Frequentist History Matching with Interval Predictor Models

Jonathan Sadeghi[a], Marco De Angelis[a], Edoardo Patelli[a,*]

[a]*Institute for Risk and Uncertainty, Chadwick Building, University of Liverpool, Peach Street, Liverpool L69 7ZF, United Kingdom*

Abstract

In this paper a novel approach is presented for history matching models without making assumptions about the measurement error. Interval Predictor Models are used to robustly model the observed data and hence a novel figure of merit is proposed to quantify the quality of matches in a frequentist probabilistic framework. The proposed method yields bounds on the p-values from frequentist inference. The method is first applied to a simple example and then to a realistic case study (the Imperial College Fault Model) in order to evaluate its applicability and efficacy. When there is no modelling error the method identifies a feasible region for the matched parameters, which for our test case contained the truth case. When attempting to match one model to data from a different model, a region close to the truth case was identified. The effect of increasing the number of data points on the history matching is also discussed.

Keywords: Interval Predictor Model, History Matching, Surrogate Model, Inverse Problem, Imprecise Probability, Frequentist Inference

1. Introduction

History Matching is a method of calibrating a model, with the aim of inferring unknown parameters of the model by matching real world observations to its output. We believe the model is reasonably physically accurate, but some input parameters of the model are unknown. In many circumstances we may not know the error in the observed data and we are forced to make assumptions which may not be justified regarding the distribution of noise in the data [1]. These assumptions may cause the derived values for the model parameters to be biased or incorrect, and hence our predictions will have incorrect uncertainty bounds. In addition, our data may be limited in the sense that we do not have

*Corresponding author
Email address: edoardo.patelli@liverpool.ac.uk (Edoardo Patelli)

enough data to uniquely match the model and hence, there may be many possible matches. Therefore, we may be unable to make unique predictions regarding future behaviour of the system [2].

In order to quantify the noise in our data we will create a metamodel. Metamodels (also known as surrogate models or emulators) are approximate, black box models which can be constructed by fitting a model to training data or simulations. Metamodels are of use when we have a model which is computationally expensive and it is therefore infeasible to run the many simulations required by the Monte Carlo method, for example. An approximate model is needed to perform many simulations quickly. In this context they are models of models. Metamodels are also of use when we have some data from a process and we wish to construct an approximate model of the Data Generating Mechanism (DGM), and this is the context in which metamodels will be used in this paper.

Interval Predictor Models (IPMs) [3] are a type of metamodel well suited to dealing with scarce and limited data. For every realisation of their input, Interval Predictor Models output an interval rather than a single value, and hence their predictions reflect the data's uncertainty. Compared to conventional regression techniques less assumptions are made about the data, and crucially no distribution is assumed for the errors.

Our proposed approach uses an Interval Predictor Model, which is trained on the observed data, and a novel figure of merit, which is used to compare the plausibility of matches, to avoid making the assumption that the error on the measurements is Gaussian when history matching. To the authors' knowledge this has not been achieved in any existing paper.

The proposed methodology is applied to solve a realistic case study, The Imperial College Fault Model [4]. The Imperial College Fault Model consists of an oil reservoir model which must be matched to limited observational data, in order to predict the future oil production from the reservoir. This enables appropriate oil reservoir management decisions to be made. The problem is challenging as there are many plausible history matches [2], and can be solved by using IPMs to robustly quantify measurement error, without making unjustified assumptions.

In Section 2 some background information about Bayesian analysis and a common Figure of merit used for history matching is described. In Section 3 an alternative method is proposed, making use of Interval Predictor models to quantify the uncertainty in the observation data. In Section 4, two applications of the method are presented. A conclusion and recommendations for future research are given in Section 5.

2. History Matching Background

2.1. Existing Method

Bayesian Inference is a popular technique used as a tool for robust history matching when we wish to determine unknown parameters of a model from observed data, in order to use the model to make predictions about future observations [1]. An approximate model M is compared to observed measurements

μ_o which are produced when the error defined by parameter σ is applied to the unobserved true model output μ_t (i.e. the output of M), i.e.

$$\mu_o = \mu_t + \sigma = M(m) + \sigma. \tag{1}$$

In history matching problems we wish to find the parameters m of M which are responsible for producing output μ_t, and hence by association μ_o. Therefore we wish to find the value of m which is most likely given the model and our observations, which is equivalent to finding the maxima of the function

$$P(m|\mu_o, \sigma, M) = P(m|M) \int \frac{P(\mu_t|m, M)P(\mu_o|\mu_t, \sigma)}{P(\mu_o|\sigma, M)} d\mu_t, \tag{2}$$

where $P(\mu_o|\mu_t, \sigma)$ is the measurement error and is usually assumed to have a Gaussian distribution in μ_o centred around μ_t [5], normalised by σ, i.e.

$$P(\mu_o|\mu_t, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{[\mu_o - \mu_t]^2}{2\sigma^2}\right). \tag{3}$$

$P(\mu_o|\sigma, M)$ may be taken to be constant. $P(\mu_t|m, M)$ represents the modelling error and various assumptions can be made for this error as in [1], including no modelling error. In some works (for example, [2]) the analyst simply found the minima of the function

$$\Lambda(m|\mu_o) = \sum_{i=1}^{i=N_o} \frac{[\mu_{i_o} - \mu_{it}]^2}{\sigma_i^2}, \tag{4}$$

where N_o is the total number of measurements to be used for history matching and μ_{it} is the ith component of μ_t. The subscripted i allow us to perform a summation for models with more than one output. This approach is followed because the minima of $\Lambda(m|\mu_o)$ coincide with the maxima of $P(\mu_o|\mu_t, \sigma)$. In this paper we wish to avoid making the assumption that the measurement error is Gaussian, and instead use Interval Predictor Models to give an unbiased estimate of the measurement error.

3. New History Matching Approaches

3.1. Interval Predictor Models

Let us consider a Data Generating Mechanism (DGM) which acts on a vector of input variables $x \in R^{n_x}$ to produce an output $y \in R^{n_y}$. We will approximate the DGM with an Interval Predictor Model (IPM) which returns an interval for each vector $x \in X$, the set of inputs, given by

$$I_y(x, P) = \big\{y = G(x, p), p \in P\big\}, \tag{5}$$

where G is an arbitrary function and p is a parameter vector. In order to be useful, the interval we create should have the smallest range possible whilst still

enclosing all data points generated by the full model. The theory for the single output multi input case ($n_x \geqslant 1$, $n_y = 1$) of IPMs with a linear parameter dependency is described by Crespo in [3], and is summarised here. By making an approximation for G Eqn. 5 becomes

$$I_y(x, P) = \left\{ y = p^T \phi(x), p \in P \right\}, \tag{6}$$

where $\phi(x)$ is a basis (polynomial and radial bases are commonly used), and p is a member of the hyper-rectangular uncertainty set

$$P = \left\{ p : \underline{p} \leq p \leq \bar{p} \right\}, \tag{7}$$

where $\underline{p}$ and $\bar{p}$ are parameter vectors specifying the defining vertices of the hyper rectangular uncertainty set. The IPM is defined by the interval

$$I_y(x, P) = [\underline{y}(x, \bar{p}, \underline{p}), \bar{y}(x, \bar{p}, \underline{p})], \tag{8}$$

where $\underline{y}$ and $\bar{y}$ are the lower and upper bounds of the IPM, respectively. Explicitly, the lower bound is given by

$$\underline{y}(x, \bar{p}, \underline{p}) = \bar{p}^T \left(\frac{\phi(x) - |\phi(x)|}{2} \right) + \underline{p}^T \left(\frac{\phi(x) + |\phi(x)|}{2} \right), \tag{9}$$

and the upper bound is given by

$$\bar{y}(x, \bar{p}, \underline{p}) = \bar{p}^T \left(\frac{\phi(x) + |\phi(x)|}{2} \right) + \underline{p}^T \left(\frac{\phi(x) - |\phi(x)|}{2} \right). \tag{10}$$

An optimal IPM is obtained by minimising the expected value, with respect to x, of

$$\delta_y(x, \bar{p}, \underline{p}) = (\bar{p} - \underline{p})^T |\phi(x)|, \tag{11}$$

by solving the linear and convex optimisation problem

$$\left\{ \hat{\underline{p}}, \hat{\bar{p}} \right\} = \underset{u,v}{\operatorname{argmin}} \left\{ \mathbf{E}_{\mathbf{x}}[\delta_y(x, v, u)] : \underline{y}(x_i, v, u) \leq y_i \leq \bar{y}(x_i, v, u), u \leq v \right\}, \tag{12}$$

where x_i and y_i for $i = 1...N$ are training data points, which in the case of a metamodel should be sampled from the full model. The constraints ensure that all data points to be fitted lie within the bounds and that the upper bound is greater than the lower bound. This combination of objective function and constraints is linear and convex, and is known as a Type 1 IPM. In the case of a radial basis, more sophisticated constraints may be added to avoid over-fitting of the data [6].

