

Bayesian calibration of fluvial flood models for risk analysis

Lucy J. Manning

Thesis submitted for the degree of Doctor of Philosophy



School of Civil Engineering and Geosciences

Newcastle University

January 2011

Abstract

Flood risk analysis is now fundamental to flood management decision making. It relies on the use of computer models to estimate flood depths for given hydrological conditions. The correct calculation of risks associated with different management options requires that the uncertainty in the computer model output is carefully estimated. There are several sources of uncertainty in flood models, including structural uncertainties in the model representation of reality, uncertainty in model parameters, and observation errors. We refer to the first of these as “model inadequacy”. The work described in this thesis concerns the calibration of computer models to describe fluvial flooding, taking into account model inadequacy and paying particular attention to the requirements of risk analysis calculations.

A methodology which has had some success in other application areas is Bayesian model calibration, using Gaussian process representation both for the error arising from model inadequacy, and to emulate the computer model output. The effectiveness of this methodology is demonstrated for steady state flood models, both of a series of laboratory experiments, and of a historical flood using a satellite image of flood outline for calibration. Extension of the methodology to calibration of dynamic models using gauged data is not straightforward, but is achieved for flood models by means of an emulator, which replaces the computationally expensive hydrodynamic model with a time-dependent transfer function. This permits calibrated prediction of floods using historical gauged data, both in the existing channel and after modelling potential modifications to the channel. It is shown that calibration without inclusion of a model inadequacy function cannot match measured data. Finally, application of the methodology is demonstrated in the context of a calculation of probability of inundation in the channel, both with and without modification.

Acknowledgements

Few endeavours can be truly described as solitary, and this has been no exception. I am grateful for help from the following:

Jim Hall shared his interest with me in the area of Bayesian model calibration, helped me to shape this project, and provided endless patient encouragement.

Dave Walshaw provided encouragement and practical suggestions, particularly during the first half of the project.

Peter Challenor through his project Probability, Uncertainty, Climate and Modelling hosted a workshop and a six-week research playground in Durham; attendance at these helped me to better understand the statistical context of the project I was undertaking and the use of Gaussian process emulators for computer codes.

Leanna House helped me to get my first Markov chain to converge.

Nigel Wright helped me with modelling the data from the Flood Channel Facility, described in Chapter 4.

Keith Beven provided the data and hydraulic model for the Severn case study used in Chapters 6 and 7.

Peter Young and his research group spent time to discuss with me the use of their software package CAPTAIN to emulate the hydraulic model of the Severn, used in Chapter 6.

Contents

1	Introduction	1
1.1	Risk analysis for decision making in flood defence planning	2
1.2	Flood model calibration	4
1.3	Bayesian statistics	6
1.4	Thesis structure	8
2	Hydrological background	9
2.1	Introduction	9
2.2	Rainfall-runoff models	10
2.3	Hydraulic models	12
2.3.1	Description of mechanisms in floodplain flow	12
2.3.2	Equations representing fluvial flow	12
2.3.2.1	LISFLOOD-FP	14
2.3.2.2	Hec-Ras	15
2.3.3	Data requirements of hydraulic models	15
2.4	Discussion	17
2.5	Characterisation of uncertainty sources for flood models	18
2.5.1	Parametric variability	19
2.5.2	Structural uncertainty	19
2.5.3	Input errors	20
2.6	Requirements of current research	22
3	Bayesian analysis of computer code output	23
3.1	The use of Gaussian processes for interpolation and emulation	24
3.2	The use of Gaussian processes for calibration	26
3.2.1	Model specification when the computer model is expensive to evaluate	27
3.2.2	Integrating out the linear regression	29
3.2.3	Specification of priors	30
3.2.4	Solution of equations	30
3.2.5	Calibrated prediction	30
3.3	Markov chain Monte Carlo	31
3.3.1	Metropolis Hastings algorithm	32
3.3.2	MCMC diagnostics	33

