

Our Calibrated Model has No Predictive Value: An Example from the Petroleum Industry

J.N. Carter^a, P.J. Ballester^a, Z. Tavassoli^a and P.R. King^a

^a Department of Earth Sciences and Engineering,
Imperial College, London, SW7 2AZ, United Kingdom.
Email: j.n.carter@imperial.ac.uk.

Abstract: It is often assumed that once a model has been calibrated to measurements then it will have some level of predictive capability, although this may be limited. If the model does not have predictive capability then the assumption is that the model needs to be improved in some way.

Using an example from the petroleum industry, we show that cases can exist where calibrated models have no predictive capability. This occurs even when there is no modelling error present. It is also shown that the introduction of a small modelling error can make it impossible to obtain any models with useful predictive capability.

We have been unable to find ways of identifying which calibrated models will have some predictive capacity and those which will not.

Keywords: Prediction, Calibration, Uncertainty, Petroleum

1. INTRODUCTION

In many studies involving numeric models of complex real world situations, for example petroleum reservoirs and climate modelling, it is implicitly assumed that if the model has been carefully calibrated to reproduce previously observed behaviour, then the model will have some predictive capacity. It is recognised that predictability may only be achievable for a finite period of time, and that any prediction will be uncertain to some extent.

Two types of error are considered in most calibration exercises: measurement error and model error. Measurement errors are fixed at the time the measurement was made, they generally have well defined statistics and can be handled appropriately. Model errors are due to approximations, such as a loss of spatial, or temporal, resolution, and the non-inclusion of all of the relevant physics. The assumption that is normally made is that if the model errors are sufficiently unimportant, so that when the model has been calibrated to measurement data, then we have some level of acceptable predictability. If the model does not have predictability, then the model errors are assumed to be too large and we need to use a “better” model. Where “better” probably means improved resolution, spatial or temporal, and/or the inclusion of more physics.

In this paper we present the results of a study, for a petroleum reservoir, where a well calibrated model has no predictive value. Even though the calibration and truth models had identical physics and identical spatial and temporal resolution. A second study shows

Correspondence to J.N. Carter

that where there are slight differences in the physics between the calibration and truth models, then the problems encountered are even worse.

In the next section the experimental set-up is described, this is followed by the results for cases with/without modelling errors. Finally we draw some conclusions from our observations.

2. EXPERIMENTAL SET-UP

In this section we describe our three parameter reservoir model and our methodology for calibrating the model against the available measurements.

2.1. Model Description

Our model is a cross-section of a simple layered reservoir, with a single vertical fault midway between an injector producer pair, as shown in figure 1. The model that we calibrate has three parameters: the vertical displacement (throw) of the fault; the permeability of the poor quality sand; and the permeability of the good quality sand. The geological layers are assumed to be homogeneous (ie they have constant physical properties). The “truth” case, which is used to generate the measurements for the calibration, is a variant of the calibration model, but with fixed parameter values. In the case of no model error, then the “truth” case is a member of the set of all possible calibration models. The size and type of model error is chosen by how a specific calibration model is perturbed to obtain the truth case. In the work presented in this paper, the model error is obtained by introducing small variations into the spatial properties of the geological layers. The permeability and porosity in each grid block are randomly perturbed. The maximum variations that are allowed is $\pm 1\%$ of the unperturbed mean values. These perturbations are much lower than would be expected for a real world rock that had been classified as homogeneous. A more extensive description of the model can be found a paper that deals with estimating model errors[4].

2.2. Calibration Methodology

Our procedure to produce a calibrated model is as follows:

1. Choose “truth” values for the three model parameters;
2. Select the level of measurement and model error to be used;
3. From the truth case produce the measurements required for the calibration process (three years of monthly data);
4. Calibrate the model against the measurements;
5. Predict the behaviour for years 4-10.