In this work, all Interval Predictor Models are Type 1 IPMs, with polynomial bases, i.e. $\phi(x) = \left[1, x^{i_2}, x^{i_3}, ...\right]$ with $x = [x_a, x_b, ...]$ and $i_j = [i_{j,a}, i_{j,b}, ...]$ with $i_j \neq i_k$ for $j \neq k$. Polynomial bases were used, without loss of generality, in order to explain and demonstrate the method in a simple fashion, and avoid the discussion of radial basis hyperparameters. Our method is equally applicable when radial basis IPMs are used, and these methods may enable tighter IPMs

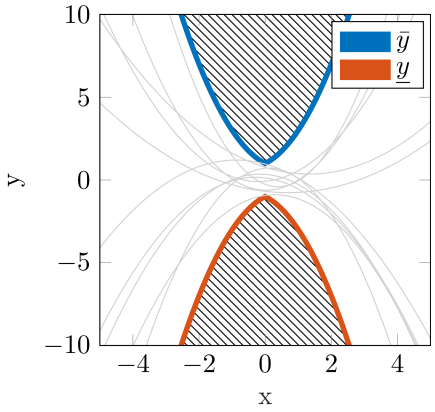

Figure 1: A degree 2 IPM in the 'data space' with $\bar{p} = [1, 1, 1]$ and $\underline{p} = [-1, -1, -1]$. Sampled polynomials within the bounds of the IPM are shown as grey lines. The blue region is outside the IPM and cannot be sampled from.

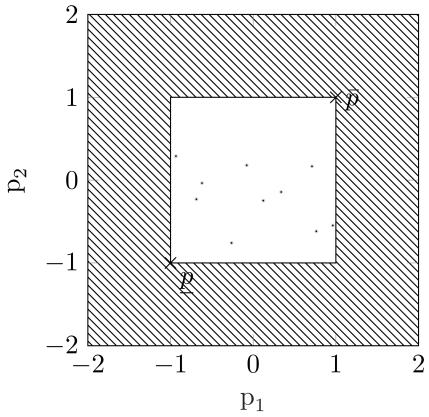

Figure 2: The IPM's hyper rectangular uncertainty set plotted in 'parameter space'. The uniformly sampled parameter vectors of the polynomials shown in Fig. 1 are displayed as points in the uncertain set.

.

to be created at equal levels of reliability for some data sets. For illustrative purposes an example degree 2 IPM is shown without training data points in Fig. 1. The hyper rectangular uncertainty set corresponding to the IPM in Fig. 1 is plotted in Fig. 2. The discontinuity observed in the upper and lower bounds is a consequence of the chosen basis, and can be avoided by choosing a basis where $\phi(x) = |\phi(x)|$.

3.2. Outliers

Outliers may be identified by finding data points satisfying two criteria. Firstly, a CDF is computed for the minimum distance of the value of p corresponding to each data point from the centre of the uncertainty set (ρ) and then, a fraction λ_p of points are identified which prevent the spread of the interval being reduced. For the ith data point the metric ρ is defined as

$$\rho_i(\bar{p}, \underline{p}) = |\hat{p}| \min_{\alpha,p} \left\{ \alpha : p_0 - \alpha\left(\hat{p}\right) \leq p \leq p_0 + \alpha\left(\hat{p}\right), y_i = p^T \phi(x_i), \alpha > 0 \right\}, \quad (13)$$

where p_0 is the centre of the hyper rectangular uncertainty set, i.e. the mean of $\underline{p}$ and $\bar{p}$, and $\hat{p}$ is the semi diagonal of the hyper rectangular uncertainty set, i.e. $\frac{\bar{p}-\underline{p}}{2}$. To illustrate why this metric is useful consider that points on the edge of the IPM in Fig. 1 will constrain the rectangle in Fig. 2.

In addition, a fraction λ_e of points are identified with the furthest squared deviation from their least squares fit predictions. The new IPM must be recalculated without the points satisfying both criteria. The analyst must make a sensible choice of λ_p and λ_e. Once data points satisfying these criteria are

removed from the training data set they are almost certainly not contained within the IPM, and this is crucial for the approach used in Section 3.3 to be valid [7]. If this is not the case then these constraints should be reinstated when performing the reliability assessment. The outlier removal algorithm is briefly demonstrated in Section 4.1.

Alternatively, outliers may be removed by constructing a Type 2 IPM. Type 2 IPMs are trained by the same objective function as Type 1 IPMs with the constraints modified to allow a fraction of points not to be enclosed by the IPM. This is formulated as

$$\left\{\hat{\underline{p}}, \hat{\bar{p}}\right\} = \underset{u,v}{\operatorname{argmin}} \left\{\mathbf{E}_{\mathbf{x}}[\delta_y(x,v,u)] : F_{\rho(v,u)}\left(\frac{||v-u||}{2}\right) \geq \lambda, u \leq v\right\}, \tag{14}$$

where $F_{\rho(v,u)}$ represents the CDF of ρ_i from Eqn. 13 and λ is a paramter to be set by the analyst representing the fraction of input data to be contained by the IPM. A full discussion of this method is given in [3]. In contrast to the previous formulation, this approach yields the removal of the optimal set of outliers.

3.3. Proposed Method

Consider a model which produces a time series of measurements as an output. Taking data points consisting of input-output pairs from the model, with time as the input, a surrogate IPM can be created for the model. The quality of the created surrogate depends upon the amount of data available, and consistency between the chosen basis and continuity of the model output. If the model has more than one output then it would be necessary to create multiple IPMs, or enumerate the data with a dummy index. The surrogate IPM will provide a robust estimate of the range of possible measurements from the model, which follows from Section 3.1. The reliability R of an IPM represents the probability that a future unobserved data point (i.e. not contained in the training data set) is contained within the IPM. If the sampled data is identically and independently distributed then R is bounded by

$$P(R \geq 1-\epsilon) \geq 1-\beta, \tag{15}$$

for complementary reliability parameter ϵ and complementary confidence parameter β satisfying

$$\binom{k+d-1}{k} \sum_{i=0}^{k+d-1} \binom{D}{i} \epsilon^i (1-\epsilon)^{D-i} \leq \beta, \tag{16}$$

where k is the number of discarded points, D is the total number of data points, and d is the number of optimisation parameters required [3]. By taking the value of reliability as confidence (i.e. $1-\beta$) asymptotically approaches 1 we can be almost certain in our lower bound for reliability. In this paper we will arbitrarily choose a 99% confidence level for the reliability and therefore the value of $1-\epsilon$ when $\beta = 0.01$ will be referred to as R^* from here on. In principle a higher level of confidence can be used, however for the examples in this paper a 99%

confidence level is an illustrative choice. The practice of allowing the confidence to approach 1 is a commonly used technique which allows us to estimate the reliability of an IPM with the single parameter R^*, which we know with near certainty [8].

In order to derive the frequentist figure of merit which we will use to rank the plausibility of history matches we will make the assumption that there is no model error, i.e. our simulations were generated by the same data generating mechanism represented by our IPM. This assumption is required for history matching in our framework, and Eqn. 16 holds regardless of this assumption. This assumption may not hold well in practice, but the process of determining the model discrepancy can be complex and therefore in many industrial case studies it is neglected [9].

Let us consider the case when the parameters of the model are matched to their true values. Let us also assume that $P(\mu_t \in I) \geq P(\mu_o \in I)$, where I is the prediction interval identified by the IPM, which is a conservative assumption for most practical cases. This is necessary because we are comparing simulations (μ_t) to an IPM describing measurements (μ_o). Although the algorithm which we will propose in this section is strictly only valid for stating the plausibility of a μ_o originating from the same process, our assumption ensures that the proposed algorithm is conservative when μ_t is history matched to the IPM of μ_o. We justify this assumption further in Appendix Appendix A. We note that in this case the probability that $\hat{C}$ of D simulated measurements fall inside the IPM is given by the binomial distribution

$$P(\hat{C} = C) = \binom{D}{C} R^C (1-R)^{D-C}, \tag{17}$$

where R is the true value for the reliability of the IPM which is not known. This is a consequence of following the frequentist approach, as the reliability bound of the IPM, which represents the probability of the DGM producing an observation outside I, must also apply to the model when the model parameters are matched to their true value.

Using the lower bound for R, i.e. R^*, and the cumulative density function for binomial distributions, we can calculate a bound for $P(\hat{C} \leq C)$, the probability that the number of simulated outputs which fall inside the IPM is less than a particular value C:

$$P(\hat{C} \leq C) \leq \sum_{i=0}^{C} \binom{D}{i} R^{*i} (1-R^*)^{D-i}. \tag{18}$$

$P(\hat{C} \leq C)$, or alternatively C, provides a frequentist figure of merit for history matches. We can compute Eqn. 18 for each simulation and discard any simulations achieving an implausible $P(\hat{C} \leq C)$ ($P(\hat{C} \leq C) \leq 0.05$ for example). A diagram of the algorithm is shown in Fig. 3.