3.3.3	Refinements of the Metropolis-Hastings algorithm	35
3.4	Application: simple algebraic example	37
3.4.1	Identifiability of the solution with sparse data	39
3.4.2	Use of an emulator to represent the computer program	42
3.4.3	Experimental design	42
3.4.4	Effect of the choice of regression basis on solution behaviour	44
3.5	Summary	46
3.6	Appendix: Use of BACCO to estimate a Gaussian process emulator	46
4	Calibration of steady state laboratory experiments	48
4.1	Introduction	48
4.2	Experimental data	48
4.3	Hydraulic model	49
4.4	Stage discharge relationship with varying floodplain width	50
4.4.1	Variation of flow with depth	51
4.4.2	Variation of flow with depth and floodplain half-width	53
4.4.3	Variation of flow with depth, floodplain half-width, and channel side slope	55
4.4.4	Comment	57
4.5	Taking into account experimental errors	57
4.5.1	Observation errors in the experimental programme	57
4.5.2	Including input errors in the calibration	58
4.6	Summary	59
5	Calibration of a steady state flood model	60
5.1	Introduction	60
5.2	Flood model	61
5.3	Calibration data	62
5.4	Calibration	63
5.4.1	Emulator construction	63
5.4.2	Calibration	66
5.5	Comparison of results with previous work	70
5.6	Alternative emulator formulations for multivariate model output	71
5.7	Summary	73
6	Calibration of a dynamic flood model	74
6.1	Introduction	74
6.2	Calibration of dynamic models	75
6.3	Extensions of the Kennedy and O'Hagan methodology to time-varying prob- lems	77
6.4	Flood model	78
6.5	Calibration	79
6.5.1	Emulator construction	79

6.5.1.1	Gaussian process emulator, with calibration in the time domain	84
6.5.1.2	Gaussian process emulator for the spline knot values, with calibration in the transformed domain	85
6.5.1.3	Spline emulator, with calibration in the transformed domain	89
6.5.2	Calibration	89
6.5.2.1	Calibration with Gaussian process emulator	91
6.5.2.2	Spline emulator	93
6.6	Discussion	95
6.6.1	Stability of the emulator formulation to spline knot position and to lag value	96
6.6.2	Comparison with Bayesian calibration, without model inadequacy .	97
6.7	Channel modification	98
6.8	Summary	101
7	Use of calibrated prediction in calculating probability of inundation	103
7.1	Introduction	103
7.2	Flood frequency curve	104
7.3	Synthetic upstream data	105
7.3.1	Selection of appropriate window length for flood peak	106
7.3.2	Parameterisation of peak shape	109
7.3.3	Joint distributions for peak characteristic parameters	110
7.4	Probability of inundation	113
7.5	Sensitivity analysis	115
7.6	Summary	118
8	Conclusions and recommendations for further work	120
8.1	Introduction	120
8.1.1	Analytical model	121
8.1.2	Steady-state flood model	121
8.1.3	Dynamic flood model	122
8.1.4	Risk calculation	123
8.1.5	General comments	123
8.2	Recommendations for further work	124
	References	126

