic optimization techniques, such as Genetic Algorithms (GA) [6-8], Simulated Annealing (SA) [9], and more recently the Neighbourhood Algorithm (NA) [10-12] and combination of these methods [13, 14]. Some of the uncertainty quantification methods have already been compared on benchmark problems [15, 16].

Gradient methods lead to the determination of a single history-matched model, and thus neglect the inherent non-uniqueness of the solution of the ill-posed history matching problem. Stochastic methods, on the other hand, have the potential to find multiple minima of the parameter space, and therefore could conceptually handle the problem better.

The posterior probability distribution might vary significantly with different sampling schemes, depending on how

the parameter space is explored. This study aims at investigating how different sampling strategies affect the estimated uncertainty interval in recovery predictions, given a certain parameterization and data.

We use Genetic Algorithms (GA) and the Neighbourhood Algorithm (NA) to generate an ensemble of reservoir models for a synthetic reservoir case, the IC Fault model [17]. These algorithms also have tuning parameters that can be adjusted to change the search behaviour, global exploration or local refinement. We then apply a Gibbs sampler for posterior inference from the generated ensemble to calculate the uncertainty interval of the predictions. The uncertainty range is reported as P10-P50-P90 values, which indicate the recovery predictions below which the truth lie with 10%, 50% and 90% probabilities respectively. Uncertainty ranges obtained from different sampling strategies are compared, with the intention to identify the factors causing the spread of the distribution, thus help improve robustness of the overall methodology.

We analyze three main points:

- Choice of algorithm: Different search algorithms might produce ensembles of different quality in terms of the degree of exploration and the number of distinct minima identified in the parameter space. In Section 2.1, we compare forecast uncertainty estimations based on different ensembles generated by GA and NA.
- Predictive capability: When sampling, the algorithms are directed by the measure of how well the model fits the history data. This misfit measure is also used when calculating the likelihood of the models. However, the uncertainty interval in the predictions is also affected by the forecast behavior of the models generated.

It is principally believed that the more distinct minima identified in the parameter space, the wider the uncertainty interval achieved. There have been efforts to improve the algorithm's performance in finding distinct multiple minima [18]. However, we demonstrate that identifying more distinct minima does not always guarantee obtaining wider uncertainty intervals; but the predictive capability of the models gives indications about the spread of the distribution. In Section 2.2, we investigate how predictive capability can be quantified and used as an indicator for the spread of the posterior probability distribution.

Number of models: Practically, the number of models to be generated for an uncertainty study is determined by the time available to run reservoir simulations. Every single model we generate is an additional data point for the probability calculations. But, does the posterior probability distribution converge with the number of models? We demonstrate how the distribution evolves throughout the sampling period in Section 2.3.

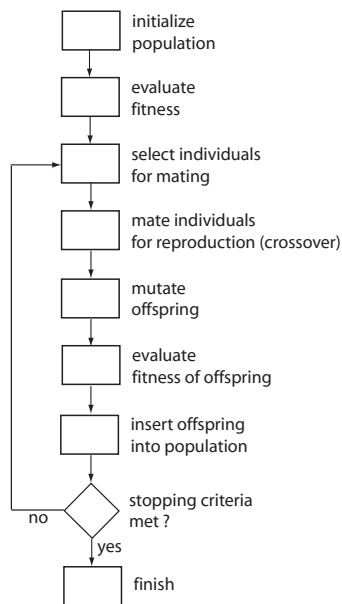


Figure 1

Genetic Algorithm work ow.

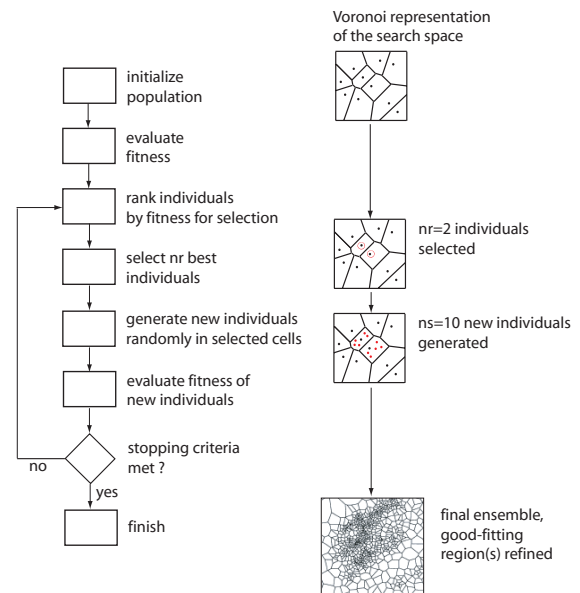


Figure 2

Neighbourhood Algorithm work ow.

1 METHODOLOGY

Our Bayesian methodology can be divided into two parts;

- Sampling the parameter space,
- Posterior inference.

Here we discuss GA and NA specifications regarding the sampling stage and briefly explain NA-Bayes (NAB) for the posterior inference.

1.1 GA Application

GAs are inspired by Darwin's theory of natural evolution [19]. The algorithm searches for possible solutions in parameter space by mimicking crossover, mutation and natural selection phenomena and employing a "survival of the fittest" strategy.

A generic GA work ow is given in Figure 1. The algorithm starts with a randomly generated initial population of individuals and computes the fitness of the individuals using the objective function. It then selects and mates the parents to produce new offspring by crossover and mutation operations. In crossover, the parameters are swapped between models, while in mutation the parameters are randomly perturbed. The new offspring is then evaluated and inserted into the population. If the stopping criteria are met the algorithm terminates, otherwise it goes back to the selection process. Stopping criteria could be the maximum number of generations to be produced or a convergence measure defined by the user.

GA can be tuned by two parameters controlling the probability of crossover and mutation operations, (p_c and p_m respectively). Crossover provides a means for information exchange between individuals (reservoir realizations in our case), while mutation adds diversity to the population when generating new realizations. Our previous sensitivity runs [20] showed that low p_c and high p_m would be the most appropriate choice for the history-matching problem. The number of models to generate at each generation, n_s , also determines the exploration capability: higher n_s leads to a more exploratory search.

We use a steady state GA (SSGA) in this study, where the algorithm replaces a certain proportion of the population in each generation, rather than replacing the entire population. Therefore SSGA works with overlapping populations and falls into "non-generational" GA category [21].

