History Matching with Robust Predictive Metamodels

Jonathan Sadeghi and Edoardo Patelli
Institute for Risk and Uncertainty, University of Liverpool

Abstract: In this paper a novel approach is presented for history matching models without making assumptions about the measurement error when the available data is limited. Interval Predictor Models are used to robustly quantify the noise in the observation data and a novel objective function is proposed to quantify the quality of matches in a frequentist probabilistic framework. The method is applied to a simple numerical example in order to evaluate its applicability and efficacy, and the proposed method identifies a reasonably small feasible region for the matched parameters. The effect of increasing the number of data points on the history matching is also discussed.

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 (known as the truth case) 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 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 [7].

Bayesian Inference is a technique frequently used to allow data (sometimes known as the evidence) to be combined with our prior belief in the model parameters. However it requires that the likelihood of observing a particular set of data is known, given a particular model, in addition to having a prior belief in the model parameters (although some techniques exist to choose so-called non-informative priors) [6]. Determining these two pieces of information can be challenging when the available data is limited.

In this paper, 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 to obtain samples for a Monte Carlo simulation, for example. An approximate model is needed to perform many simulations quickly. In this context metamodels are models of models. Metamodels are also of use when we have some data from a process and we wish to construct an approximate black box model of the Data Generating Mechanism (DGM), and this is the context in which metamodels will be mainly used in this paper.

Interval Predictor Models (IPMs) [2] are a type of metamodel well suited to dealing with scarce and limited data. For each initialization of their input parameters, 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 and a novel objective function 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.
In Section 2 some background information about Bayesian analysis and a common objective function used for history matching is described. In Section 3.1 an alternative method is proposed, making use of Interval Predictor models to quantify the uncertainty in the observation data. In Section 4 the proposed methodology is applied to solve a simple test case, where we would like to fit a power law to the data is generated from a different analytic function with added noise. 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]. These techniques will be outlined in this section. 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. In history matching problems we wish to find the parameters m of M which are responsible for producing output 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

$$P(m|\mu_o,\sigma,M) = P(m|M)\int \frac{P(\mu_t|m,M)P(\mu_o|\mu_t,\sigma)}{P(\mu_o|\sigma,M)}d\mu_t, \tag{1}$$

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, normalised by σ, i.e.

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

$P(\mu_o|\sigma,M)$ 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, [7]) the analyst simply found the minima of the function

$$\Lambda(m|\mu_o) = \sum_{i=1}^{i=N_o}\frac{[\mu_{i_o}-\omega_i(m)]^2}{\sigma_i^2}, \tag{3}$$

where N_o is the total number of measurements to be used for history matching and $\omega(m)$ is the prediction of μ by M. The subscripted i allow us to perform a summation for models with more than one output. This approach is followed because thee 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 Approach

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

$$I_y(x,P) = \{y = G(x,p), p \in P\}, \tag{4}$$

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 [2], and is summarised here. By making an approximation for G Eqn. 4 becomes

$$I_y(x,P) = \{y = p^T \phi(x), p \in P\}, \tag{5}$$

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

$$P = \{p : \underline{p} \leq p \leq \bar{p}\}, \tag{6}$$

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

$$I_y(x,P) = [\underline{y}(x,\bar{p},\underline{p}), \bar{y}(x,\bar{p},\underline{p})], \tag{7}$$

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

$$\underline{y}(x,\bar{p},\underline{p}) = \bar{p}^T \left(\frac{\phi(x) - |\phi(x)|}{2}\right) + \underline{p}^T \left(\frac{\phi(x) + |\phi(x)|}{2}\right), \tag{8}$$

and the upper bound is given by

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

An optimal IPM is yielded by minimising the expected value of

$$\delta_y(x,\bar{p},\underline{p}) = (\bar{p} - \underline{p})^T |\phi(x)|, \tag{10}$$

by solving the linear and convex optimisation problem

$$\{\hat{\underline{p}}, \hat{\bar{p}}\} = \underset{u,v}{\operatorname{argmin}} \{\mathbf{E_x}[\delta_y(x,v,u)] : \underline{y}(x_i,v,u) \leq y_i \leq \bar{y}(x_i,v,u), u \leq v\}, \tag{11}$$

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 [3]. In this work all Interval Predictor Models are Type 1 IPMs, with polynomial bases, i.e. $\phi(x) = [1, x^{i_2}, x^{i_3}, ...]$ with $x = [x_a, x_b, ...]$ and $i_j = [i_{j,a}, i_{j,b}, ...]$ with $i_j \neq i_k$ for $j \neq k$.

3.2 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 provided enough data points are available. 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. R is bounded by

$$P(R \geq 1-\varepsilon) \geq 1-\beta, \tag{12}$$

for reliability parameter ε and confidence parameter β satisfying

$$\binom{k+d-1}{k} \sum_{i=0}^{k+d-1} \binom{D}{i} \varepsilon^{i}(1-\varepsilon)^{D-i} \leq \beta, \tag{13}$$

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 [2]. By taking the value of reliability as confidence 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 $1-\varepsilon$ 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.