List of Figures

1.1	Data dependency model (after Huard and Mailhot, 2006)	7
3.1	Illustration of a Gaussian process, with given values of ω and σ^2 , and known mean function, conditional on given data points.	25
3.2	Estimated Gaussian approximation to given data points	26
3.3	Example chains to show MCMC problems	35
3.4	Example correlations corresponding to the chains in Figure 3.3: Upper right, pairwise correlation; Diagonal, individual variable density; Lower left, 2-dimensional density plots	35
3.5	Reference Gaussian processes for use in determining prior distributions for “roughness” values: three realisations of each.	37
3.6	Chains for toy example	38
3.7	Correlations corresponding to the chains in Figure 3.6: Upper right, pairwise correlation; Diagonal, individual variable density; Lower left, 2-dimensional density plots	39
3.8	Prior and posterior distributions for parameter θ	40
3.9	Calibrated prediction of algebraic example, for a) 10 data points and b) 20 data points	40
3.10	Greyscale plot of the likelihood over a 3-dimensional grid of hyperparameter values	41
3.11	Testing the code emulator: a) contour map of the emulator mean response surface; solid points are where the code has been run, circles are points where the emulator accuracy has been tested, b) 95% prediction interval for the emulator at locations labelled in a (central points are true code output values)	43
3.12	Effect of different emulator and model inadequacy regression bases on the calibrated prediction in extrapolation: a) calibration only, model inadequacy basis (x); b) emulator basis $(1, x, \theta)$, model inadequacy basis (x); c) emulator basis (1) , model inadequacy basis (x); d) emulator basis $(1, x, \theta)$, model inadequacy basis (x^3)	45
4.1	View of the Flood Channel Facility experimental setup.	49
4.2	Geometry of the Flood Channel Facility experimental setup for half channel width.	50

4.3	Comparison of measured data with model predictions: a) least-squares best fit of Manning’s equation for different floodplain widths b) error in measured data, by comparison with least squares fit to inbank data	51
4.4	Bayesian calibrated prediction using data for a single floodplain width using a model inadequacy: a) calibrated prediction; b) difference between calibrated prediction and Manning’s equation, fitted to the in-bank data.	53
4.5	Calibrated prediction: calibration undertaken using data from two floodplain widths, and comparing the calibrated prediction with the third. Validation dataset a) $b_w=3.15\text{m}$ b) $b_w=5\text{m}$	55
4.6	FCF straight channel experimental design	55
4.7	Calibrated prediction for floodplain half-width $b_w=3\text{m}$., comparing the effect of in-bank covariance structures in Equations: a_i) and a_{ii}) (4.3a) and b) (4.3b)	57
5.1	Digital elevation map of the study region, with river course superimposed; direction of flow, left to right	61
5.2	Flood extent, obtained from processed SAR data	62
5.3	Modelled flood extent for different values of Manning’s n	63
5.4	LISFLOOD calibration and emulation data a) Flood water elevation inferred from the SAR flood outline image b) Points where flood elevation values were extracted from SAR observations c) Typical LISFLOOD water surface profiles, at different values of Manning’s n d) Points used to construct the LISFLOOD emulator	64
5.5	Absolute error in mean emulator prediction, and standard deviation of the predicted emulator uncertainty for three different values of Manning’s n : close to, in between, and away from the training runs. Values of test runs and training runs are illustrated	65
5.6	Deviation of model output from linear slope	65
5.7	Emulator error distribution; with superimposed t distribution with 5 degrees of freedom.	66
5.8	Prediction of water levels along the measured shorelines, with different prior assumptions about measurement error, and numbers of data points: a) mean error 35cm, precise distribution, b) mean error 25cm, precise distribution, c) mean error 10cm, precise distribution, d) mean error 25cm, vague distribution, i) 26 data points, ii) 61 data points	69
5.9	Probability of inundation	70
5.10	Probability of inundation obtained by Aronica et al. (2002)	71
6.1	Digital elevation map of the Severn catchment in the area around Shrewsbury, with the river course superimposed, and gauging stations marked in red. Modelled flood relief channel in Shrewsbury marked with dotted line.	79