The proposed method enables the identification of implausible regions of the space of the parameters of the model (the simulator) which we are trying to

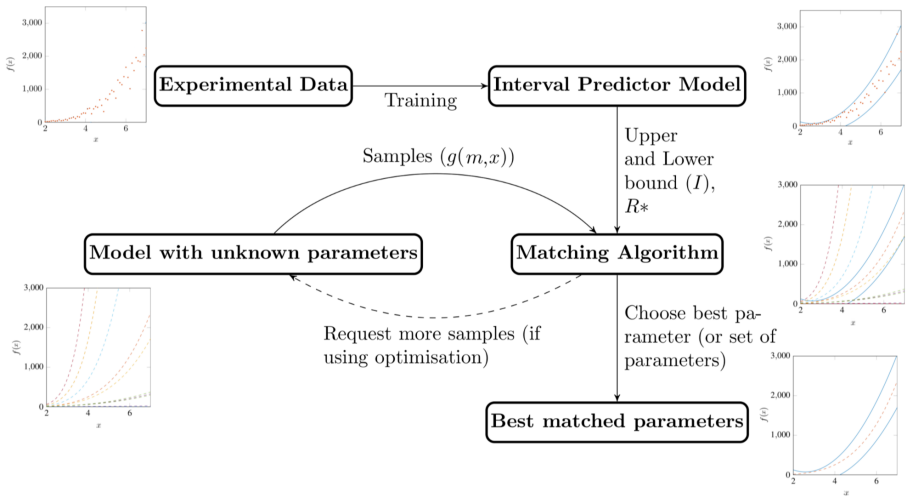

Figure 3: A diagram of the algorithm workflow

Algorithm 1 Matching Algorithm

Calculate $C(m)$, the number of simulated outputs falling inside the obtained IPM interval I, for each sampled m.
Calculate bound of $P(\hat{C} \leq C(m))$ from Eqn. 18 for each sampled m.
Find m giving $P(\hat{C} \leq C(m))$ greater than desired confidence level.

match to the observed data. This is the essence of history matching [10]. For example, we can uniformly sample points in the space of the simulator parameters and evaluate them with the figure of merit (Eqn. 18). Once this process is complete we are left with a plausible domain (similar to a frequentist confidence interval). If the IPM which we trained on the observed data is not a good representation of the true DGM then we may overestimate the plausible domain. In this way the method may be seen as a way to use imprecise probabilities to compute bounds on the p-values from frequentist inference, since p-values are simply percentiles of the CDF of the likelihood function, and in this sense the proposed method is novel. A Bayesian approach would not be applicable as we only have a bound on the CDF, but not the likelihood function itself. In addition, to use Bayesian inference we would require a prior. Making assumptions on a prior would defeat the object of not making assumptions for the likelihood function. Therefore the feasible parameter regions found in this paper are confidence intervals and not credible intervals.

It should be noted that as the output from our simulations (μ_t) has not yet been distorted by the measurement error, running the model forward with all of the matched parameter values will yield an interval μ_t, which will not necessarily contain every realisation of the random variable μ_o. ,

3.4. Implementation

OpenCOSSAN is an open source and free toolbox for uncertainty quantification in MATLAB [11] [12]. Users can download the engine, make modifications and easily quantify uncertainties in many disciplines. An OpenCOSSAN class, inheriting the abstract metamodel class, was created to implement the IPM as described in Section 3.1. The class constructor allows the user to provide any data set or full model (with choice of sampling method) and generate an IPM. The user must specify necessary parameters for the type of IPM being created. Rapid stochastic analysis can then be performed. In this paper the interior point linear optimisation algorithm in MATLAB was used to solve the optimisation program in Eqn. 12. However, using OpenCOSSAN the user is able to select different solvers such as COBYLA.

4. Numerical Applications

4.1. Function Example

4.1.1. Method

As in [1], the following function will be taken as a black box representing a DGM

$$f(x) = (x^2 + 0.1x)^2 + \eta_1, \tag{19}$$

where η_1 is normally distributed noise with standard deviation $0.2(x^2 + 0.1x)^2$, and mean zero. The history we will attempt to match will be between the times $x = 2$ and $x = 7$, at intervals of 0.1 (i.e. $x_1 = 2$, $x_2 = 2.1$ etc.). Therefore

$D = 51$ data points are available for matching. We will attempt to fit the function

$$g(m, x) = x^m, \tag{20}$$

where we wish to find plausible values of the scalar parameter m which we believe will most successfully allow us to reproduce unobserved $f(x)$ with $g(m, x)$.

We firstly create an IPM from the sampled data (i.e. samples of $f(x)$ at $x_1 = 2$, $x_2 = 2.1$ etc.) with input x and output $f(x)$. This allows us to find R^* and I. We can then compute the following figure of merit for varying m, which will allow us to compare the plausibility of different values for m:

$$\Delta(m) = \sum_{i=0}^{C(m)} \binom{D}{i} R^{*i}(1 - R^*)^{D-i}, \tag{21}$$

with

$$C(m) = \sum_{i=1}^{n_z} \begin{cases} 1 & \text{if } g(m, x_i) \in I_i \\ 0 & \text{otherwise,} \end{cases} \tag{22}$$

where I_i is the output interval for the IPM for quantised x, x_i. For simplicity we computed the figure of merit whilst varying m between 0 and 6 since we knew for this problem plausible matches would not be found outside of this range. In reality bounds on parameters are usually available from physical considerations.

The identified set of possible m was then used to compare $g(m, 10)$ with $f(10)$, by comparing the output interval for $g(m, 10)$ with 1000 realisations of $f(10)$. The output interval for $g(m, 10)$ was computed by taking the maximum and minimum values of $g(m, 10)$ for all feasible m. If the set of m is large and $g(m, 10)$ is a monotonically increasing function of m then this procedure may be completed more efficiently by only using the minimum and maximum values for m.

The example was then repeated with an artificial outlier added to the dataset to demonstrate the algorithm described in Section 3.2. The outlier data point was added at $f(x = 7) = 0$. The outlier removal algorithm was then applied to the IPM with $\lambda_e = 0.95$ and $\lambda_p = 0.9$. This means that at most 5% of data points could be removed from the IPM, i.e. no more than 2. Our choice of λ_e and λ_p implies a search for the 2 furthest data points from the least squares fit and the 5 data points which most prevented the spread of the IPM being reduced. In this example it is visually obvious when outliers are successfully removed, however for higher dimensional data sets it may be difficult to determine appropriate values for λ_e and λ_p. In this case it may be preferable to use a Type 2 IPM, where only one coefficient must be set.

4.1.2. Results

A plot of the history data with the fitted IPM of degree 2 is shown in Fig. 4, with the corresponding reliability plot shown in Fig. 5. In this example, Fig. 5

shows that $R > 0.76$ with confidence 0.99 and so $R^* = 0.76$. If a more robust prediction interval was desired then the analyst could simply choose a higher confidence and consequently a lower reliability (for example, $R > 0.711$ with confidence 0.999).

Fig. 6 shows $\Delta(m)$, the figure of merit, plotted against m. It was found that $\Delta(m) > 0.01$ in the interval between $m = 3.557$ and $m = 4.229$, and therefore values of m inside this interval were possible matches for the model parameters. Fig. 7 shows a comparison between the obtained prediction interval for and 1000 sampled values of $f(10)$. No samples fall outside the prediction interval, which is better than expected for two reasons. Firstly, the confidence in the reliability of the IPM was only 0.99 and the probability threshold we use for $\Delta(m)$ was 0.01. Therefore at each of these stages of the calculation we will lose 1% of feasible matches. Secondly, we should note that we have made the assumption that there is no model error which is clearly not true in this example as $f(x)$ has a different functional form to $g(x)$. Even if this assumption were true, using the feasible values of m to give an interval for $g(10)$ corresponds to a feasible interval for μ_t, which is not the same as our samples of $f(10)$ (which correspond to μ_o). Crucially, and as discussed in the Section 3.3, μ_t is yet to be affected by the measurement error. Consequently, the interval we have found should be regarded as a prediction interval for μ_t, and not μ_o.

Fig. 8 shows the least squares figure of merit, $\Lambda(m|\mu_o)$, plotted against m for the same realisation of the history data. All σ_i were set equal to 1, but could be set equal to g in order to rescale the figure of merit. The minima found was similar to the maxima in Fig. 6, as was expected for this example due to the Gaussian noise. Fig. 9 shows the identified minimum ($m = 4.012$) from the least squares fit with the history data.

As $g(m, x)$ is monotonically increasing in m, it is acceptable to use an interval for m, i.e. $\underline{m} = 3.557$ and $\bar{m} = 4.229$. For $g(10)$ this gives a prediction of $\bar{g}(10) = 16943$ and $\underline{g}(10) = 3605.8$.

A plot of the history data with the fitted IPM before an artificial outlier was removed is shown in Fig. 10. Fig. 12 shows the same plot after the two outliers were removed. The confidence-reliability plot for the IPM before the outlier was removed is shown in Fig. 11. Fig. 13 shows the same plot after the outlier was removed. Fig. 14 shows the obtained prediction interval for 1000 sampled values of $f(10)$ before the artificial outlier was removed. Fig. 15 shows the same plot after the outlier was removed. Before the outlier was removed the prediction interval contained 100% of the sampled values. After the outlier was removed the prediction interval still contained 99% of the sampled values, however the range was reduced from 20006 to 11301. Fig. 16 shows the CDF for ρ before and after the outliers were removed. It is clear that after the outliers are removed the IPM is a better representation of the data, as removing a small subset of data points reduced the value of ρ for most data points.