1.2 NA Application

NA is a stochastic search method for finding models of acceptable data fit in a multidimensional parameter space. It was originally developed for solving a seismic waveform inversion problem by Sambridge [22].

A brief NA work ow is given in Figure 2. Starting from a randomly generated set of models, similar to GA, the algorithm first computes the fitness of the models using the objective function. It then determines the nearest neighbour region of each model in the parameter space by constructing

the Voronoi diagram [23]. An example of the Voronoi representation and the evolution of a 2-parameter space is given beside the flow chart in Figure 2.

The algorithm then chooses the best models according to the fitness scores and generates new models by a uniform random walk within the Voronoi cells of the models chosen. Therefore, at each iteration models are re-sampled and n_s/n_r models are generated in each Voronoi cell chosen. After evaluating the new models, the Voronoi diagram is updated and the same procedure is repeated until the stopping criterion, i.e. the maximum number of iterations specified by the user, is met.

The total number of models generated after a number of iterations can be calculated by Equation 1:

$$N = n_{init} + (n_{iter} \times n_s) \quad (1)$$

Our previous studies with NA demonstrated that the algorithm identified the good-fitting regions in the parameter space and further refined them locally. However, this refinement sometimes resulted in trapping in a local minimum. The algorithm over-refined the region after this point. We developed a methodology [20] to detect when NA got trapped and started over-refining the local minimum. The method was based on quantifying how much extra information was gained by going from one iteration to the next. The information contained at each iteration was quantified by the “average misfit (M_{avg})” and the “minimum misfit (M_{min})” over the current iteration. Then, taking the first derivative of these indicators gave us how much extra information we gained going from one iteration to the next. Setting a tolerance level, which could be determined depending on how much refinement was required, we could identify the iteration where NA started over-refining. The method is summarised in Equation 2.

If,

$$M_{avg}(j) \dot{\sim} M_{avg}(j \dot{\sim} 1) \quad \epsilon_1 \quad j \in \{1 \dots n_{iter}\} \quad (2)$$

and

$$M_{min}(j) \dot{\sim} M_{min}(j \dot{\sim} 1) \quad \epsilon_2 \quad j \in \{1 \dots n_{iter}\} \quad (3)$$

trapped at j -th iteration, where ϵ_1 and ϵ_2 are the tolerances

Algorithm parameters n_s and n_r can also be tuned to change the search behavior; high n_{init} increases the exploration, while high n_s/n_r ratio results in faster convergence to a local minimum, but reduces the exploration. Consequently, several restarts from different initial populations with high n_s/n_r ratio is an efficient way of using NA to locate distinct multiple minima.

1.3 Posterior Inference (NAB)

Having generated an ensemble of models, the next step is to infer the information in this ensemble in terms of a probability distribution to calculate the forecast uncertainties.

We use NA-Bayes (NAB) for posterior inference [24]. NAB uses a Gibbs sampler to importance sample the posterior probability density of a given ensemble. The Gibbs sampler performs a random walk on the Voronoi surface and samples the models based on the conditional probability distribution proportional to the likelihood of the models. The likelihood is defined by Equation 4 assuming Gaussian statistics.

$$L = e^{-\sum M} \quad (4)$$

where, M is the least square misfit.

The main advantage of NAB is that it approximates the posterior probability distribution without the need for further running reservoir simulation of each model re-sampled by NAB.

2 IC FAULT MODEL CASE STUDY

The IC Fault model is a simple 3-parameter model set up by Tavasoliet al. [17] as a test example for automated history matching. It has proved extremely difficult to history match and Carter et al. [25] concludes that the best history matched model obtained from an exhaustive run (1691 models) is a very poor predictor. We use IC Fault model in this work to demonstrate how different sampling strategies affect uncertainty estimations.

The geological model consists of 6 layers of alternating high and low permeability sands with gradually decreasing thicknesses downward (Fig. 3). There is a water injector on the left and oil producer on the right hand side of the model. The simple fault in the middle alters the contact between good and poor sands, so determines the efficiency of the water flood along with the layer permeabilities. Taking the throw thickness and the two permeabilities as the uncertain parameters, we applied our methodology to estimate oil recovery prediction uncertainties.

The prior ranges of the uncertain parameters and the chosen truth case are given in Table 1.

The misfit (M) of the models is calculated by the sum of least-squared error (Equation 5).

$$M = \frac{1}{n} \sum_p \sum_{t=1}^n \frac{(\text{obs} - \text{sim})_t^2}{2\sigma_t^2} \quad (5)$$

TABLE 1

IC Fault model uncertain parameter ranges

Parameter	Minimum	Maximum	Truth Case
k_{high} , md	100	200	131.6
k_{low} , md	0	50	1.3
h (throw), ft	0	60	10.4

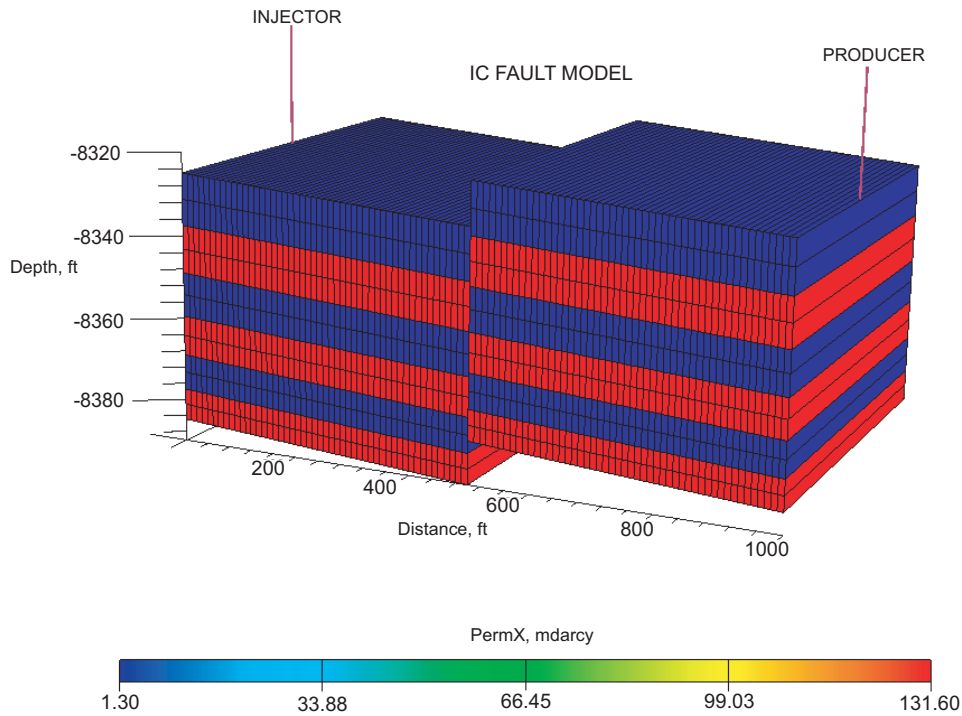


Figure 3
IC Fault model.

