Oil Reservoir Production Forecasting with Uncertainty Estimation Using Genetic Algorithms

Harald H. Soleng
Norwegian Computing Center
P.O. Box 114 Blindern
N-0314 Oslo, Norway
Harald.Soleng@nr.no

Abstract: A genetic algorithm is applied to the problem of conditioning the petrophysical rock properties of a reservoir model on historic production data. This is a difficult optimization problem where each evaluation of the objective function implies a flow simulation of the whole reservoir. Due to the high computing cost of this function, it is imperative to make use of an efficient optimization method to find a near optimal solution using as few iterations as possible. In this study we have applied a genetic algorithm to this problem. Ten independent runs are used to give a prediction with an uncertainty estimate for the total future oil production using two different production strategies.

1 Oil production history matching

In order to be able to forecast the oil production using different production strategies, one needs realistic geological models of the oil reservoir. The geological models are formulated on grids with thousands or millions of grid cells. Each grid cell has several physical variables, and hence, the number of unknown parametres in a realistic reservoir characterization is formidable.

The *history matching* problem is the problem of finding geological models which are consistent with both static data—such as permeabilities and porosities as measured in wellbore plugs—and with dynamic data such as production rates, bottom hole pressures, and gas oil ratios throughout the production history of the field.

In general, the history matching problem is a non-unique inverse problem; several combinations of parametre values representing the geology could give the same production performance. In a full scale heterogeneous reservoir, the number of unknown parametres is often as high as several millions while the number of observables is much smaller. The task is to find a set of parametres so that the difference between the results of flow simulations and the true production history is as small as possible. This is a hard optimization problem. Since the computational cost of differentiation within a flow simulator is very high, we have chosen to experiment with a *genetic algorithm* (GA) [1, 2] as an optimization tool.

For any realistic case one expect that there are many local optima in the history matching problem. Since production parametres to a large extent are determined by the geology near the wells, regions far from wells are not very well determined by this kind of inverse modelling. This induces large uncertainties in the predicted production, especially if new wells are drilled. In order to estimate this prediction uncertainty, we have generated a population of history matched models where each individual is generated by an independent GA optimization. In this way it is hoped that we span a significant part of the parametre space compatible with the known production history and thus, that we can estimate the prediction uncertainty.

2 The PUNQ S3 case

The method is tested on a synthetic oil field prepared as part of the PUNQ (production forecasting with uncertainty quantification) project sponsored by the European Community. In the PUNQ project ten partners from industry, research institutes and universities are collaborating on research on uncertainty quantification methods for oil production forecasting [3]. The 'historic data' of the case studied in this paper were generated using the Eclipse oil simulator. Gaussian noise was added to both the historic data and the well observations before the data sets with uncertainties were presented to the partners. At the time the history matching was carried out, the true reservoir was not known to the partners. In the present work we used the More simulator for history matching.

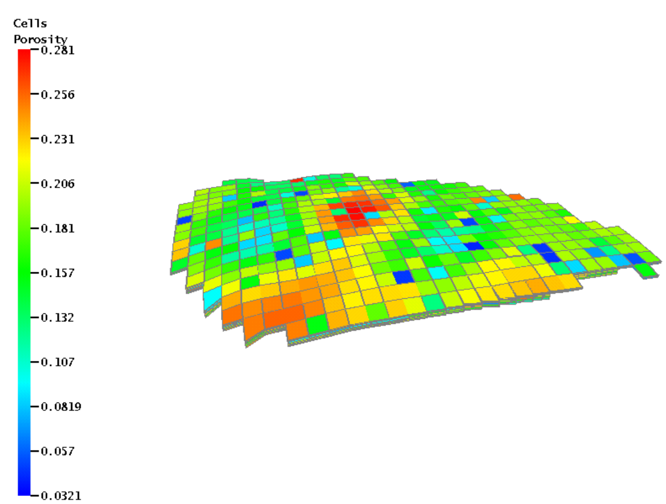

Figure 1: A reservoir model optimized with GA.