4.2. Imperial College Fault Model

The Imperial College Fault Model is a model of a reservoir which simulates the oil production and water production from a reservoir which has been pro-

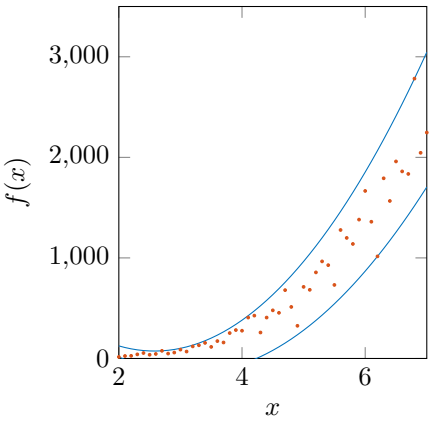

Figure 4: A history from $f(x)$ with degree 2 IPM fitted.

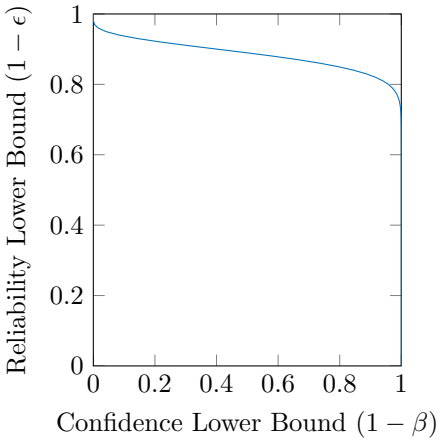

Figure 5: Confidence Parameter - Reliability Parameter plots for polynomial IPMs of degree 2, with $D = 51$ and $k = 0$.

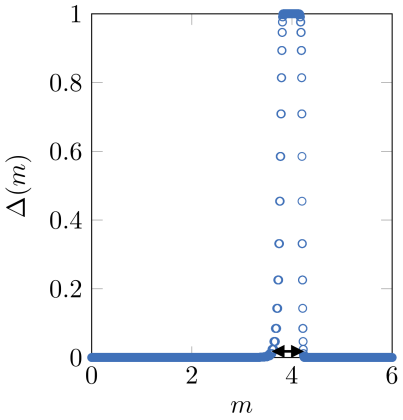

Figure 6: Plot of history matching figure of merit for uniform sample of q when matching $f(x)$ to $g(m, x)$. The identified interval for m is shown as a black double headed arrow.

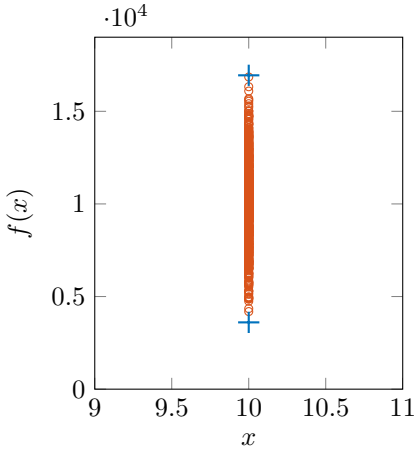

Figure 7: Plot to compare 1000 samples of $f(10)$ (circles) with a prediction interval of $g(\bar{m}, 10)$ and $g(\underline{m}, 10)$ (crosses).

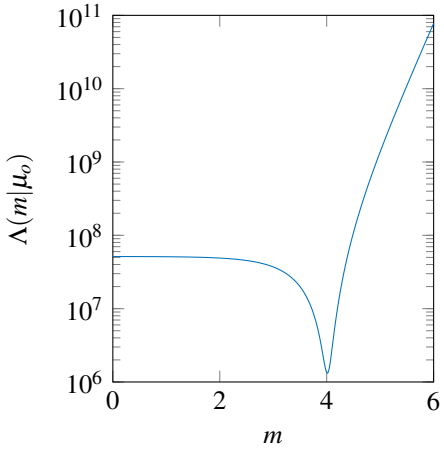

Figure 8: A least squares figure of merit computed for the same realisation of the history data.

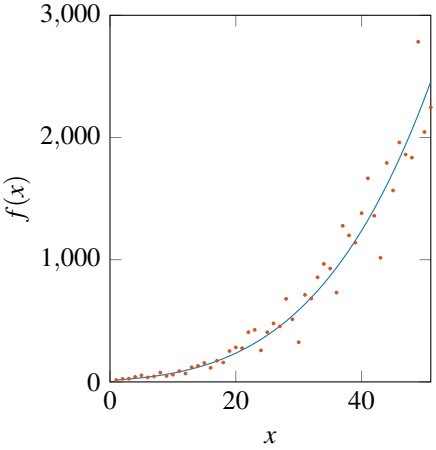

Figure 9: The minimum of the least squares figure of merit ($m = 4.012$), shown with the history data.

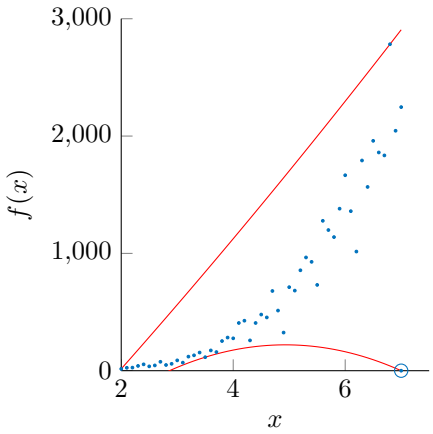

Figure 10: A history from $f(x)$ with degree 2 IPM fitted. The artificial outlier is circled.

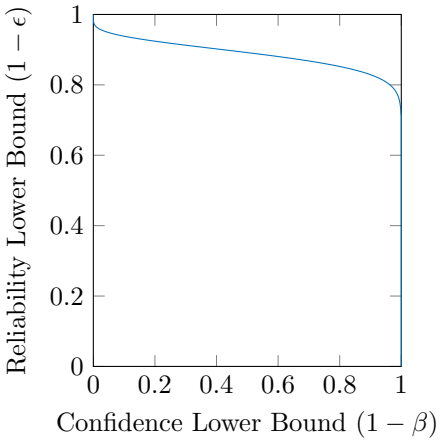

Figure 11: Confidence Parameter - Reliability Parameter plots for polynomial IPMs of degree 2, with $D = 52$ and $k = 0$.

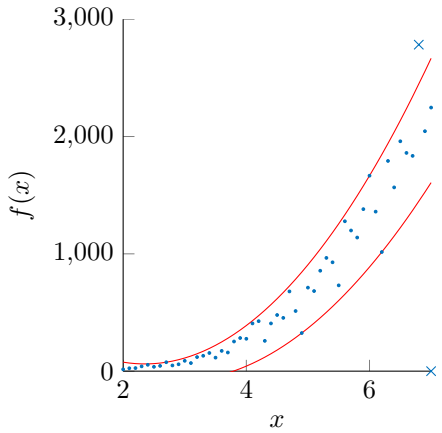

Figure 12: A history from $f(x)$ with degree 2 IPM fitted. The removed outliers are shown as crosses.

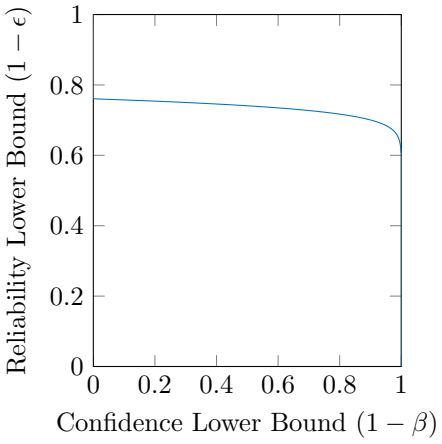

Figure 13: Confidence Parameter - Reliability Parameter plots for polynomial IPMs of degree 2, with $D = 52$ and $k = 2$.

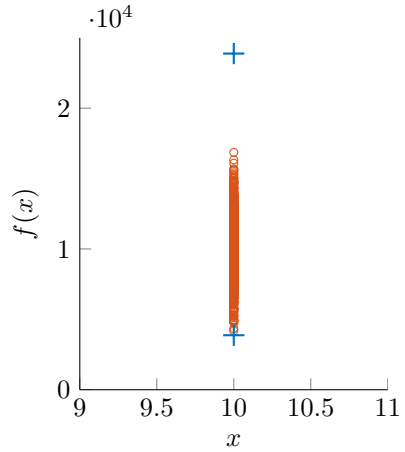

Figure 14: Plot to compare 1000 samples of $f(10)$ (circles) with a prediction interval of $g(\bar{m}, 10)$ and $g(\underline{m}, 10)$ (crosses).