where,
 p: oil and water production rates
 obs: observed data (truth case)
 t: time
 sim: simulated result
 n: number of observed data
 σ_t : $0.03 \times \text{obs}$

the throw-axis (Fig. 5d), the mis t landscape is more complicated as it has several steep minima mainly within three different regions (0-104, 2332-2927 and around 499). Because of the complex mis t surface structure, it is a very challenging problem for a search algorithm to locate these local minima. The next section demonstrates how NA and GA performed in searching good-ting regions and how

The truth case oil and water production results (Fig. 4) are used as the “observed data” in mis t calculation. Oil and water production results from the generated models are referred to a “simulated result” in this report.

Figure 5a shows the search volume (3D cube) constructed by the database of 1561 uniformly sampled models. The colors indicate the model mis t. When calculating the mis t, we cut o the simulated water production at the observed breakthrough time, as specified by Carter.

The shape of the mis t surface determines the quality in sampling. Figures 5b, 5c and 5d are 1D cross-sections of the mis t surface in each of the 3 axes, where two of the parameters are fixed at the truth and the other is varied. All three cross-sections show that the mis t surface has very steep minima. Along the k_{ow} -axis (Fig. 5b) there is only one sharp minimum, which is at the truth. However, along the k_{ow} -axis (Fig. 5c) the truth lies within a flat region of good-ting models where k_{ow} varies between .3 and 68. Along

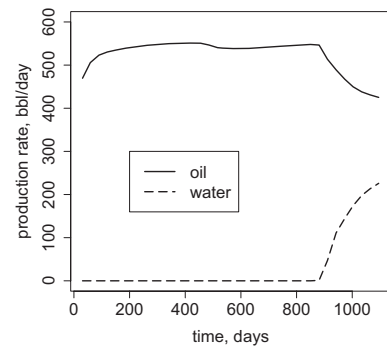


Figure 4
Truth case production data (observed data).

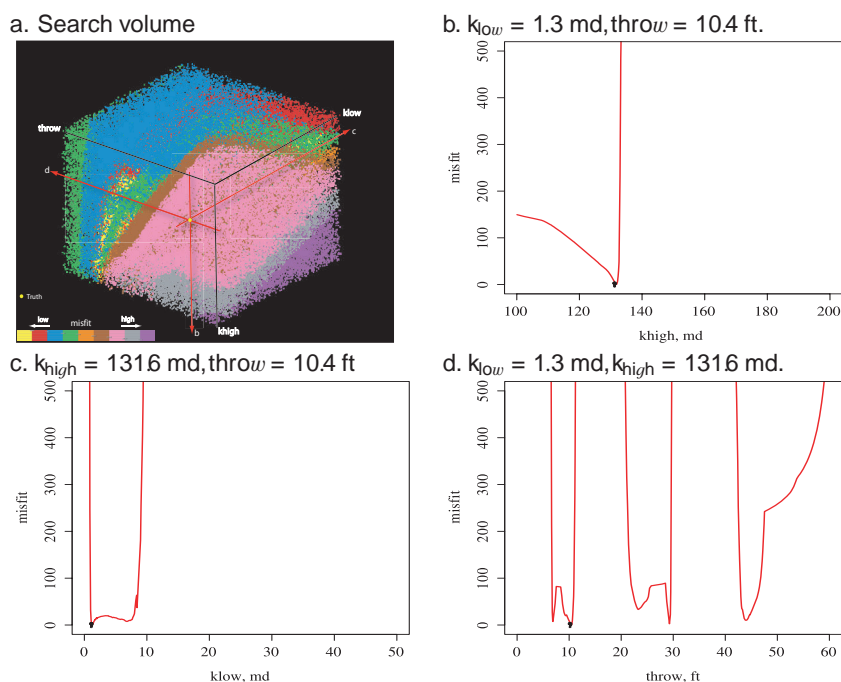


Figure 5

a. The search volume of the IC Fault problem. Three plots are the cross-sections of the misfit surface at the truth along the three parameter axes as shown. b: High permeability. c: Low permeability. d: Throw axes. Black circles show the truth.

TABLE 2

GA and NA sampling parameters and results

	GA	NA1	NA2
n_{init}	50	50	200
n_s (popsize for GA)	50	50	100
n_r	-	50	100
total number of models	3522	1700	3400
models with $M \leq 25$	3%	58%	67%
best model misfit	0.95	0.29	0.09

different sampling strategies resulted in different uncertainty interval estimations.

2.1 Choice of Algorithm

Taking the database as the reference case, we ran 3 sampling cases: GA, NA1 and NA2. Algorithm specifications of these cases are given in Table 2. In NA1 and GA the number of models generated at each iteration were the same (50), while in NA2 initial population was larger ($n_{init} = 200$) and 100 models were generated at each iteration. Therefore, NA2 was expected to perform more exploratory search than NA1. Note that in both NA cases, n_s/n_r ratio was 1, which is the most exploratory way of running NA in terms of n_s/n_r ratio.

GA was stopped at 140th generation. With 50% replacement scheme in SSGA, 3522 models were generated in total.

Both NA cases were stopped when we detected that the algorithm got trapped in a local minimum by the methodology explained in Section 1.2. When monitoring the minimum and the average misfits, the tolerances ϵ_1 and ϵ_2 were equal to 0.1. NA1 ended up an ensemble of 1700 models, while NA2 did 3400. The reason for NA2 generating more models until the over-representation stage was because NA2 was exploring more extensively than NA1, as it had higher n_{init} and n_s .