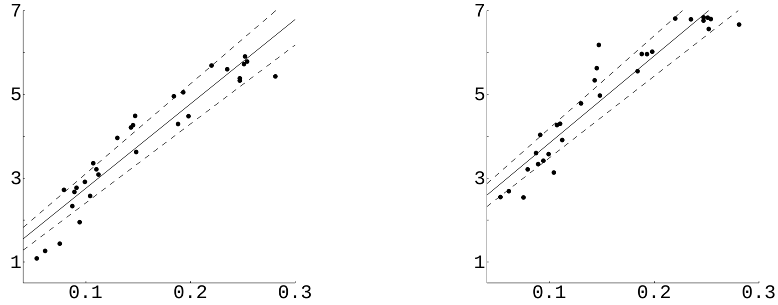

Figure 2: Wellbore plug data: logarithm of vertical and horizontal permeability $\log(K_\text{v})$ and $\log(K_\text{h})$ versus porosity ϕ, respectively.

The production history for the oil field was known for a period of 8 years. The field has 6 wells. In the present case study we work on a relatively small block-centered Cartesian grid with dimensions $19 \times 28 \times 5$. The geological parametres are the horizontal and vertical permeabilities, K_h and K_v, and the porosities, ϕ. About one third of the grid blocks are inactive. Figure 1 shows the porosity field of an optimized model. The horizontal size of the reservoir is much larger than its vertical size, and hence, the grid blocks are very thin slivers. Consequently, the third dimension is hardly resolved in Fig. 1.

There are six wells with well bore plug observations for all five layers. The well data indicates a 95% correlation between horizontal and vertical permeabilities and the porosity, cf. Fig. 2. For this reason we used three highly correlated Gaussian random fields conditioned on well observations to intialize the population before starting the optimization.

The pressure of the field was maintained by a number of aquifers. It was therefore not necessary to use injection wells. In the test case, an analytic aquifer model of the Carter–Tracy type was specified as part of the Eclipse model file. Since there is no such aquifer model in the current version of the More simulator, we had to construct one by using sets of isolated inactive blocks to represent huge water sources and non-neighbour connections to model water flow into the grid blocks specified by the Eclipse aquifer model.

After 8 years of production, two different recovery strategies should be considered. The first alternative was to continue production for another 8.5 years using the original 6 wells. The second alternative was to add 5 new wells for the next 8.5 years. The aim was to give forecasts with uncertainty estimates for the total oil production for each of the production strategies.

3 The genetic algorithm

The objective of history matching is to find a geological model giving simulation results as close as possible to the real production history. The main idea of evolution programming is to search for the optimum from a population of possible solutions. New states are proposed by recombination of *genetic* material in the population. By application of Darwin's *survival of the fittest* the population increases in fitness making efficient use of gradient information implicitly encoded in the population. The genetic algorithm is based on a genetic representation of possible solutions, a mating operator for producing offspring, and a mutation rule.

3.1 History mismatch and fitness

In history matching we try to minimize the deviation from the true production history. Thus the most *fit* models are those with a minimal deviation from the historic data. Hence, let us quantify this deviation by a weighted sum of squared deviations

$$Q(r) = \frac{1}{n_s} \sum_{i=1}^{n_s} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{w_{ij}\,(p_{ij}(r) - o_{ij})^2}{\sigma_{ij}^2} \tag{1}$$

where r is the reservoir characteristics vector, n_s is the number of time series, n_i is number of elements in time series i, p_{ij} are the different simulator output results for the bottom hole pressure, the gas/oil ratio, and the water cut, o_{ij} are the corresponding observations, w_{ij} are the weights, and σ_{ij}^2 are the variances. The weights and variances were specified in the PUNQ S3 case.

Then we define the *fitness* of a particular set of reservoir characteristics r as

$$f(r) = \frac{1}{1 + \sqrt{Q(r)}} \tag{2}$$

where the history mismatch $Q(r)$ is given in Eq. (1). If for some reason, the flow simulator crashed during the simulation, the fitness was set equal zero. Thus the fitness is normalized to the interval $[0, 1]$.

3.2 Genetic representation

In order to preserve the high degree of correlation between the petrophysical parametres as seen in the wells (cf. Fig. 2), the reservoir genome was coded as an array of active grid blocks. Each grid block carries a set of three petrophysical block values, namely the horizontal permeability, the vertical permeability, and the porosity. Thus in this model each gene codes the complete set of petrophysical data for a reservoir block.

3.2.1 Genetic operators

The initial population was created by using highly correlated transformed Gaussian random fields (conditioned on the well observations) for the geological parametres. The transformation maps the Gaussian random field to porosity values between 0 and 0.3. For the permabilities we used exponential transforms of the Gaussian random fields to get only positive values.