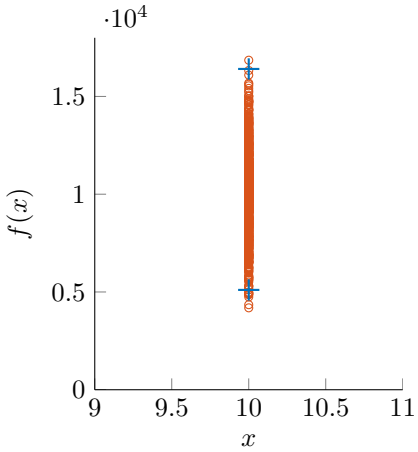

Figure 15: Plot to compare 1000 samples of $f(10)$ (circles) with a prediction interval of $g(\bar{m}, 10)$ and $g(\underline{m}, 10)$ (crosses).

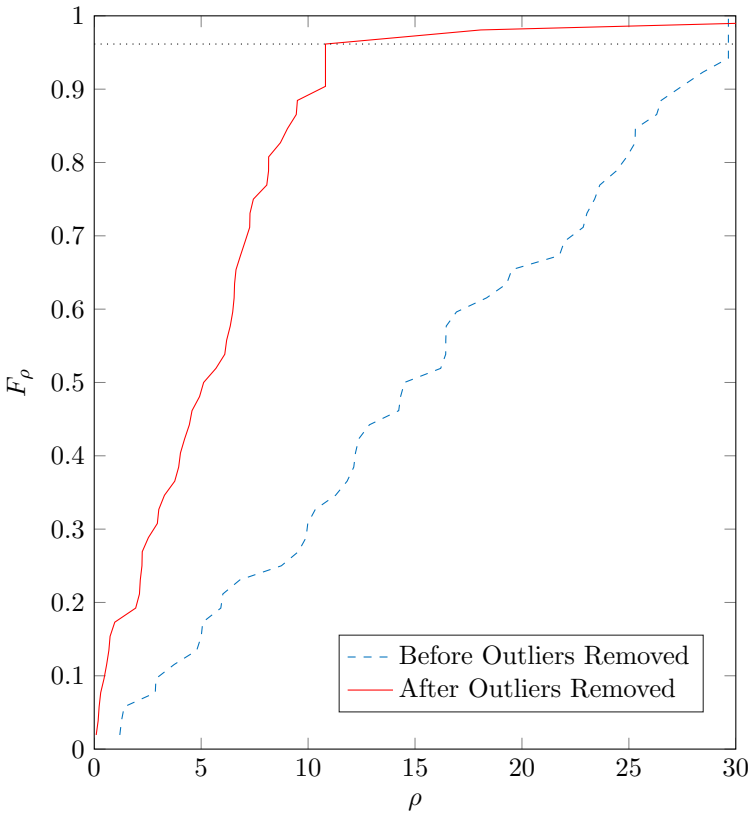

Figure 16: A plot of the empirical CDF for ρ from Eqn. 13 before and after the outliers were removed. The dotted line represents fraction of points which were removed as outliers.

ducing oil for 36 months and has now started producing water. It is important to history match the reservoir model in order to predict future oil production and to decide if parts of the well should be closed.

The history data from the reservoir was generated by the Eclipse reservoir simulator [13] with a model consisting of six layers of alternating good and poor quality sand, with a total thickness of 60 feet (18.3 m). The model is 1000 feet (304.8 m) wide with a fault at the mid-point which offsets the layers. There is a water injector well on the left edge and a producer on the right hand edge. The 'truth' history data consists of the values of the three output parameters (monthly oil production rate, water production rate and water injection rate) each month for 36 months, when the input parameters (fault throw, high quality sand permeability and low quality sand permeability) for the heterogeneous model were set to their 'true' values of 10.3 feet (3.14 m), 158.6 mD and 2.0 mD [4]. The model is heterogeneous in the sense that the porosities and permeabilities for each grid block are drawn from uniform distributions.

The published data set consists of the production history data of 159645 sampled input points from a homogeneous model, i.e. a model where the porosity and permeability in each grid block is set to the same fixed value. The future production data is also included for each sampled input point and hence if the value of the three input parameters are known then it is possible to determine the future production. Therefore we wish to use the production history data to match the three unknown input parameters.

The truth case history data has random Gaussian noise added with a standard deviation of 3% of the true value, in order to make matching the truth case to the simulation data more challenging. Although there are only three input parameters in the model there are many possible geologies offering a potential match to the truth case history data [2].

4.2.1. Method

The method described in Section 3.3 was applied to the Imperial College Fault Model problem in two ways. Firstly by matching the homogeneous model simulations to a synthetic history from the homogeneous model and then by matching the truth case data from the heterogeneous model to the simulations from the homogeneous model. For the homogeneous model a simulation was taken at random from the data set and Gaussian noise was added with standard deviation 3% of the absolute value of each measurement in order to create a realistic observation. This allowed the method to be tested without any model error being present. In both cases two separate polynomial IPMs of degree 2 were created for the oil production rate and water production rate observations. The effect of increasing the degree of the polynomial to 4 was studied for the match between the heterogeneous data and homogeneous model.

Following from Eqn. 18, the figure of merit

$$\Delta(m) = \sum_{i=0}^{C(m)} \binom{2n_t}{i} R^{*i}(1-R^*)^{2n_t-i}, \tag{23}$$

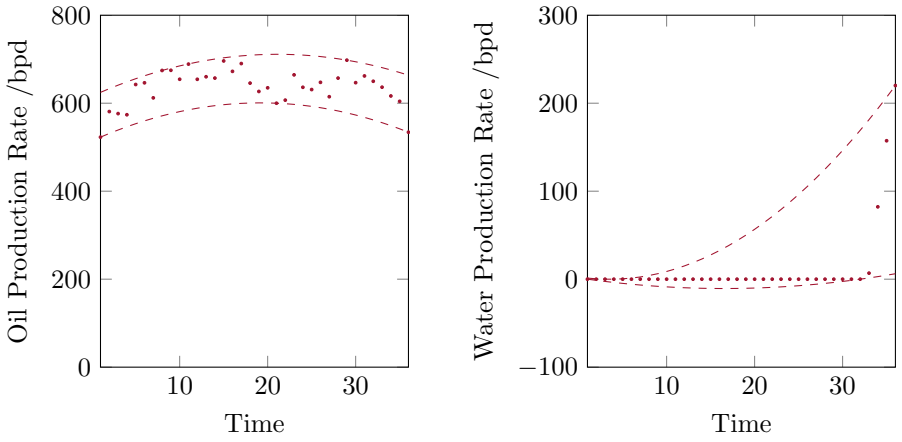

Figure 17: Oil (left) and Water (right) Production (barrels per day) Truth Case Interval Predictor Models (Synthetic - homogeneous).

was computed for all simulations in the data set to perform the match with

$$C(m) = \sum_{i=1}^{n_t} \sum_{j=1}^{n_o} \begin{cases} 1 & \text{if } y_j(m, t_i) \in I_{true-ij} \\ 0 & \text{otherwise,} \end{cases} \tag{24}$$

where $y_j(m, t_i)$ is the simulation output variable j at time t_i for model parameters m, and $I_{true-ij}$ is the output interval for the IPM for variable j at time t_i. Measurements of n_o outputs were taken at n_t points in time.

4.2.2. Results

The Interval Predictor Models created for the synthetic truth data (from the homogeneous model) are shown in Fig. 17. The Interval Predictor Models created for the truth data from the heterogeneous model are shown in Fig. 18. The confidence and reliability parameters for the IPMs of degree 2 are shown in Fig. 19. From this plot it is clear that $R > 0.675$ with confidence 0.99 and so $R^* = 0.675$.

Fig. 20 shows $\Delta(m)$ plotted against fraction of simulation points. This allows the analyst to easily estimate how many simulations will be deemed implausible by choosing a particular p-value threshold for high-dimensional data sets.

Fig. 21 shows a plot of model parameters corresponding to simulations with a value of $\Delta(m) > 0.01$ (i.e. $C \geq 39$) when matching to the synthetic truth case from the homogeneous model. This corresponds to a probability of less than 0.01 that the matches we have discarded (and not shown in the plot) are a possible match to the model. This probability could be adjusted by an analyst to any desired level of confidence. A reasonable number of potential matches are shown and the truth case is contained within the bulk of these. Fig. 22 shows a plot of model parameters with $C \geq 70$. These are the best matches to the model, however matches with a high value of $\Delta(m)$ were discarded, and hence

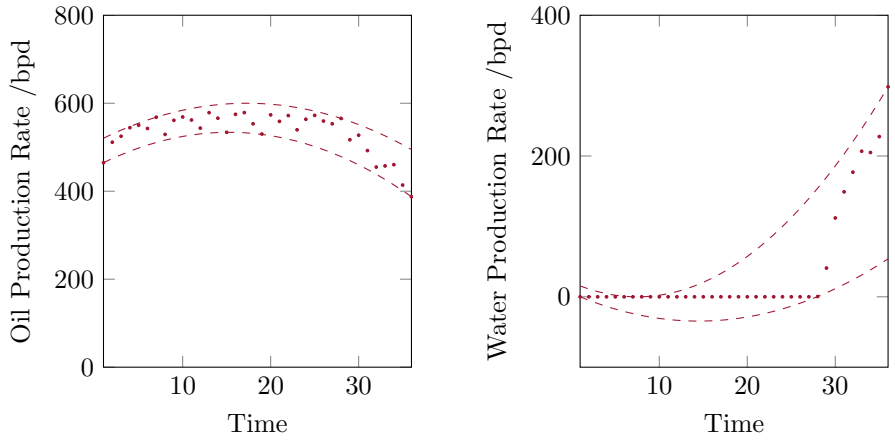

Figure 18: Oil (left) and Water (right) Production (barrels per day) Truth Case Interval Predictor Models (Truth - heterogeneous).