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. Therefore, for any simulation, the probability that $\hat{C}$ of $\hat{D}$ simulated measurements fall inside the IPM is given by the binomial distribution,

$$P(\hat{C} = C) = \binom{\hat{D}}{C} R^{C}(1-R)^{\hat{D}-C}, \tag{14}$$

where R is the true value for the reliability of the IPM which is not known. Using the lower bound for R, 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 measurements which fall inside the IPM is less than a particular value C:

$$P(\hat{C} \leq C) \leq \sum_{i=0}^{C} \binom{\hat{D}}{i} R^{*i}(1-R^{*})^{\hat{D}-i}. \tag{15}$$

$P(\hat{C} \leq C)$, or alternatively C, provides a figure of merit for history matches. We can compute Eqn. 15 for each simulation and discard any simulations achieving an unsatisfactory $P(\hat{C} \leq C)$ ($P(\hat{C} \leq C) \leq 0.05$ for example). By using this method we simply search for feasible values of the model parameters in a robust manner, rather than attempting to search for a most likely estimate of the parameters. In other words we are attempting to find and discard values of the model parameters which we are sure do not fall within the confidence interval. Depending on the quality of our reliability bounds it may only be possible to identify a fraction of these values. 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. 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 distribution. Making assumptions about a prior distribution 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 error parameter (σ), then the assumption that they are from the same DGM as the observations used to construct the IPM (μ_o) may be unjustified. The uncertainty from the measurement error is causing an uncertainty in the matched parameters of the model when we violate this assumption. Although we can be reasonably sure that the samples of the input parameters we discarded were not matches, the set of potential matches we have found is an overestimate. Running the model forward with all of the matched parameter values will yield an interval μ_t, which will not necessarily contain all of μ_o. A more robust approach could be achieved by training a second IPM with the model parameters to be matched and time as inputs. Then robust predictions could be made by propagating the matched parameter values through the second IPM.

3.3 Implementation

OpenCOSSAN is an open source and free toolbox for uncertainty quantification in MATLAB [4] [5]. 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. In this paper the interior point linear optimisation algorithm in MATLAB was used to solve the optimisation program in Eqn. 11.

4 Numerical Applications

4.1 Method

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

$$f(z) = (z^2 + 0.1z)^2 + \eta_1, \tag{16}$$

where η_1 is normally distributed noise with standard deviation $0.2f(z)$. The history we will attempt to match will be between the times $z = 2$ and $z = 7$, at intervals of 0.1 (i.e. $z_1 = 2$, $z_2 = 2.1$ etc.). Therefore $D = 51$ data points are available for matching. We will attempt to fit the function

$$g(q,z) = z^q, \tag{17}$$

where we wish to find the value of the parameter q which we believe will most successfully allow us to reproduce unobserved $f(z)$ with $g(q,z)$. In practice $g(q,z)$ would usually be a complex computational model, chosen to represent the known physics of the process $f(z)$ as closely as possible.

We will vary q between 0 and 6, with an interval between samples of 0.001. The following objective function was computed for all simulations in the data set to perform the match:

$$\Delta(q) = \sum_{i=0}^{C(q)} \binom{\hat{D}}{i} {R^*}^i (1 - R^*)^{\hat{D}-i}, \tag{18}$$

with

$$C(q) = \sum_{i=1}^{n_z} \begin{cases} 1 & \text{if } g(q,z_i) \in I_i \\ 0 & \text{otherwise,} \end{cases} \tag{19}$$

where I_i is the output interval for the IPM for quantised z, z_i. In this example $\hat{D} = 51$.

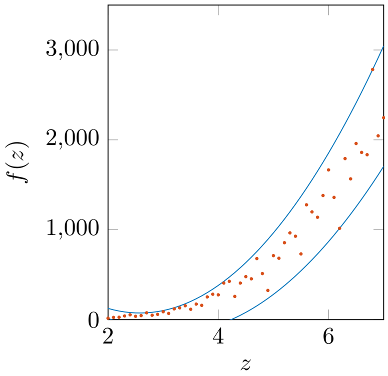

Figure 1: A history from $f(z)$ with degree 2 IPM fitted.

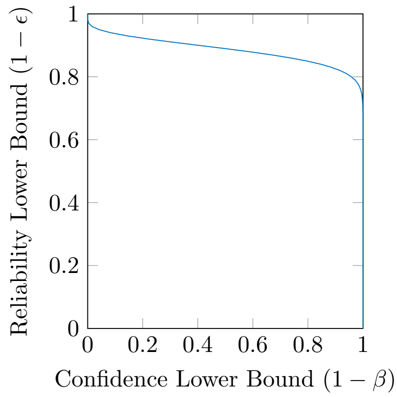

Figure 2: Confidence Parameter - Reliability Parameter plots for polynomial IPMs of degree 2, with $D = 51$ and $k = 0$.

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

4.2 Results

A plot of the history data with the fitted IPM of degree 2 is shown in Fig. 1, with the corresponding reliability plot shown in Fig. 2. In this example, Fig. 2 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. 3 shows $\Delta(q)$, the objective function, plotted against q for $\hat{D} = 51$. It was found that $\Delta(q) > 0.01$ in the interval between $q = 3.557$ and $q = 4.229$, and therefore values of q inside this interval were possible matches for the model parameters.