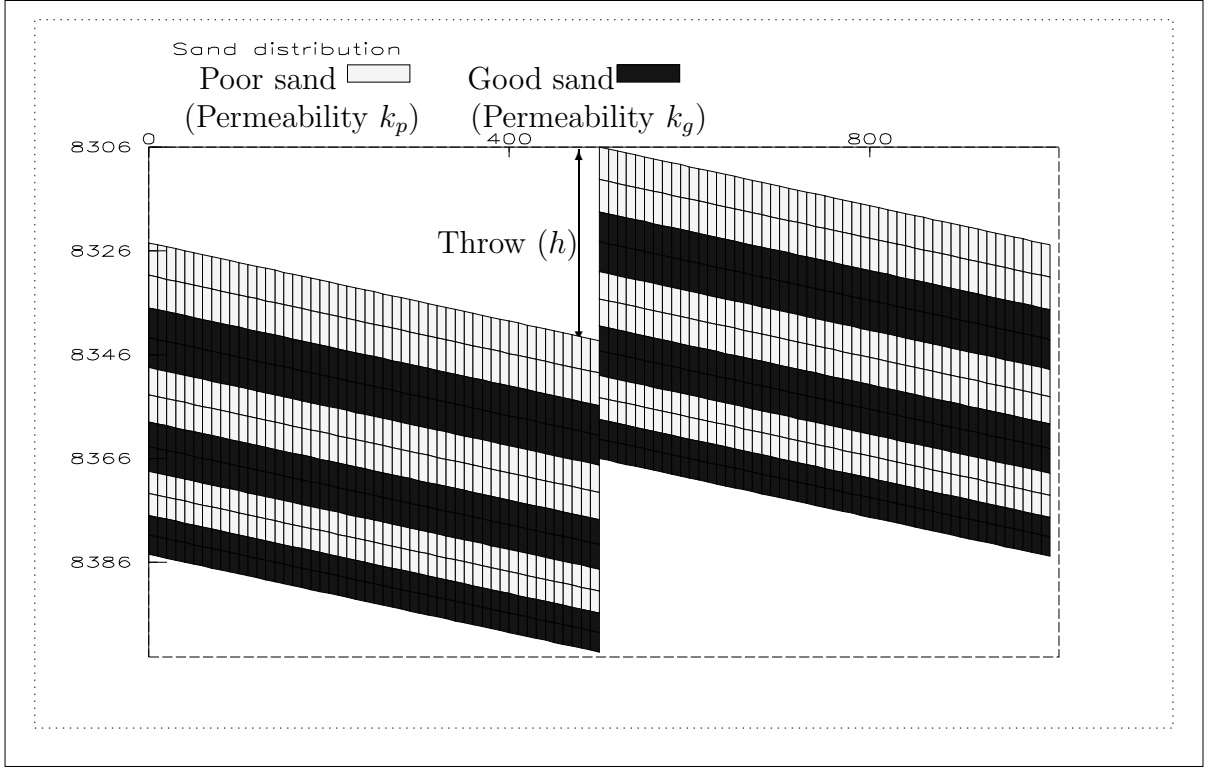


Figure 1. Reservoir model showing the fault throw and the geological, and simulation, layers.

We have considered the truth case: $h = 10.4$, $k_p = 1.31$ and $k_g = 131.7$ with and without model error. No measurement error was added, but we assumed Gaussian noise with a 1% standard deviation when calculating the likelihood that a proposed calibration matches the truth.

In order to quantify the degree of the model calibration against measurements, we define first an objective function for the calibration period, Δ_m , as follows

$$\Delta_m = \frac{1}{36} \sum_{j=1}^{36} \sum_{k=1}^3 \frac{|sim(j, k) - obj(j, k)|}{2\sigma_{jk}} \quad (1)$$

where $sim(j, k)$ is the simulated response for production series k of the model at time j , $obj(j, k)$ is the corresponding true value and σ_{jk} , an estimation of what would be the associated measurement error. We consider three production series: Oil Production Rate, Water Production Rate (or Water Cut) and Water Injection Rate.

Likewise, the objective function for the prediction period, Δ_f , is

$$\Delta_f = \frac{1}{7} \sum_{j=37}^{43} \sum_{k=1}^3 \frac{|sim(j, k) - obj(j, k)|}{2\sigma_{jk}} \quad (2)$$

The ranges that the model parameters were allowed to take are: $h \in (0, 60)$, $k_g \in (100, 200)$ and $k_p \in (0, 50)$.