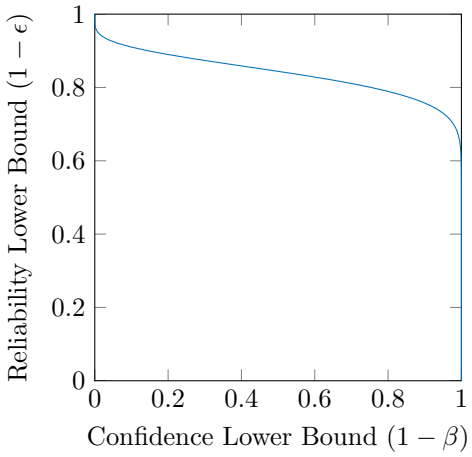

Figure 19: Confidence Parameter - Reliability Parameter plots for polynomial IPMs of degree 2, with $D = 72$ and $k = 0$.

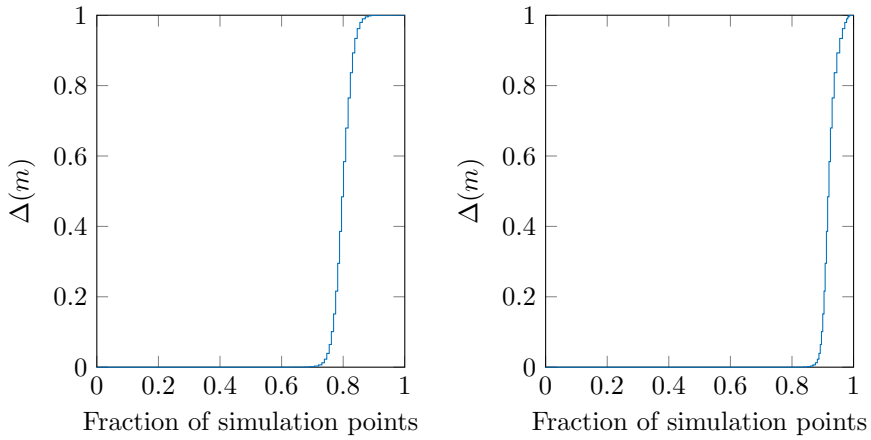

Figure 20: Plot of history matching figure of merit for Synthetic - homogeneous (left) and Truth - heterogeneous (right) data.

an analyst should be suspicious of the exhaustiveness of this search. Although there are still many matches, they are confined to a relatively compact area of the total parameter space. Crucially the truth case is contained within this area.

Fig. 23 shows a plot of possible matches, i.e. parameter values, m, which achieve $\Delta(m) > 0.01$ when they are matched to the truth case from the heterogeneous model. 13% of parameter value samples are feasible matches, which is fewer than in the corresponding plot for the synthetic truth case. The true value of the parameters is on the edge of the main bulk of matches. Fig. 24 shows a plot of model parameters with $C \geq 70$. The feasible region is smaller, as expected due to the incorrect assumption that there is no model error. The truth case is not contained within the feasible region, however the low quality sand permeability parameter is well matched and the good quality sand permeability is approximately 40% of the range of the search space of this parameter away from its true value. The Fault Throw is approximately 20% of the range of the search space of this parameter away from its true value, and is not well matched.

Visually the IPM in Fig. 18 does not represent the data well, as for a long time the water production is zero but the IPM suggests a higher upper bound. With the consequence of reducing the reliability lower bound to $R^* = 0.5$ an IPM of degree 4 can be fitted to the data as shown in Fig. 25 which, visually, is a better fit for the data. Fig. 26 shows a plot of model parameters with $C \geq 67$ for these IPMs. In this plot fault throw and low quality sand permeability are well matched but good quality sand permeability is not well matched. Fig. 27 shows the figure of merit, $\Delta(m)$, plotted against fraction of simulation points for the IPMs of degree 4. This plot represents a CDF for $\Delta(m)$ (with CDF on the horizontal axis), which allows us to observe how many of the sampled

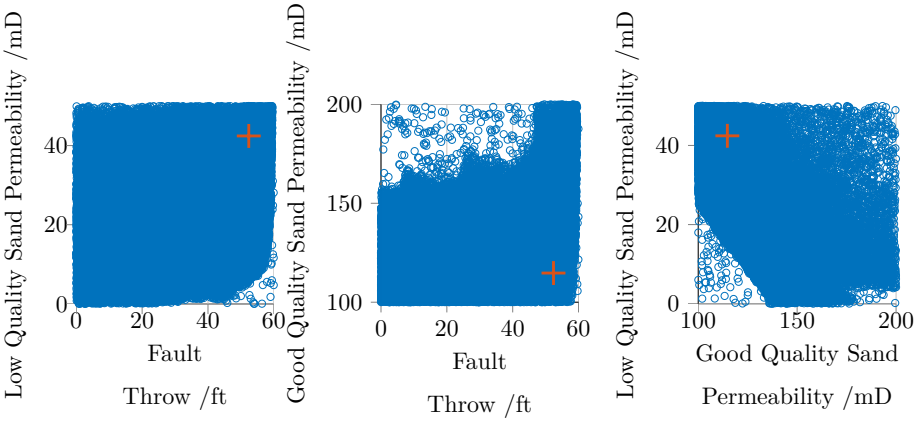

Figure 21: Plot of History Matched parameters with $\Delta(m) > 0.01$ for Synthetic case. The True value of the parameters is shown as a red cross for comparison.

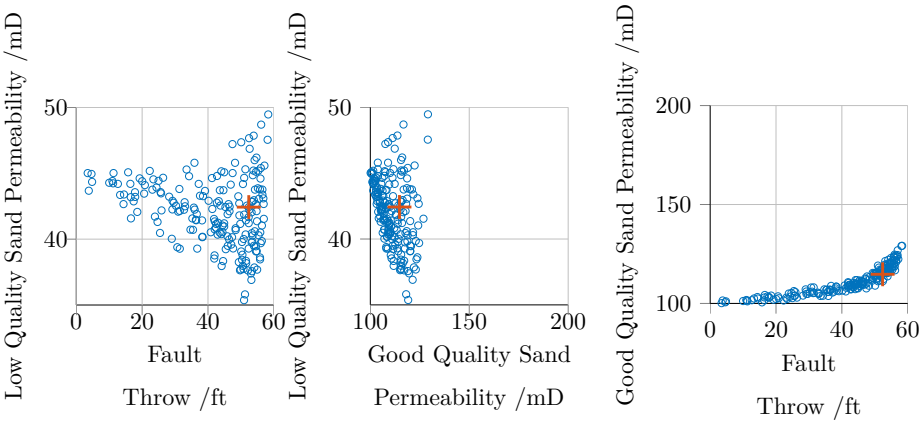

Figure 22: Plot of History Matched parameters with $C \geq 70$ for Synthetic case. The True value of the parameters is shown as a red cross for comparison.

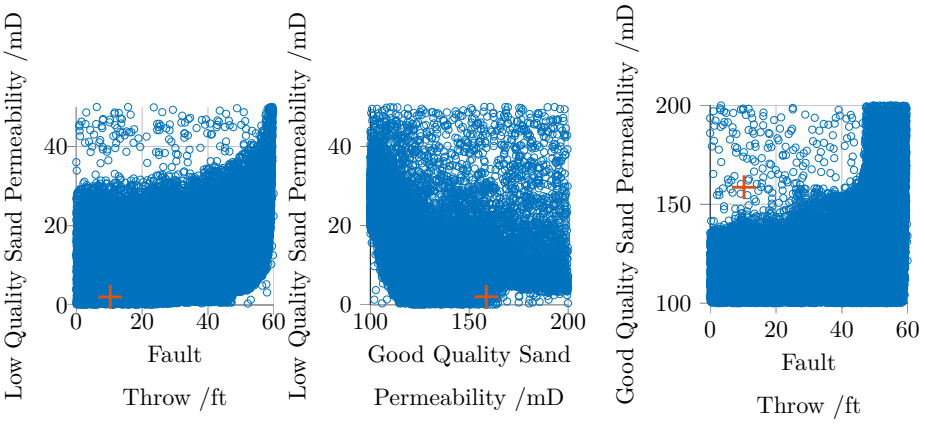

Figure 23: Plot of History Matched parameters with $\Delta(m) > 0.01$ for truth case (IPM degree 2). The True value of the parameters is shown as a red cross for comparison.