The *crossing operator* is a simple one point crossover on a one-dimensional list of blocks. Such a simple cut-and-paste operator in state space could lead to some realizations with very high density contrasts. One could avoid this by implementing a smoother interpolation between genetic material from the two parents, but we have chosen to let the objective function take care of it; realizations that make the flow simulator crash are given zero fitness. However, using a three-dimensional crossover method with smooth interpolations would probably improve the method. The *mutation operator* is a simple swap operator interchanging pairs of grid blocks. The probability of swapping the grid blocks observed in the wells is zero. Thus, the conditioning on well observations is preserved by the algorithm.

3.2.2 GA algorithm and parametres

We used a steady state genetic algorithm with overlapping populations [4, p. 32]. We used a population size of $N = 50$, a mutation probability of 1%, a crossover probability of 90%, and a replacement percentage of 25% in each generation.

Thus, in each generation 12 children are born, and 12 individuals pass away. We let 90% of the children be formed by crossover followed by mutation. The remaining 10% were created as mutated clones of existing individuals. The mutation probability for each gene in a chromosome is 1%. With around 1700 genes in each chromosome, the genome of a typical child has undergone 17 mutations.

The evolution was stopped when the mean fitness was within 98% of the fitness of the best individual. Children replaced the worst individuals of the population. For mating, a roulette wheel selector was used. Let f_i be the fitness of individual number i. If the total fitness of the population is

$$F \equiv \sum_{i=1}^{N} f_i, \tag{3}$$

then the probability for the individual j to be chosen for mating is

$$p_j = \frac{f_j}{F}. \tag{4}$$

A total of ten independent runs were made in order to be able to give an estimate of the forecast uncertainty.

4 Results

The objective of this study was not to find a single solution matching the observed history. Rather it was to give a good forecast of the total oil production and to estimate the uncertainty of this forecast. Thus, it was necessary to produce a population of history matched reservoir models where the individuals may represent different local optima of the fitness landscape of history matching. This was obtained by picking the best individual of the last generation of ten independent GA runs. In this section we report on the results for history matching and forecasts using this set of optimized reservoir realizations. In the optimized population, the mean value of the weighted squared deviation Q, as defined in Eq. (1), was 1.25 (ranging from 1.10–1.46). The mean of means of the 10×50 realizations drawn from the prior was 3.23. The minimal and maximal values of the history mismatch of the unoptimized realizations were 1.58 and 5.30, respectively. Thus, a significant improvement of Q is observed in the optimized population compared to the prior distribution.

4.1 History match

Since the stopping criterium of the GA was that the mean fitness is with 98% of the best, the number of flow simulation runs before convergence was reached varied among the GA runs. The mean number of reservoir simulation runs in the genetic algorithm was 805 spanning the range of 566–1154. The two runs with lowest and highest final value for $Q(r)$ are

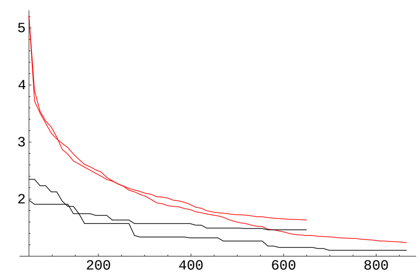

Figure 3: Evolution of $Q(r)$ of the best (black) and worst (red) individuals of two independent GA runs as functions of number of flow simulations.

shown in Fig. 3. Observe that the run with the highest final history mismatch stopped early. There is a general tendency of early stoppers being less well adapted than those running longer. This may be taken as sign of premature convergence of the GA.

Details about the history match for the optimized realizations are shown in Figs. 4–6.