From the optimization point of view, minimizing the misfit, NA cases worked better than GA, as the best model misfits for both NA cases are lower than the GA case (Table 2). The proportion of the models with $M \leq 25$ is much lower in the GA ensemble than NA cases. This is because NA refines good-fitting regions locally and generates clusters of models in these regions.

Figure 6 shows the good-fitting models of the database case (M = 25), (Fig. 6a), and the ensembles generated by GA, NA1 and NA2 (Figs. 6b, 6c and 6d respectively) in 3D parameter space. We take the database as the reference case representing the true misfit surface, so we expect GA and NA to identify the ribbon-like structure shown in Figure 6a. The notable difference between the GA ensemble and the NA cases is that GA models spread all over the cube, while NA forms clusters around the good-fitting regions. NA1 has 4 clusters that don't encapsulate the truth (red dot). How-

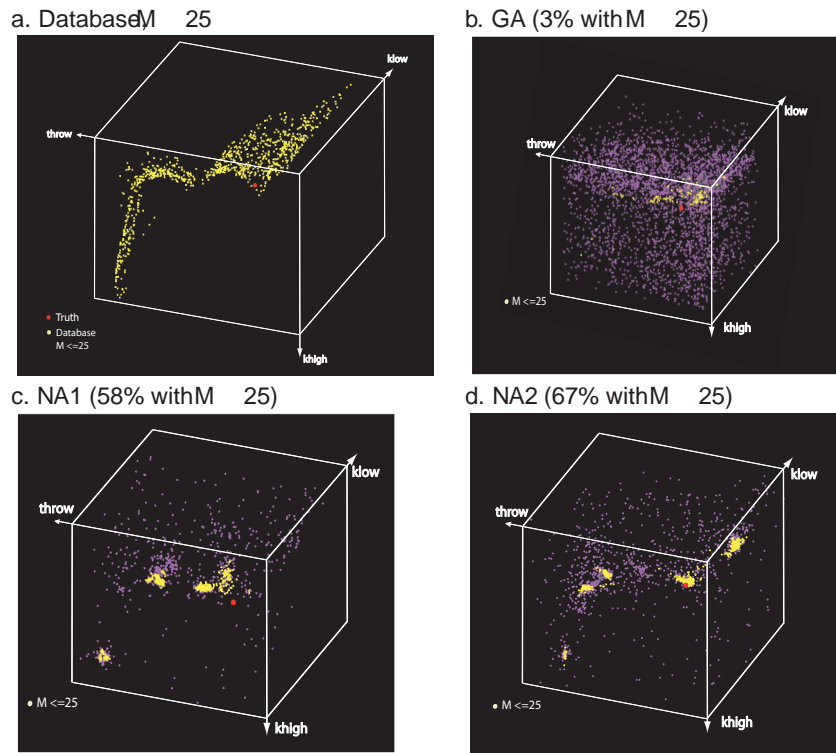


Figure 6

a: Database models with $M \leq 25$. b: GA ensemble. c: NA1 ensemble. d: NA2 ensemble. Purple dots are all the models with $M > 25$.

ever, NA2 has a cluster around the truth, and identifies different probability distributions. The next section explains distinct clusters in total, one of which is further back on the k_{low} axis, which is different from the NA1 ensemble.

2.2 Predictive Capability of the Models

Having examined the main differences in the generated ensembles, we then compare the uncertainty estimations computed based on them. Figures 7a, 7b, 7c and 7d are recovery prediction uncertainties computed based on the pre-database, GA, NA1 and NA2 ensembles respectively. The P90-P10 difference at the end of the prediction period is given on top of each plot for comparison.

The distributions are also displayed as boxplots in Figure 8. All 3 cases underestimate the uncertainty compared with the database result. According to the inference from the database, the truth is below the P10 line, while GA and NA2 results encapsulate the truth within their distribution. The NA1 ensemble produces a very narrow distribution.

Linking to the ensemble characteristics, we could say that high diversity in the GA ensemble results in a wider distribution, as opposed to localized clusters in the NA1 ensemble. However, the NA2 ensemble also produces a wide spread similar to GA, although it contains localized clusters. Looking at the ensemble characteristics like the diversity and the number of good-fitting regions identified is not sufficient to understand why NA1 and NA2 result in significantly different forecast behavior of the models. Good predictors are the ones that forecast similarly to the truth, while bad predictors diverge from the truth in the forecast period.

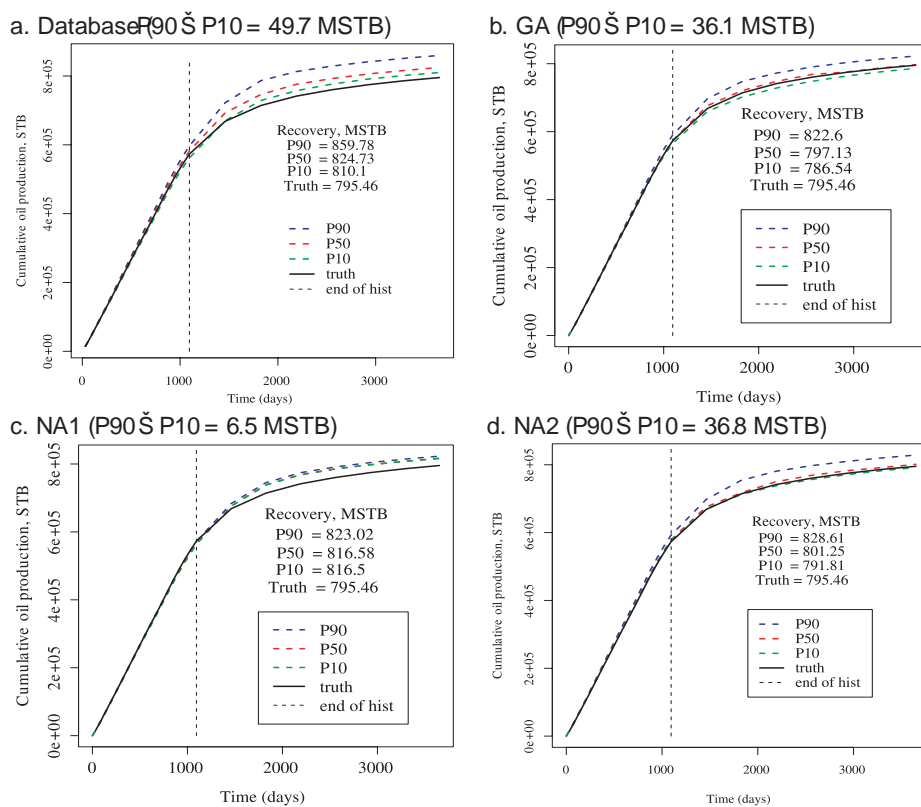


Figure 7

Recovery prediction uncertainties based on: a. The database. b. GA ensemble. c. NA1 ensemble. d. NA2 ensemble. Between P90 and P10 estimations at the end of the prediction period are given on top of each plot for comparison.