6.2	a) Gauged upstream stage at Montford, 15th Jan - 7th Mar 2002, b) Output of Hec-Ras model, with input of gauged upstream stage, for different values of Manning's n , compared with gauged downstream stage at Welsh Bridge in Shrewsbury, showing that no value of the parameter will allow the model output to correspond with the data	80
6.3	Heuristic motivation for emulator strategy: a) input-output plot, when output is determined by Equation (6.2); b) input-output plot, when output is determined by Equation (6.3)	81
6.4	Input-output plots: a) downstream v. upstream gauged stage (relative to local data- 47mAOD and 52mAOD respectively); b) downstream stage v. transformed upstream gauged stage	81
6.5	Schematic of identification procedure for nonlinear transfer function	82
6.6	Components of nonlinear transfer function for computer output with different values of Manning's n and for data: a) nonlinear function values, $\frac{b(\cdot)}{(1-a)}$, and b) autoregressive coefficient, a	83
6.7	Comparison with hydraulic model output of time series recovered from nonlinear functions and autoregressive coefficients applied to the upstream hydrographs, for a) time period (15th January to 7th March 2002) used for estimation and b) validation time period (25th October - 15th December, 2002).	83
6.8	Predicted a) nonlinear functions and b) autoregressive coefficients for values of Manning's n between those used to construct the emulator, compared with values obtained directly from the hydraulic model output	87
6.9	Predicted time series for values of Manning's n away from those used to construct the emulator, compared with hydraulic model output for a) original time period (15th January to 7th March 2002) and b) validation time period (25th October - 15th December, 2002).	88
6.10	95% prediction interval for Manning's $n = 0.045$, compared with the slope of the predicted stage for a) original time period (15th January to 7th March 2002) and b) validation time period (25th October - 15th December, 2002).	89
6.11	Projected a) nonlinear functions and b) autoregressive coefficients, for values of Manning's n between those used to construct the spline, compared with values obtained directly from the hydraulic model output	90
6.12	Projected time series for values of Manning's n away from those used to construct the spline, compared with hydraulic model output for a) original time period (15th January to 7th March 2002) and b) validation time period (25th October - 15th December, 2002).	90
6.13	Calibrated prediction, using first calibration method, for a) nonlinear function, b) autoregressive coefficient a and c) predictive distribution for Manning's n	93

6.14	Calibrated prediction for stage at Welsh Bridge in Shrewsbury, using first calibration method, compared with model predictions and observed data: a) time period used in calibration, b) validation time period	93
6.15	Calibrated prediction, using second calibration method, for a) nonlinear function, b) autoregressive coefficient a and c) predictive distribution for Manning's n	95
6.16	Calibrated prediction for stage at Welsh Bridge in Shrewsbury, using second calibration method, compared with model predictions and observed data: a) time period used in calibration, b) validation time period	96
6.17	Posterior distribution of parameter Manning's n , found with simple calibration method	97
6.18	Calibrated prediction for stage at Welsh Bridge in Shrewsbury, using simple calibration method, compared with model predictions and observed data: a) time period used in calibration, b) validation time period	98
6.19	Calibrated prediction for a) original river channel, b) channel modified with triangular relief channel and c) channel modified with rectangular relief channel: i) nonlinear function and ii) autoregressive coefficient	100
6.20	Calibrated prediction for a) original river channel, b) channel modified with triangular relief channel and c) channel modified with rectangular relief channel	101
7.1	Variation of flood height with return period at Montford, as predicted by fitted Generalised Extreme Value model	106
7.2	Flood height at Montford, at annual maximum peaks (scaled by maximum height), for a window of 10 days before and after the peak	107
7.3	Scaled data for annual maximum peaks in 1960 and 1995, compared with synthetic waveforms used for sensitivity analysis	108
7.4	Illustration of the truncation of the flood peak waveform used in the sensitivity analysis	108
7.5	Scaled flood height at Montford, at annual maximum peaks, for a window of 36 hours before and 24 hours after the peak. Hydrograph model superimposed.	110
7.6	Correlation of transformed variables describing the flood peak shape at Montford. Darker shades correspond to higher peaks.	112
7.7	Choice of covariance matrix for correlated variables.	112
7.8	Correlation of simulated transformed variables describing the flood peak shape at Montford. Darker shades correspond to higher peaks.	113
7.9	Probability of issuing a flood warning at Welsh Bridge in Shrewsbury, for unmodified and modified channels.	115