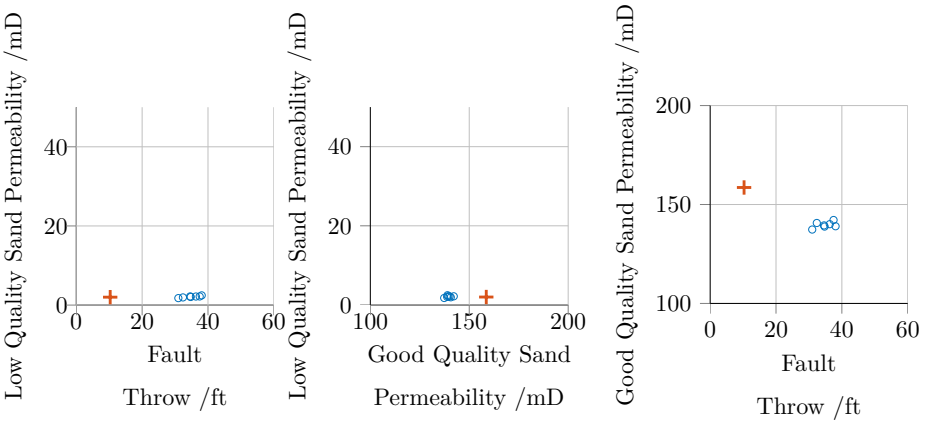

Figure 24: Plot of History Matched parameters with $C \geq 70$ for truth case (IPM degree 2). The True value of the parameters is shown as a red cross for comparison.

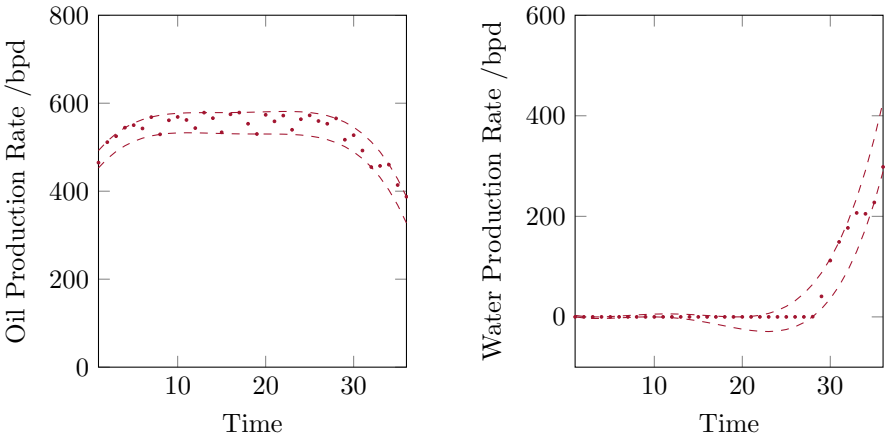

Figure 25: Oil (left) and Water (right) Production Truth Case Interval Predictor Models (Heterogeneous truth case IPM degree 4)

parameter values are good matches to the true data.

4.2.3. Discussion

Matching the 'synthetic' truth case model was more successful than matching the actual truth case model. This should be expected as the matching technique made no attempt to account for modelling error. This error is clearly significant in the later case, as the fraction of simulation points achieving a high value of $\Delta(m)$ is lower. It would be desirable to integrate a technique to account for this error into the framework in order to achieve more accurate values for the matched parameters. Interestingly, although the value of R^* is reduced when the degree of the IPM is increased, the fraction of simulation points achieving a high value for $\Delta(m)$ is roughly similar, but the fraction of points achieving a value for $\Delta(m) > 0.01$ increases. Therefore if a robust evaluation of feasible matches is required the analyst should attempt to move the "knee" of this graph to the right in order to reduce the number of feasible matches. However, if the analyst is simply interested in very close history matches then increasing the degree of the IPM may be useful.

The technique would also be of use if some data has been lost, for example if a reading at a particular period in time is not available. In this case the IPM should provide an unbiased estimate of the lost reading and so the history match would still be possible, provided that the requirements for Eqn. 16 to be valid are true.

The IPM illustrates the effect of taking more data as the analyst will be able to calculate how many more data points are required to increase the value of R^* to a particular value, and therefore improve the method's ability to discriminate between different potential values for the matched parameters, although the analyst would not be able to calculate exactly how much this would reduce the

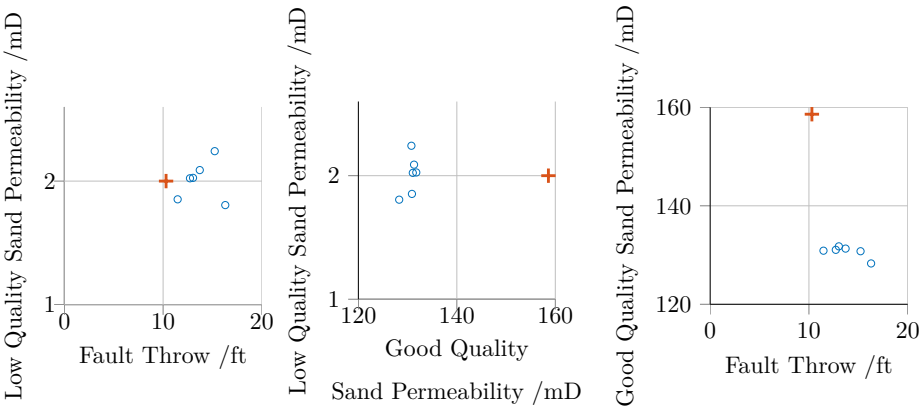

Figure 26: Plot of History Matched parameters with $C \geq 67$ for truth case (IPM degree 4). The True value of the parameters is shown as a red cross for comparison.

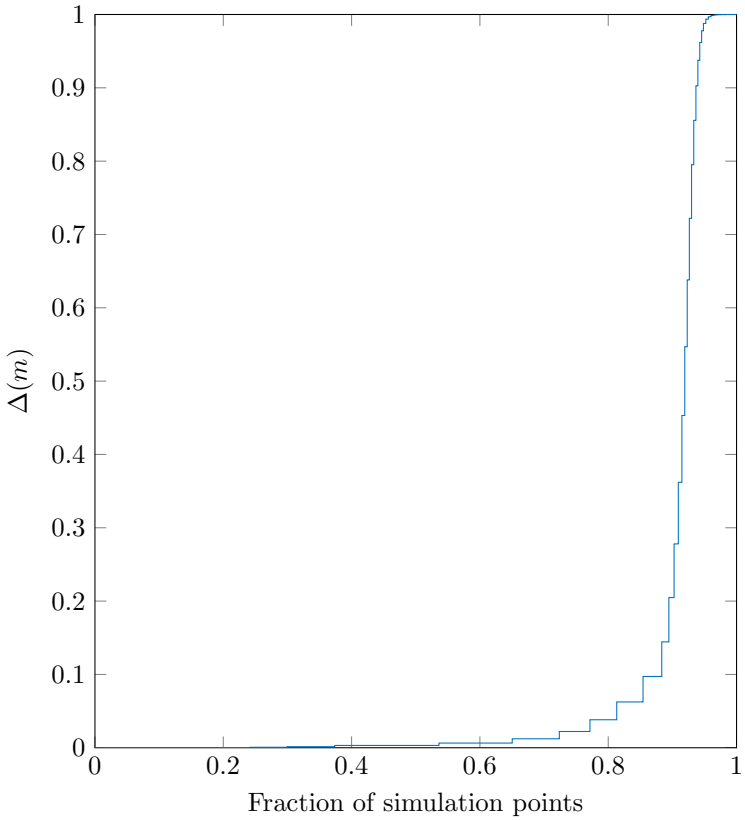

Figure 27: Plot of history matching figure of merit for and Truth - heterogeneous data (IPM degree 4).

number of feasible matches by. In the case that a large number of measurement data points are available, the analyst could also remove outliers from the IPM using the procedure described in [3], if the reduction in R^* was acceptable. In particular, when matching the homogeneous observations to the homogeneous model, it is likely that a better match could have been found by increasing the degree of the IPM, and potentially removing outliers.

4.3. Computational Considerations

In the second example there is a significant computational burden required to generate the simulations from the homogeneous model which we are history matching. One possibility to speed up the calculation would be to introduce a metamodel after some initial samples of the figure of merit have been calculated. A metamodel (for example a Gaussian process) could be made of the figure of merit, and then used to determine where future samples should be generated in order to find the minimum [14].

All of these techniques require linear optimisation to train the necessary IPMs, which in some scenarios would be more time consuming than simply performing more simulations. The number of constraints in the IPMs used in this paper is $D + 2d$ [3], and so IPMs requiring many parameters and data points (such as those with many inputs) will take longer to train, because solving optimisation problems with more constraints is more time consuming. In practice the complexity of an IPM with d inputs and $D+2d$ constraints scales with order $(2d)^2(2d + D)$ (this can be found by considering that in practice a linear optimisation problem with h design variables and H constraints scales with order h^2H) [15].