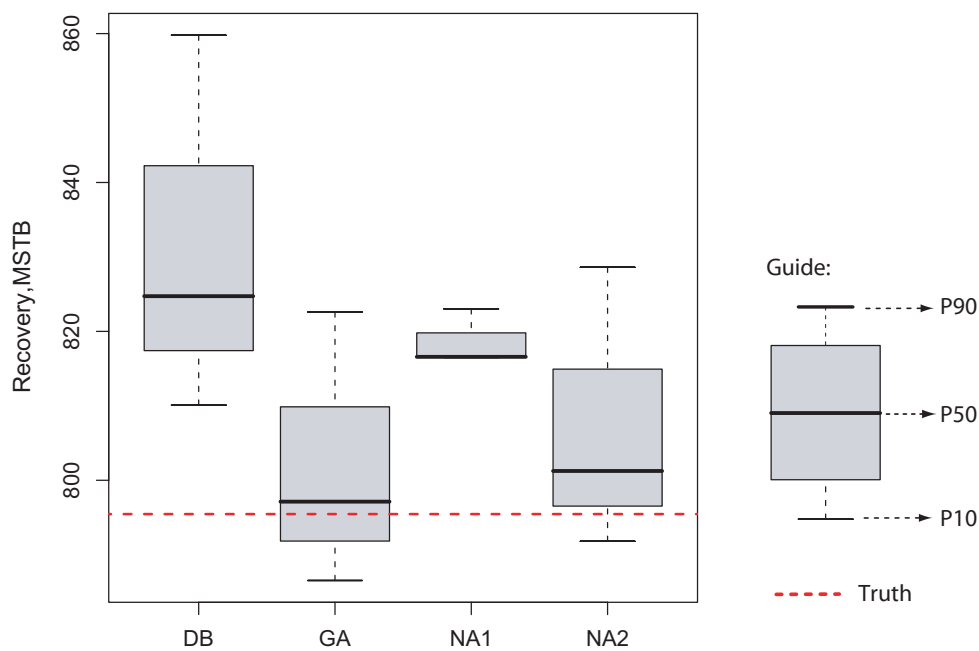


Figure 8

Boxplots of the prediction uncertainties obtained based on the database, GA, NA1 and NA2.

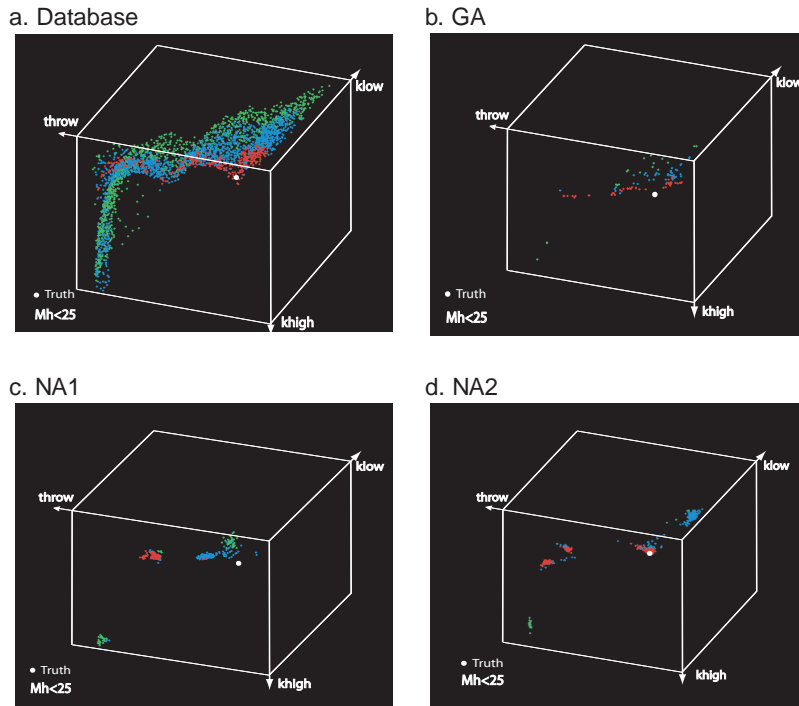


Figure 9

Forecast misfit (M_f) of good history-matched ($M_h < 25$) in a. the database, b. GA ensemble, c. NA1 ensemble, d. NA2 ensemble. Red: $0 < M_f < 25$, Blue: $25 < M_f < 50$, Green: $M_f > 50$.

In the database good predictors (red dots) spread along the edge of the ribbon, close to the throw axis (Fig. 9a). In other words, k_{low} and k_{high} of these good predictors are almost fixed around the truth, but the $throw$ varies all over the range. We can see this aligned structure in the GA ensemble as well (Fig. 9b). Among the clusters that NA1 generated, one of them contains good predictors, while NA2 3 out of 5 clusters contain good predictors.

In order to capture the true uncertainty in our predictions it is necessary to find the “bad” predictors (if there is any), i.e. models that diverge from the truth in the forecast period, which are shown in blue and green in these pictures. Cross-plot of M_h and M_f , Figure 10, shows more clearly how our good history-matched models behaved differently in the forecast period. Basically, the models aligned near to the y -axis have very low M_h , i.e. fit history data very well, but diverge in the forecast. These models are called “diverging models” in this paper. Circled models on these plots indicate NAB re-sampled models. Note that mostly low misfit (or high likelihood) models are re-sampled by NAB (recall Equation 4).

In the database (Fig. 10a), the number of diverging models is higher than the other cases, so the uncertainty ranges are the widest. P90 - P10 differences are given on top of each plot.