Figure 4 shows the simulated bottom hole pressure versus time for the ten optimized realizations. The figure shows the historic data with $1\,\sigma$ error bars. The initial step-like behaviour is due to well tests carried out in the first part of the simulation. The long plateau represents a shut-down period of 1461 days. In this period there is a slow increase in the pressure. In the production profiles of the optimized reservoirs this increase is slightly smaller than in the historic data. Thus, the simulated bottom hole pressures (Fig. 4) at the end of the shut-down period fall outside the $1\,\sigma$ error bar at this

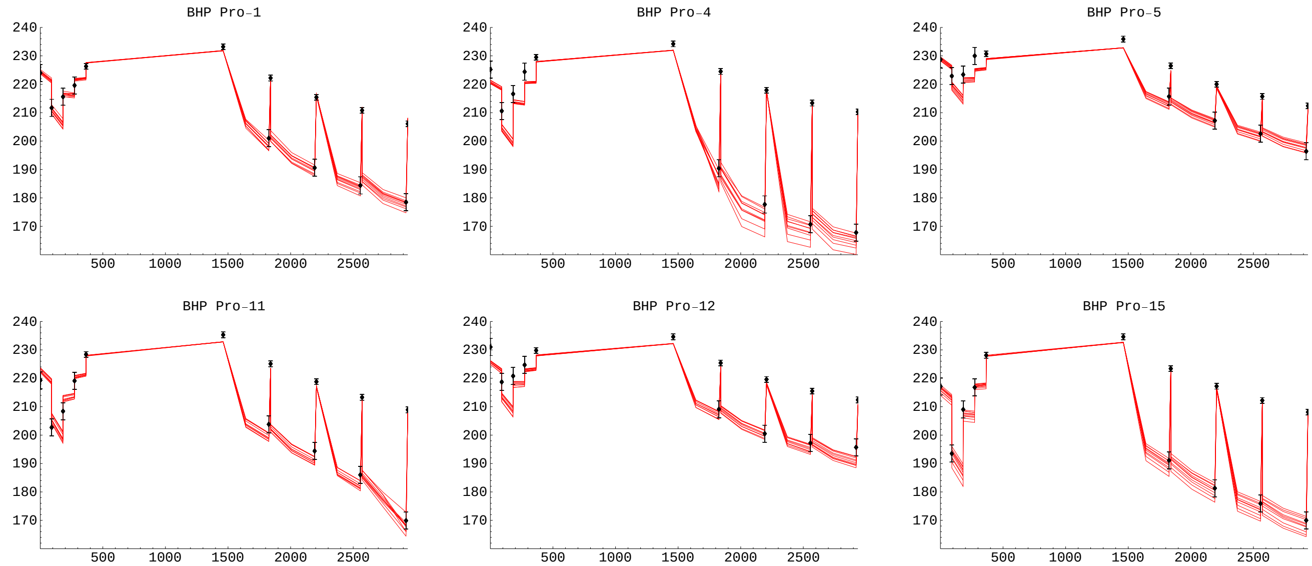

Figure 4: Bottom hole pressure versus days for six production wells. The spikes are due to periodic testing and maintainance. Shutting the producing wells causes a rapid increase in their pressures followed by a large drop when production is restarted.