5. Conclusions

A method for history matching has been proposed which does not require the analyst to make the assumption that the measurement error has a Gaussian distribution. Instead, Interval Predictor Models have been used to robustly quantify the measurement error in observation data. The method computes the reliability of an IPM, and hence the calculate bounds on the p-values for potential matches. Therefore, it is no longer necessary to make unjustified assumptions about the distribution of measurement error when history matching. The proposed method is general and could be applied to other history matching problems, regardless of the amount of measurement error.

However, this method does not attempt to address the model form error of the computational model being matched (the discrepancy between the functional form of the matched model and the true data generating mechanism). We make a distinction between this epistemic uncertainty which cannot be quantified using our method and the aleatory 'noise' in the observed data which we analyse using the Interval Predictor Model. This distinction is a controversial issue in the wider Uncertainty Quantification community.

Therefore, the proposed technique is most applicable in cases where the model form error is small or not present. However, if this discrepancy is very

large and the number of data points used to train the IPM is high, then in some circumstances there will be no feasible matches. This is reassuring as such a discrepancy would mean very few measurements were falling inside the IPM, and hence the probability of a match should be very low.

The proposed method has been applied successfully to a simple example and a realistic model of a reservoir. A feasible region was identified for both of the history matching problems.

6. Acknowledgements

Funding: This work was supported by the *EPSRC Centre for Doctoral Training in Nuclear Fission - Next Generation Nuclear* [grant number EP/L015390/1] which is gratefully acknowledged by the authors.

References

[1] J. Carter, Using bayesian statistics to capture the effects of modelling errors in inverse problems, Mathematical geology 36 (2) (2004) 187–216. doi:10.1023/B:MATG.0000020470.51595.6d.

[2] Z. Tavassoli, J. N. Carter, P. R. King, An analysis of history matching errors, Computational Geosciences 9 (2-3) (2005) 99–123. doi:10.1007/s10596-005-9001-7.
URL http://link.springer.com/article/10.1007/s10596-005-9001-7

[3] L. G. Crespo, S. P. Kenny, D. P. Giesy, Interval predictor models with a linear parameter dependency, Journal of Verification, Validation and Uncertainty Quantification 1 (2) (2016) 021007–021007–10. doi:10.1115/1.4032070.

[4] Standard Models | Imperial College London [cited 2016-04-14].
URL http://www.imperial.ac.uk/engineering/departments/earth-science/research/research-groups/perm/standard-models/

[5] J. L. Beck, L. S. Katafygiotis, Updating models and their uncertainties. i: Bayesian statistical framework, Journal of Engineering Mechanics 124 (4) (1998) 455–461. doi:10.1061/(ASCE)0733-9399(1998)124:4(455).

[6] L. G. Crespo, S. P. Kenny, D. P. Giesy, R. B. Norman, S. Blattnig, Application of interval predictor models to space radiation shielding, in: 18th AIAA Non-Deterministic Approaches Conference, American Institute of Aeronautics and Astronautics, 2016. doi:10.2514/6.2016-0431.

[7] M. C. Campi, S. Garatti, A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality, Journal of Optimization Theory and Applications 148 (2) (2011) 257–280.

doi:10.1007/s10957-010-9754-6.
URL http://link.springer.com/article/10.1007/s10957-010-9754-6

[8] M. C. Campi, G. Calafiore, S. Garatti, Interval predictor models: Identification and reliability, Automatica 45 (2) (2009) 382–392.

[9] M. Kathrada, J. N. Carter, others, Case studies of successfully history matched reservoir simulation models using a powerful optimization algorithm being of limited predictive value, in: Abu Dhabi International Petroleum Exhibition and Conference, Society of Petroleum Engineers, 2010. doi:10.2118/136659-MS.
URL https://www.onepetro.org/conference-paper/SPE-136659-MS

[10] Z. Gong, F. A. DiazDelaO, M. Beer, Sampling schemes for history matching using subset simulation, in: 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, 2017, pp. 154–164. doi:10.7712/120217.5359.16948.

[11] E. Patelli, Handbook of Uncertainty Quantification, Springer International Publishing, Cham, 2016, Ch. COSSAN: A Multidisciplinary Software Suite for Uncertainty Quantification and Risk Management, pp. 1–69. doi:10.1007/978-3-319-11259-6_59-1.
URL http://dx.doi.org/10.1007/978-3-319-11259-6_59-1

[12] E. Patelli, M. Broggi, S. Tolo, J. Sadeghi, Cossan software: A multidisciplinary and collaborative software for uncertainty quantification, in: Proceedings of the 2nd ECCOMAS thematic conference on uncertainty quantification in computational sciences and engineering, UNCECOMP 2017, Rhodes Island, Greece, 2017. doi:10.7712/120217.5364.16982.

[13] S. GeoQuest, Eclipse reservoir simulator, Technical description.

[14] M. Frean, P. Boyle, Using gaussian processes to optimize expensive functions, in: Australasian Joint Conference on Artificial Intelligence, Springer, 2008, pp. 258–267. doi:10.1007/978-3-540-89378-3_25.

[15] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004. doi:10.1017/CBO9780511804441.

Appendix A. Connection between $P(\mu_o \in I)$ and $P(\mu_t \in I)$

In order for the algorithm in Section 3.3 to be valid we require $P(\mu_t \in I) \geq R^*$, or alternatively $P(\mu_t \in I) \geq P(\mu_o \in I)$ (which is itself equivalent to $P(\mu_t \in I|\mu_o \in I) \geq P(\mu_o \in I|\mu_t \in I)$). We do not deem this assumption to be particularly strong since $P(\mu_o|\mu_t)$ is a distribution of noise which 'scatters' the observations, but should not bias them in any particular direction. In fact, it is

satisfied in the trivial case that σ is additive and the median is zero. However, in this section we show that the assumption can hold in more general cases. An example of $P(\mu_o)$ is shown in Fig. A.28.

To demonstrate why we believe that this assumption is conservative, consider a general probability distribution on the real numbers, $\mathrm{P}(\mu_o)$, i.e. in this example μ_o is a scalar. Let us assume that an interval, I, has been found which contains at least R of the probability density of μ_o. In other words we have an IPM where

$$P(\mu_o \in I) \geq R^*. \tag{A.1}$$

Now let us assume that the distribution $P(\mu_o|\mu_t)$ is such that μ_t does not reside in the tails (where we define the complement of I, (i.e. the tails) as having density less than $1 - R^*$). In other words, we impose the following constraint on $P(\mu_o|\mu_t)$:

$$\min(P(\mu_o > \mu_t), P(\mu_t > \mu_o)) \geq 1 - R^*. \tag{A.2}$$

Combining Equation A.1 with Equation A.2 we are forced to conclude that $P(\mu_o \in I) \geq \max(P(\mu_o > \mu_t), P(\mu_t > \mu_o))$. Hence $\mu_t \in I$, which means that $P(\mu_t \in I) \geq R^*$ is conservative.

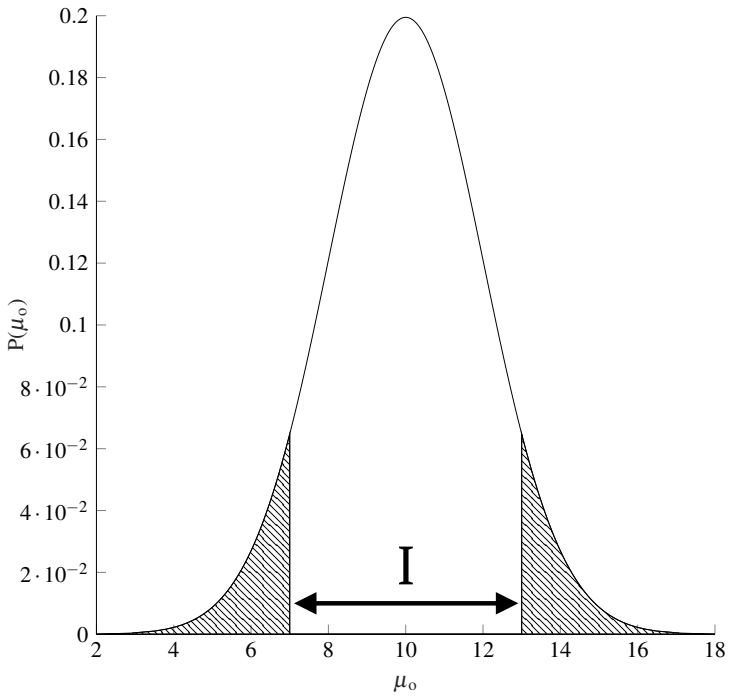

Figure A.28: An example distribution $P(\mu_o)$. The distribution shown was chosen to be Gaussian for illustrative purposes only - Gaussianity plays no part in the results obtained in this section. In the example shown I is centred on the mean of the distribution but this need not be the case. I can be anywhere, provided that $\int_{\mu_o \in I} P(\mu_o) d\mu_o \geq R$.