Comparing the two NA cases (Figs. 10c and 10d), the NA1 models with low history misfit (less than 5) are clustered within a narrow range (20-25) on the y -axis, however the NA2 models mostly lie between 0 and 26. This explains the big difference between the uncertainty range inferred from NA1 and NA2. GA models are also spread over the y -axis similar to the NA2 picture. Although there are fewer models in GA than NA2, the uncertainty ranges are almost the same. In this particular problem, with only a few models from GA ensemble (note the black circled dots) we obtain almost the same uncertainty range with more refined NA2 case, where NAB resamples much more models.

It is important to note here that in order to obtain a wider spread in the uncertainty range, we need to pick up diverging models from different regions in the sampling stage.

We further analyze the forecast behavior of the models in the database to see how predictive capability in short and long term forecast changes by the length of the history period. Changing the length of the history period, we plot M_h vs M_f in Figure 11. Left hand side plots show the history and two forecast periods ($f1$: short-term, $f2$: long-term) chosen for each case and right hand side pictures are the cross plots. As the misfit is very sensitive to the breakthrough time in this problem, we move the history period “before, at and after” the water breakthrough.

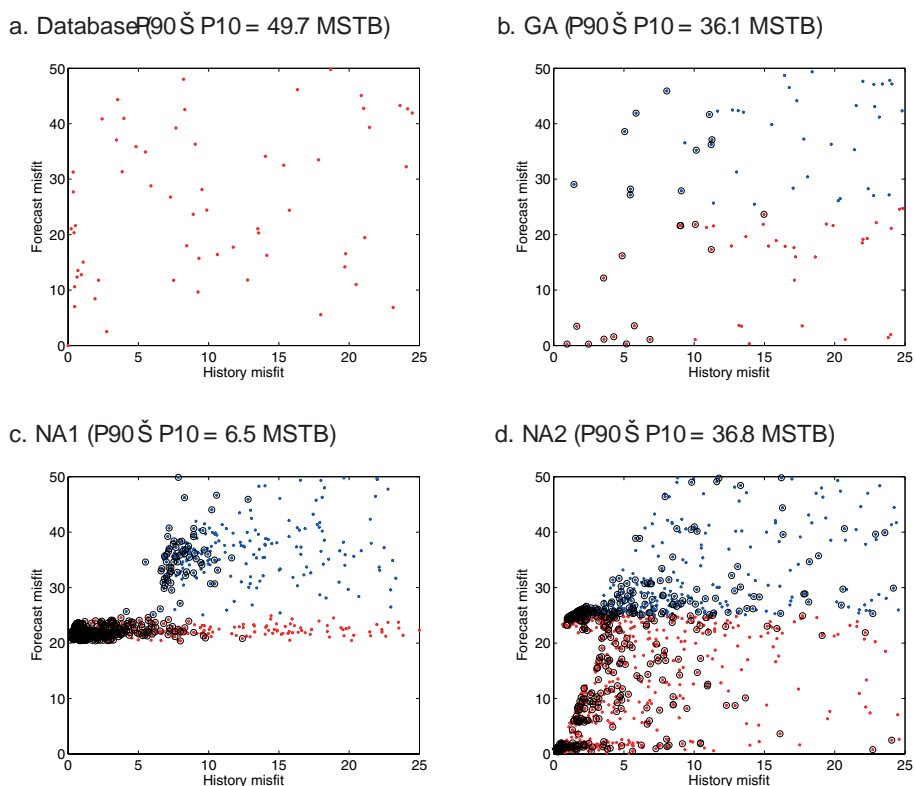


Figure 10

M_h and M_f cross-plots showing diverging models for a. the database, b. GA, c. NA1 and d. NA2. Black circles on b, c and d plots indicate the models resampled by NAB to determine the posterior probability distribution.

In case 1 the history period ends after the breakthrough (by the inner picture). Consequently, we would expect even time. Long term forecast of this case is the case we have wider uncertainty interval in this case. conducted all the analysis presented so far in this paper. The overall shape of the plots for f_1 and f_2 are similar with two distinct branches. The main difference is that f_1 is squeezed down, while f_2 has an extra branch lying on the axis. That means in long-term forecast we have more diverging models, as the inner picture, zoomed into the good history, matched models, also displays.

The case 2 history period ends at the breakthrough time (inclusive). In this case, the overall shape of f_1 and f_2 almost overlap, but the zoomed-in picture shows that there are more diverging models in the long-term forecast. Comparing the inner picture of this case with the previous one, both short and long term plots are closer to the axis in this case, which would result in a wider uncertainty range than case 1

Finally in case 3 history ends before the breakthrough. Similarly to the previous case, the short and long term predictive capability of the models are not significantly different from each other, and we have even more diverging models in this case than the previous two cases (indicated by

2.3 Number of Models

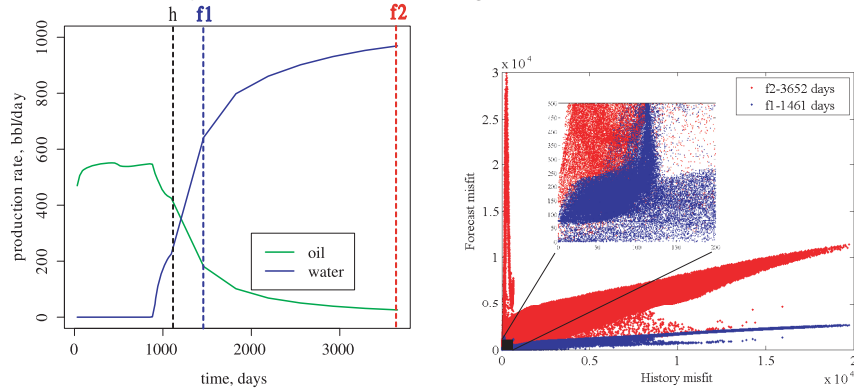
The number of models to be generated in an uncertainty study is always questioned. The common answer to that question is that the more the better, so the main practical limitation is the time available to run reservoir simulations.

However, in this section we investigate how the posterior distribution changes with the number of models generated.