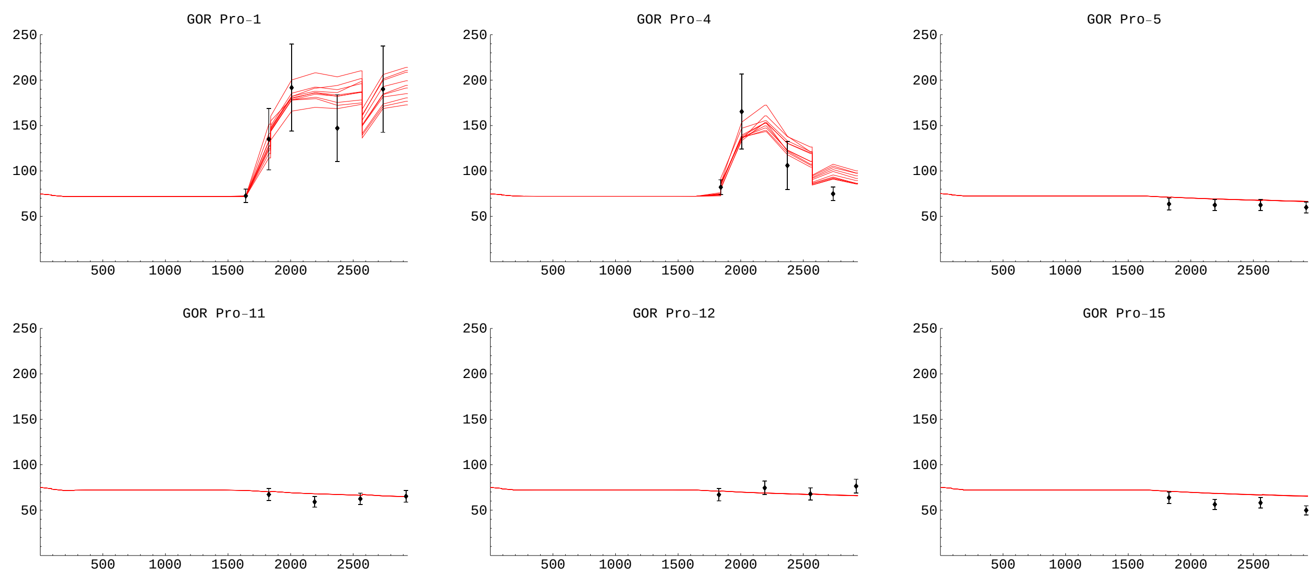

Figure 5: Gas oil ratios versus days for six production wells.

data point. The periodic spikes in the plots are due to well tests; at certain times the wells are shut in for tests and maintainance. During shut-in the pressure increases, followed by a quick drop when the production is resumed.

Figure 5 shows the simulated gas oil ratios versus time together with observed values with 1 σ error bars. The fluctuations in the gas oil ratio of wells Pro-1 and Pro-4 are hard to match. The simulation results for the gas oil ratio (Fig. 5) are outside the 1 σ error bar for several data points.

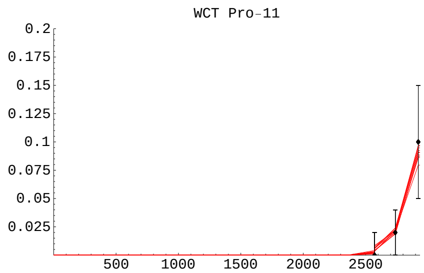

Figure 6: Water oil ratio versus days for the only production well with a nonvanishing water production.

In all wells except one, the water content is almost zero throughout the whole production history. This is the case both in the 'true' history and the simulation data. The sudden appearance of water in well Pro-11 is well matched by the simulations for all ten realizations (cf. Figure 6).

4.2 Forecasts

One of the objectives of history matching is to be able to predict the production of a field. After 16.5 years of production with the 6 original wells, we found a mean total production of 3.81×10^6 m^3 oil ($\sigma = 0.04 \times 10^6$ m^3). To study the

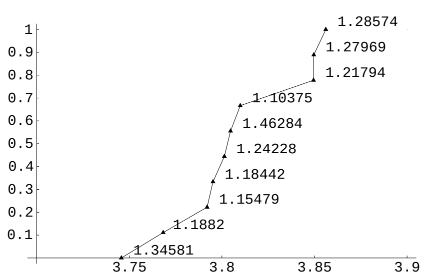

Figure 7: Estimated probability distribution curves for total oil production in units of 10^6 m^3 with the original wells.

effect of incremental drilling, five new production wells were added. Their production started at the end of the "history matching period". With these addional wells, the forecasted total production is 4.69×10^6 m^3 oil ($\sigma = 0.09 \times 10^6$ m^3). The estimated probability distribution curves for the total oil production and the increase in the oil production with 5 incremental wells are depicted in Figs. 7 and 8, respectively. The numeric labels on the data points are the corresponding $Q(r)$ values.

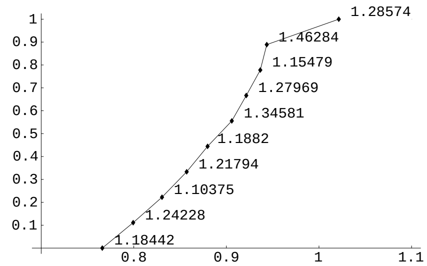

Figure 8: Estimated probability distribution curve for increased oil production with five additional wells in units of 10^6 m^3.

The predictions of Figs. 7 and 8 should be compared with the 'true total production' obtained by running More on the 'true reservoir'. With the original wells the 'true' production was 3.85×10^6 m^3. This agrees well with the prediction of $3.81 \pm 0.04 \times 10^6$ m^3. With five infill wells, the true production increased to 5.15×10^6 m^3 compared to a prediction of $4.69 \pm 0.09 \times 10^6$ m^3. In this case the predicted range falls far below the true answer.

4.3 History mismatch and production

One could ask if there is a correlation between total oil production and values of $Q(r)$. As seen in Figs. 9 and 10 this seems not to be the case. Hence, the premature convergence of the genetic algorithm has not biased the predictions.