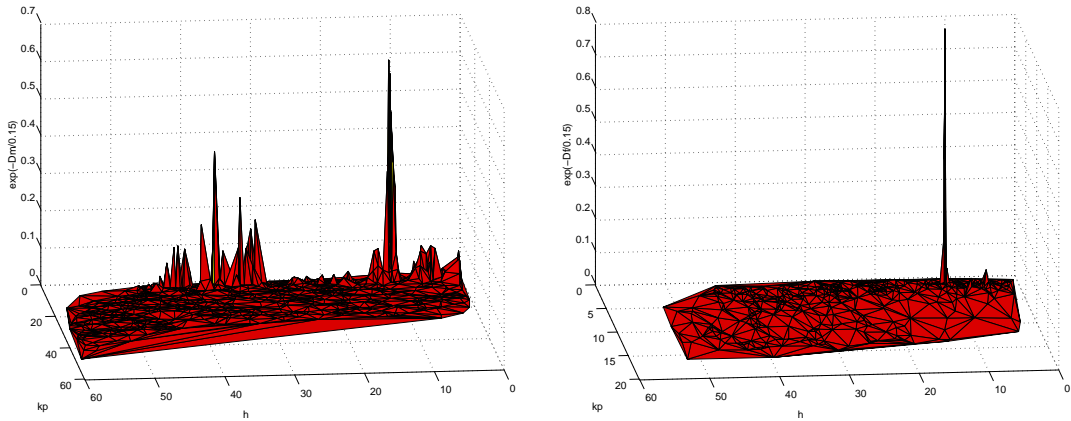


Figure 2. Calibrations of the model (with no modelling error) to a) history period, b) prediction period.

2.3. Genetic Algorithm

Our chosen search method is a Steady-state Real-parameter Genetic Algorithm. This is used because we need to search for multiple good optima within a parameter space that seems to contain very many local optima. It is a development of a previously published study [1] and has been developed to solve the type of problem described in this paper.

In brief the details are: a steady-state population of 50 individuals is used, parents are selected randomly (without reference to their fitness), crossover is performed using vSBX[1, 3], and culling is carried out using a form of tournament selection involving 10 individuals, a total of 7000 individuals are generated.

3. RESULTS

In this section we present the results of two studies: the first is with no modelling error present; the second has a low level of modelling error.

3.1. Calibration with No Modelling Error

Figure 2a shows the result of calibrating the model against the data for the first 36 months. The truth model has exactly the same physics and structure as the calibration models, and the truth model is a member of the set of possible calibration models.

The very large spike, with $h \approx 10$, corresponds to the truth case. We can also see notable local optima with $0 < h < 8$, $30 < h < 38$ and $40 < h < 45$. The global optimum has a small basin of attraction around it and has proved difficult to identify in previous work[2], the easiest optimum to find has been the one with $30 < h < 38$. The rather noisy structure of the objective surface is largely an artifact of the way that k_g is sampled. Any point with an acceptable objective value is plotted no matter what value of k_g was used. This means that it is possible for two points to have identical values for h and k_p but different values of the objective function. Hence a vertical line would be plotted. Figure 3 shows a contour plot, centred on $h = 5.0$ and $k_p = 1.65$, of the objective

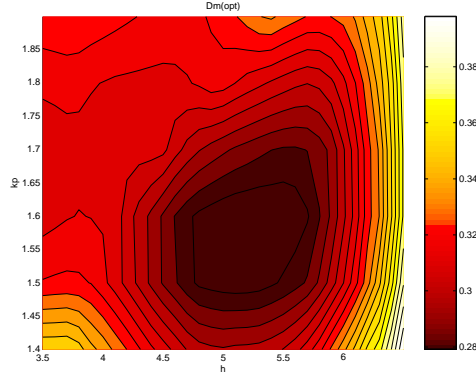


Figure 3. Surface plot for Δ_m , where k_g has been optimised so as to minimise Δ_m , $h \in (3.5, 6.5)$ and $k_p \in (1.40, 1.90)$

function Δ_m . The figure was generated by conducting a grid search on a fine grid. At each point on the grid, k_g was optimised, this results in a much smoother representation of the objective.

Figure 2b shows the result of calibrating the model to the prediction period. The only substantial point found corresponds to the truth model. All of the other local optima that can be seen in figure 2a are unable to match the observations during the prediction period. We conclude that for this model you can only obtain a good prediction from the truth case, and that good matches from the history matching phase have no predictive value.

3.2. Calibration with Modelling Error

The result of matching the calibration model to data generated by a truth case that includes modelling errors is shown in figure 4a. Superficially the figure is similar to figure 2a. The important difference is that the global optima now occurs for $h \approx 32$. This is within the largest basin of attraction and is usually found by most search algorithms. The optima associated with the “true” parameter values is of much lower quality.

If we now look at the calibration to the prediction period, figure 4b, we see that the global optimum for the history matching period has no predictive value. None of the models that have some predictive value correspond to the truth case (the spike at $h \approx 10$ has the wrong values for k_p and k_g). The objective values obtained are low compared to those in figure 2b.