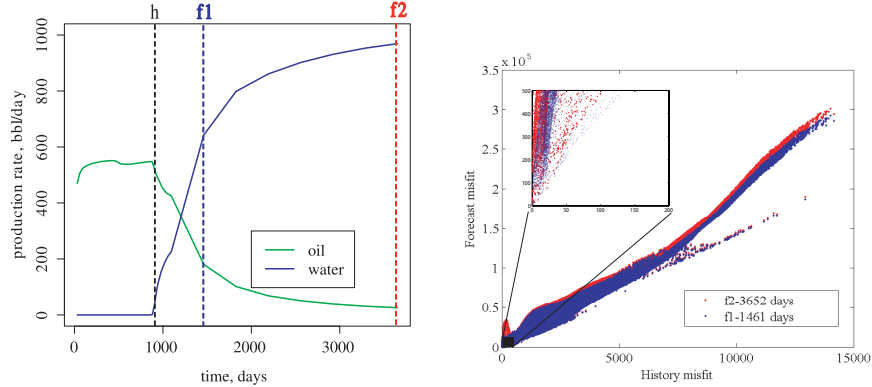
Figure 12a shows how the P10 - P50 - P90 estimates change with the number of models throughout the GA search discussed in Section 2. Uncertainty ranges are estimated at 40th, 80th, 120th and 140th generations, which correspond to 1000, 2000, 3000 and 3522 on the x-axis of this figure. We also add the result of a longer run, until 500th generation (12550 models) to see how the posterior distribution evolves. The database results are also shown on this figure for reference.

The distribution changes significantly when new models are added. In theory, when the number of models approaches infinity, the posterior distribution of the ensemble converges to the true distribution. Considering the

a. Case 1: History ends after the breakthrough



b. Case 2: History ends at the breakthrough



c. Case 3: History ends before the breakthrough

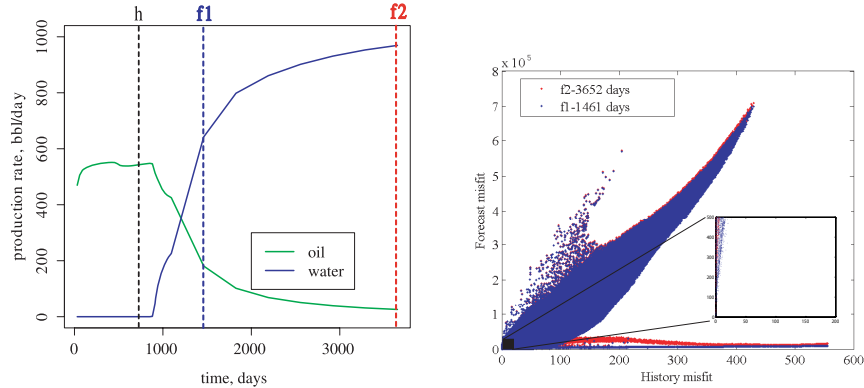


Figure 11

Variation in the predictive capability by the length of the history period. Both short (blue) and long (red) term forecasts are shown.

CONCLUSION

As the number of points get closer to the database results, we don't observe convergence with this number of samples. Plotting the same results against $1/\bar{n}$ (Fig. 12b) where \bar{n} is the number of samples, we can extrapolate the line to the point where it crosses the x-axis at 0, which would give the ranges at the synthetic reservoir model, the IC Fault model. In our case such an approximation would give us unreasonable high estimations, as it hasn't converged.

We investigated how different sampling strategies affected the uncertainty estimations of recovery predictions in a Bayesian framework. We applied the ideas on a 3-parameter model, the IC Fault model. We generated 3 ensembles of different quality using Genetic Algorithms (GA) and two variations of Neighbour-

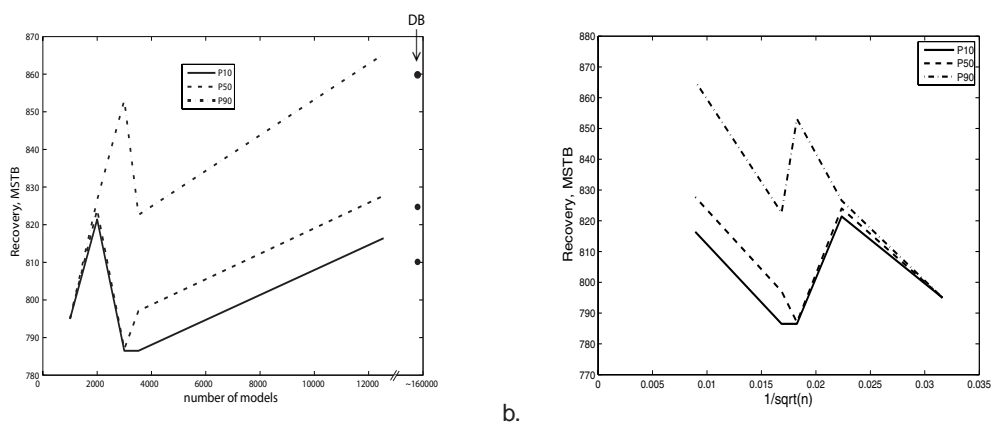


Figure 12

Change in the posterior distribution while the GA search evolves.

hood Algorithm (NA). We also had a database of nearly 160 000 uniformly sampled models as a reference case. We assessed the quality in terms of the degree of exploration, i.e. the spread of the models and the quality of minimization, i.e. number of good- tting regions identified.

- GA produced an ensemble of models spread all over the parameter space, as a result of an exploratory search:
 - no distinct clusters were observed,
 - the region that the truth case lay within was identified.
- In both NA cases, the degree of exploration was much lower than the GA case:
 - both NA cases created distinct clusters around the good- tting regions,
 - NA1 didn't locate the region around the truth, while NA2 did.
- NA cases were different in terms of their tuning parameters. NA2 had higher n_{init} , n_s and n_r values than NA1, hence performed more exploratory search. As a result, NA2 ended up with more clusters at the cost of doubling the simulation runs.
- 3% of the GA models were good- tting, based on a mis- t cut-o of 25, compare with 58% in NA1 and 67% in NA2. The best- tting models from both NA1 and NA2 cases had better match quality than the GA best model. So, NA worked better than GA in terms of the quality of minimization.
- Comparing the recovery uncertainty estimations, GA and NA2 produced much wider P10-P90 ranges that encapsulated the truth, while the NA1 ensemble resulted in a very narrow range. As a conclusion, more exploration in sampling would more likely result in a wider uncertainty range.
- Apart from the quality of the ensemble, the predictive capability of the models also acted the uncertainty range. The good- tting models that diverged in the fore-

cast period determine the shape of the posterior probability distribution. The more diverging models an ensemble contains, the wider the uncertainty range it produces.