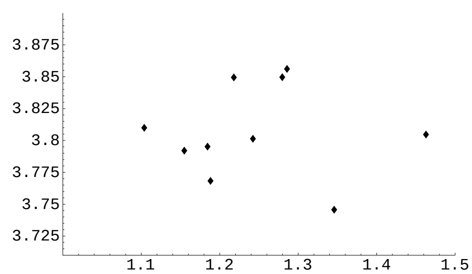

Figure 9: The total oil production without incremental wells as a function of $Q(r)$.

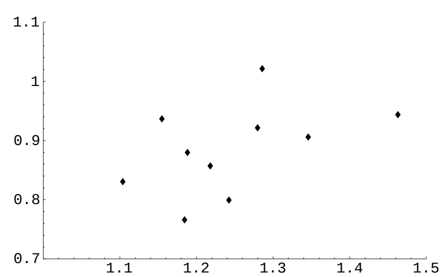

Figure 10: The increment of the total oil production with additional wells as function of $Q(r)$.

Figure 11 shows a cross plot of total oil production with and without incremental wells. As one should expect, there

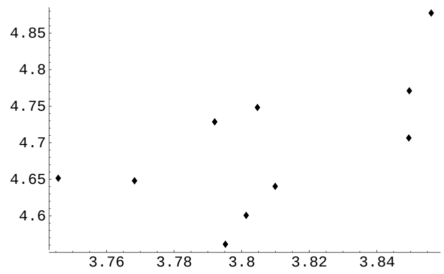

Figure 11: Total oil production with incremental wells versus total oil production with the original wells.

seems to be a weak linear correlation between total production with the two production strategies. A good reservoir with 6 wells is still good with 11.

4.4 Discussion

The true value for the case that no new wells are drilled is within the 1 σ uncertainty of the forecast. However with five new production wells in addition to the original six, the forecast is significantly lower than the true production. This mismatch can be explained from the following arguments:

a) Since the 'true reservoir' was non-Gaussian, initialization using Gaussian fields may be too restrictive, because of the finite correlation length of the Gaussian fields.

b) The true production history was generated with a different flow simulator than the one used for history matching.

c) The regions far from wells with known production history are not much constrained by inverse modelling, and thus, the production uncertainty of new wells in such regions is much higher than the production uncertainty of the history matched wells.

Argument (a) is strengthened by a study using an adaptive Metropolis–Hastings algorithm confined to Gaussian fields [5] where the predicted uncertainty is even smaller than in the present case. Thus, the use of Gaussian fields focus the search on a too small part of the parametre space and forces the geological models into a small subspace which may not cover that of the true reservoir. We now know that the synthetic reservoir is non-Gaussian.

(b) Using a different simulator for matching than the one used to generate the 'true' history may also produce a bias. For regions near wells with a known production history, the history matching will compensate for the simualtor bias, but not in other regions.

The point raised in argument (c) comes into play when new wells are drilled. Genes representing petrophysical data of grid blocks far from wells with known production history are subject to a very weak selective pressure. Consequently, genetic operators lead to mediocre permeability properties in such regions. If high-permeable grid cells are required near the existing wells, then the grid cell swapping of the mutation operator would tend to put low-permeable cells in less important regions. Furthermore, the grid cell swapping produces white noise in the permablity fields. Due to the lack of correlation in white noise, such fields have lower long range permeability than smoother Gaussian fields. Thus, the predicted production of new wells are biased towards mediocre or low yield.

Although the biasing of predictions can be reduced, it can never vanish. The starting point for optimization must always be a class of geological models. The advantage of narrowing the search by a restrictive model has to be traded against the risk of missing important properties of the true geology. In addition, the flow simulator has lots of simplifying assumptions. As a consequence, it will never reproduce the true flow of a real reservoir even if we had exact knowledge about its petrophysical rock properties. However, by making sure that our history matching methods respect the prior geological knowledge about the reservoir, *i.e.*, by doing better than producing white noise in regions with weak selective pressure, the predictability of the yield of new wells should improve.