As $g(q,z)$ is monotonically increasing in q, it is acceptable to use an interval for q, i.e. $q = 3.557$ and 4.229. For $g(10)$ this gives a prediction of $\bar{g}(10) = 16943$ and $\underline{g}(10) = 3605.8$. Fig. 4 shows a comparison between the obtained prediction interval for $\hat{D} = 51$ and 1000 sampled values of $f(10)$. All of the samples fall inside the prediction interval, which is an unexpectadly good result for two reasons. Firstly, the confidence in the reliability of the IPM was only 0.99 and the probability threshold we use for $\Delta(q)$ 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 $f(z)$ has a different functional form to $g(z)$. Even if this assumption were true, using the feasible values of q 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 3.2, μ_t is yet to be affected by the error defined by the parameter σ. Consequently, the interval we have found should be regarded as a prediction interval for μ_t, and not μ_o.

4.3 Discussion

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

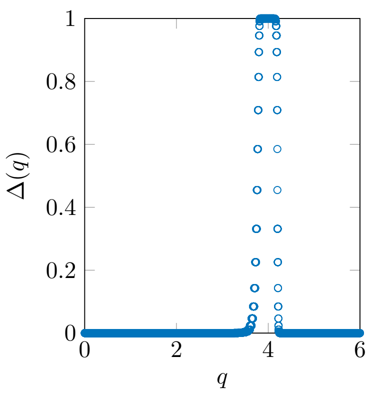

Figure 3: Plot of history matching figure of merit for uniform sample of q when matching $f(z)$ to $g(q,z)$ for $\hat{D} = 51$.

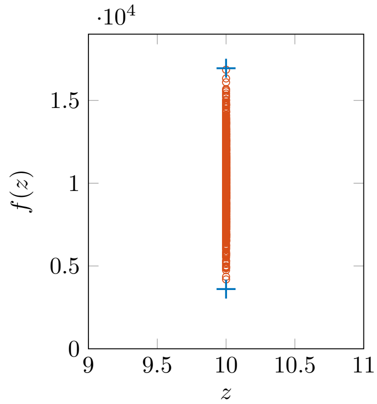

Figure 4: Plot to compare 1000 samples of $f(10)$ (red) with a prediction interval of $g(\bar{q}, 10)$ and $g(\underline{q}, 10)$ (blue) for $\hat{D} = 51$.

improve the method's ability to discriminate between different potential values for the matched parameters. 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 [2], if the reduction in R^* was acceptable.

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 relies on our ability to bound the reliability of an IPM, and hence to calculate bounds on the p-values for potential matches using a bound for the binomial CDF. The proposed method has been applied successfully to a simple test case where a feasible parameter interval was identified. The proposed method is general and could be applied to other history matching problems, regardless of the amount of measurement error.

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.

Currently the discrepancy between the functional form of the matched model and the true data generating mechanism is not accounted for in this framework and the technique is most applicable in cases where this error is small or not present. 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.

Acknowledgement

This research has been generously supported by the *EPSRC Centre for Doctoral Training in Nuclear Fission - Next Generation Nuclear* which is gratefully acknowledged by the authors. The authors are grateful for programming assistance from Dr. Marco de Angelis.

References

[1] J. Carter. "Using Bayesian statistics to capture the effects of modelling errors in inverse problems". In: *Mathematical geology* 36.2 (2004), pp. 187–216.

[2] L. G. Crespo, S. P. Kenny, and D. P. Giesy. "Interval predictor models with a linear parameter dependency". In: *Journal of Verification, Validation and Uncertainty Quantification* 1.1 (2016).

[3] L. G. Crespo et al. "Application of Interval Predictor Models to Space Radiation Shielding". In: *18th AIAA Non-Deterministic Approaches Conference*. American Institute of Aeronautics and Astronautics, 2016. URL: http://arc.aiaa.org/doi/abs/10.2514/6.2016-0431 (visited on 02/17/2016).

[4] E. Patelli. "Handbook of Uncertainty Quantification". In: ed. by R. Ghanem, D. Higdon, and H. Owhadi. Cham: Springer International Publishing, 2016. Chap. COSSAN: A Multidisciplinary Software Suite for Uncertainty Quantification and Risk Management, pp. 1–69. ISBN: 978-3-319-11259-6. DOI: 10.1007/978-3-319-11259-6_59-1. URL: http://dx.doi.org/10.1007/978-3-319-11259-6_59-1.

[5] E. Patelli et al. "OpenCossan: An efficient open tool for dealing with epistemic and aleatory uncertainties". In: *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*. ASCE, pp. 2564–2573.

[6] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2004. ISBN: 0898715725.

[7] Z. Tavassoli, J. N. Carter, and P. R. King. "An analysis of history matching errors". en. In: *Computational Geosciences* 9.2-3 (Nov. 2005), pp. 99–123. ISSN: 1420-0597, 1573-1499. DOI: 10.1007/s10596-005-9001-7. URL: http://link.springer.com/article/10.1007/s10596-005-9001-7.