Based on the learning from this study, the robustness of the methodology can be improved by a multi-objective optimization scheme in the sampling stage, where the history mis t is minimized while the forecast divergence is maximized at the same time. In a real eld study the forecast divergence can be calculated from a chosen reference case. The application of the new scheme is left for future work.

ACKNOWLEDGEMENTS

This work was carried out under the Uncertainty Project at Heriot-Watt University. We would like to acknowledge the project sponsors: Anadarko BG, BP, ConocoPhillips DTL, ENI, JOGMEC Landmark Graphics, Shell, and Statoil. The authors also thank Schlumberger-GeoQuest for use of the ECLIPSE reservoir simulator.

REFERENCES

- 1 Sivia, D.S. (1996) Data Analysis - A Bayesian Tutorial, Clarendon Press, Oxford.
- 2 Lépine, O. et al. (1999) Uncertainty Analysis in Predictive Reservoir Simulation Using Gradient Information, SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana number 57594.
- 3 Liu, N. and Oliver, D.S. (2004) Experimental Assessment of Gradual Deformation Method, Math. Geol. 36, 65.
- 4 Feraille, M. and Roggero, F. (2004) Uncertainty Quantification for Mature Field Combining the Bayesian Inversion Formalism and Experimental Design Approach, 11th European Conference on the Mathematics of Oil Recovery, Cannes, France.

- 5 Manceau, E., Mezghani, M., Zabalza-Mezghani, I. and Rogero, F. (2001) Combination of Experimental Design and Joint Modeling Methods for Quantifying the Risk Associated with Deterministic and Stochastic Uncertainties - An Integrated Test Study, SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, number 71620.
- 6 Romero, C.E., Carter, J.N., Gringarten, A.C. and Zimmerman, R.W. (2000) A Modified Genetic Algorithm for Reservoir Characterisation, SPE International Oil and Gas Conference and Exhibition in China, Beijing, China, number 64765.
- 7 Williams, G.J.J., Mansfield, M., MacDonald, D.G. and Bush, M.D. (2004) Top-Down Reservoir Modelling, SPE Annual Technical Conference and Exhibition, Houston, Texas, number 89974.
- 8 Schulze-Riegert, R.W., Haase, O. and Nekrassov, A. (2003) Combined Global and Local Optimization Techniques Applied to History Matching, SPE Reservoir Simulation Symposium, Houston, Texas, number 79668.
- 9 Sen, M.K. et al. (1995) Stochastic Reservoir Modeling Using Simulated Annealing and Genetic Algorithms, SPE Annual Technical Conference and Exhibition, Washington, US, number 24754.
- 10 Subbey, S., Christie, M. and Sambridge, M. (2002) Uncertainty Reduction in Reservoir Modelling, *Contemp. Math.*, 295, 457-467.
- 11 Litvak, M., Christie, M., Johnson, D. and Sambridge, M. (February 2005) Uncertainty Estimation in Production Predictions Constrained by Production History and Time-Lapse Seismic in a GOM Oil Field, SPE Reservoir Simulation Symposium, Houston, Texas, number 93146.
- 12 Demyanov, V., Subbey, S. and Christie, M. (2004) Uncertainty Assessment in PUNQ-33 Neighbourhood Algorithm Framework for Geostatistical Modelling, 9th European Conference on the Mathematics of Oil Recovery, Cannes, France.
- 13 Castellini, A., Landa, J.L. and Kikani, J. (2004) Practical Methods for Uncertainty Assessment of Flow Predictions for Reservoirs with Significant History - A Field Case Study, 9th European Conference on the Mathematics of Oil Recovery, Cannes, France.
- 14 Mantica, S., Cominelli, A. and Mantica, G. (2001) Combining Global and Local Optimization Techniques for Automatic History Matching Production and Seismic Data, SPE Reservoir Simulation Symposium, Houston, Texas, number 66355.
- 15 Floris, F.J. et al. (2001) Methods for Quantifying the Uncertainty of Production Forecasts: A Comparative Study, *Int. J. Geosci.*, 7, S87.
- 16 Liu, N., Betancourt, S. and Oliver, D.S. (2001) Assessment of Uncertainty Assessment Methods, SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, number 71624.
- 17 Tavasoli, Z., Carter, J.N. and King, P.R. (2004) Errors in History Matching, *SPE J.*, 86883.
- 18 Ballester, P.J. and Carter, J.N. (2003) Real-Parameter Genetic Algorithms for Finding Multiple Optimal Solutions in Multimodal Optimization, GECCO-03: Proceedings of the Genetic and Evolutionary Computation Conference
- 19 Goldberg, D.E. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company Inc., Reading, Massachusetts.
- 20 Erbas, D. (2006) Sampling Strategies for Uncertainty Quantification in Oil Recovery Prediction, PhD. Thesis, Institute of Petroleum Engineering, Heriot Watt University, Edinburgh, UK.
- 21 Valenzuela-Rendon, M. and Uresti-Charre, E. (1997) A Non-Generational Genetic Algorithm for Multiobjective Optimization, Proceedings of the 7th International Conference on Genetic Algorithms
- 22 Sambridge, M. (1999) Geophysical Inversion with a Neighbourhood Algorithm - I. Searching a Parameter Space, *Geophys. J. Int.*, 138, 479.
- 23 Okabe, A., Boots, B. and Sugihara, K. (1992) *Spatial Tesselations - Concepts and Applications of Voronoi diagrams*, John Wiley & Sons, England.
- 24 Sambridge, M. (1999) Geophysical Inversion with a Neighbourhood Algorithm - II. Appraising the Ensemble, *Geophys. J. Int.*, 138, 727.
- 25 Carter, J., Ballester, P.J., Tavasoli, Z. and King, P.R. (2004) Our Calibrated Model has No Predictive Value: An Example from the Petroleum Industry, Proceedings of the SAMO Conference

Final manuscript received in October 2006

Copyright © 2007 Institut français du pétrole

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than IFP must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee: Request permission from Documentation, Institut français du pétrole, fax. +33 1 47 52 70 78, or revueogst@ifp.fr.