5 Conclusions

A steady state genetic algorithm has been applied to a history matching problem in oil reservoir exploration. The method has proven to be reasonably fast in obtaining near optimal solutions, *i.e.*, geological models giving flow simulation results close to the observed production history. To be able to quantify the uncertainty of the predicted oil production, we created a population of history matched realizations using the best individuals of ten independent GA runs. Using these ten geological models as input to the flow simulator, production forecasts were made for two different production strategies. These forecasts are similar enough to indicate that all the end realizations of the reservoir are geologically similar. Albeit this method is *not* a statistically correct method for drawing from the posterior, it gives a valuable estimate of the forecasting uncertainty.

This work has shown that genetic algorithms are promising in history matching for real petroleum reservoirs. In this connection the inherent parallelism of the algorithm—and the huge potential for speeding up the calculations on parallel machines—will be essential when attacking larger models. It is worth mentioning that very powerful parallel machines can be built as Linux clusters using off-the-shelf PC hardware which to a large extent already is part of the inventory of most oil companies.

However, there are still a few problems with the method as implemented here. Most of these are small and solvable; we have, for example, not experimented much with parametres of the genetic algorithm itself. For instance, it is believed that premature convergence can be avoided by using several separate populations with migration instead of the single steady state population used in this paper. Also, as mentioned earlier, the disruptive effect of crossover can be reduced by using a 3D crossover operator. It can be further refined by interpolating between genetic material from the two parents near the crossover surfaces. This would reduce the risk for ending up with superficial high contrast surfaces in the children.

Other problems are more fundamental and harder to solve. For example, we need to find a genetic algorithm respecting prior geological knowledge about the reservoir such as the correlation structure of the petrophysical fields. In this study we used a population of 50 transformed Gaussian fields as the starting point for our genetic algorithm. These fields had certain trends representing prior information about the geology. However, by breaking up the correlation structure the genetic operators (especially the block swapping mutation operator) tend to bias the production from new wells towards low yield. Similar and more severe problems occur if one tries to combine genetic algorithms with more advanced reservoir characterization methods making use of more detailed geological knowledge about the reservoir. For example, a so-called fluvial reservoir [6] consists of a network of channels in a low permeable background (a fluvial reservoir is the result of river deposits built up over millions of years by the changing position of a river-bed). Such reservoirs can be represented by object models. In order to apply a genetic algorithm to reservoir models of this type, one needs to find suitable parametrizations and genetic representations of object models so that one can define meaningful crossing and mutation operators without loosing the prior geological knowledge inherent in these models.

Ideally, one would like to find not only geological realizations that match the history well, but their statistically correct distribution. To meet this demand, the genetic algorithm could be combined with a Markov chain Monte Carlo sampler [7, 5].

Acknowledgements

The author would like to thank Anne Randi Syversveen and Lars Holden for comments and suggestions, Håkon Tjelmeland for permission to use his code for Gaussian fields, and Smedvig Technologies for support and for donating a lisence for the Modular Oil Reservoir Evaluation programme (More). The author especially wants to thank David Ponting of Smedvig Technologies for his kind assistance in setting up the model files for More. The software for this work used the GAlib genetic algorithm package, written by Matthew Wall at the Massachusetts Institute of Technology. The author is grateful for financial support from the European Community through the Joule-III non-nuclear energy programme.

Bibliography

[1] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, Massachusetts: Addison-Wesley, 1989.

[2] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, Berlin: Springer, 3 ed., 1996.

[3] F. J. T. Floris, M. D. Bush, M. Cuypers, F. Roggero, and A. R. Syversveen, “Comparison of production forecast uncertainty quantification methods: an integrated study,” *work in progress*, 1999.

[4] M. Wall, *GAlib: A C++ Library of Genetic Algorithm Components*. MIT, Mechanical Engineering Department, Massachusetts Institute of Technology, Aug. 1996. http://lancet.mit.edu/ga.

[5] L. Holden, H. H. Soleng, and A. R. Syversveen, “History matching with uncertainty quantification,” in *Proceedings of the 1999 Annual Conference of the International Association for Mathematical Geology*, 1999. Accepted for publication.

[6] L. Holden, R. Hauge, Ø. Skare, and A. Skorstad, “Modeling of fluivial reservoir with object models,” *Mathematical Geology*, vol. 30, pp. 473–496, July 1998.

[7] C. C. Holmes and B. K. Mallick, “Parallel Markov chain Monte Carlo sampling: an evolutionary approach,” tech. rep., Imperial College, London, 1998.