4. CONCLUSIONS

In this paper we have examined, for a particular case, our ability to calibrate a model and then to make accurate predictions. This has been carried out for cases with and without modelling errors, but no measurement error.

From these studies we make the following observations:

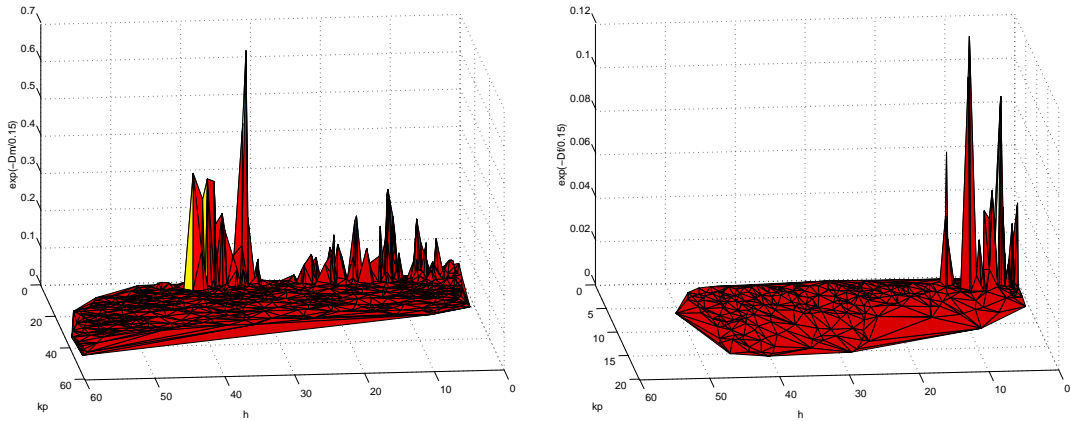


Figure 4. Calibrations of the model (with modelling error) to a) history period, b) prediction period.

- The basin of attraction around a global optimum may be sufficiently small that search algorithms may not find them. The basins of attraction associated with other local optima may be much larger and hence easier to find.
- When there is no modelling error present, some of the non-global optima may be of quite good quality. However only the global optimum is able to make an accurate prediction.
- When small amounts of modelling error are present, then the global optimum is no longer associated with the truth. The local optimum that has parameter values of the truth case is not of significant quality and could easily be disregarded.
- None of the models tested in the presence of modelling errors have valuable predictive power. In particular the global optima from the history matching period was unable to provide an accurate prediction.

In summary: in the absence of model errors, and with very low measurement errors, it is possible to obtain calibrated models that do not have any predictive capability; such models may be significantly easier to identify than the correct model; we are unable to differentiate between calibrated models with or without predictive capabilities; the introduction of even small model errors may make it impossible to obtain a calibrated model with predictive value.

In this analysis there is nothing that seems to be unique to this model. In particular there is the issue of data availability, adding more measurements does not appear to offer a guaranty of avoiding this dilemma. If the observations made with this model are not unique to the model, and we have no reason to believe that the model is unique, then this presents a potentially serious obstacle to the use of models of this type for prediction.

Our concern is that if we cannot successfully calibrate and make predictions with a model as simple as this, where does this leave us when are models are more complex, have substantive modelling errors, and we have poor quality measurement data.

REFERENCES

1. P.J. Ballester and J.N. Carter, Real-parameter Genetic Algorithms for Finding Multiple Optimal Solutions in Multi-modal *Optimisation, Genetic and Evolutionary Computation Conference 2003*:706, Lecture Notes in Computer Science 2723, Pub Springer-Verlag, Heidelberg, 2003.
2. M.D. Bush and J.N. Carter, Applications of a Modified Genetic Algorithm to Parameter Estimation in the Petroleum Industry, *Intelligent Engineering Systems Through Artificial Neural Networks* 6:397, 1996.
3. J.N. Carter, Introduction to using Genetic Algorithms, in Soft Computing and Intelligent Data Analysis in Oil Exploration, Eds M. Nikravesh, F. Aminzadeh and L.A. Zadeh, *Developments in Petroleum Science* 51, Elsevier Science, 2003.
4. J.N. Carter, Using Bayesian Statistics to Capture the Effects of Modelling Errors in Inverse Problems, accepted for publication in *Mathematical Geology*, expected to be published in 2004.