List of Tables

3.1	Prior ranges for MCMC example	34
3.2	Prior and posterior distributions for calibration	38
3.3	Parameters affecting the posterior credible interval	41
4.1	Prior and posterior distributions for calibration with respect to depth . . .	52
4.2	Prior and posterior distributions for calibration with respect to depth and floodplain half-width	54
4.3	Prior and posterior distributions for calibration with respect to depth, flood- plain half-width and channel side slope tangent	56
5.1	Prior and posterior distributions for calibration	67
5.2	Variance explained by different numbers of principal components for a flood model emulator	72
6.1	Prior and posterior distributions for emulation	86
6.2	Prior and posterior distributions for calibration (first method)	92
6.3	Prior and posterior distributions for calibration (second method)	94
6.4	Lag found in best ARX(1,0) model applied to Hec-Ras output	97
7.1	Parameter estimates for Generalised Extreme Value function fitted to his- torical annual maxima of flood height at Montford	105
7.2	Maximum difference in quantile peak height (m), for truncated compared with untruncated upstream waveform	109
7.3	Nature of marginal distributions for variables describing flood peak shapes .	111
7.4	Annual probability of issuing flood warnings, before and after channel mod- ification	115
7.5	First order sensitivity analysis	117
7.6	First order sensitivity analysis, conditional on upstream peak height	119

Chapter 1

Introduction

Since 1900, floods have affected 28.6 million people per year around the globe, causing an average 63,500 deaths per year, and giving rise to annual economic damage of US\$4bn. However, within the last 30 years, the annual death rate due to flooding has been 7,000, while the mean number affected annually has been 94 million, with annual costs of US\$14bn. Natural disasters from other causes show a similar pattern when a comparison is made between these time scales, that is the approximate tripling of lives affected annually and of annual costs, while the numbers of lives lost has reduced, or at worst remained constant (International Disaster Database, 2010). While the costs reported do not take inflation into account, an increase in cost has been highlighted by the insurance industry; costs of extreme weather events around the world have doubled each decade since the 1970s (Coomber, 2006).

Indeed, there is a perception, arising from a number of factors, that flood risk is increasing. There is some concern that flooding may increase as a result of the increased rainfall intensity associated with a rise in global temperature. In addition, development of floodplains not only increases the property vulnerable to flooding, but inhibits drainage within the floodplain, increasing the proportion of rainfall contributing to surface flow during heavy storms. Flood defences for such developments may also increase the risk to property damage by restricting the area of the floodplain available for the river in times of high flow.

In England, the Environment Agency estimate that 2.4 million properties are at risk of flooding from rivers and the sea, with annual economic damage estimated at £1billion (Environment Agency, 2009a). In addition, substantial infrastructure services are in areas at risk of flooding. Current annual spending on flood defences is £800million. In this context, there is a need for reliable flood prediction not only for the improvement of flood warnings, but also for for appropriate allocation of such resources. The execution of both of these employs computer models.

Floods are routinely predicted using computer models, both for operational warnings, and for risk analysis for planning, design and asset management. The requirements of a flood

modelling system depends on the end use. The emphasis for operational warnings is on speed for real time prediction. Real-time flood forecasting models can be either based on a physical process description or purely on historical data analysis, will start from a known state, and can be updated during operation by assimilation of weather and flood severity data as they become available. Their efficacy is easily assessed.

Conversely, flood models for risk analysis need to be largely physical process based, as such models take as a starting point an unknown state, and may be run for conditions substantially different to those for which the model has been calibrated. While the representation for forecasting involves predicting flood levels from weather conditions, a greater diversity of modelling processes may be used for risk analysis, depending on the emphasis of the study. A risk analysis study may examine flood levels predicted from weather conditions, or from upstream flow conditions. It may also take into account the effect of climate change on the weather conditions. Flood severity data for such an application are historical, and incorporated during calibration, and models do not need to be run in real time. However, since the results concern an integral over the probability distribution of an event and its consequences, assessment of the accuracy of such a study could only be made by observation, under unchanged conditions, over an extended period of time, and is thus difficult to achieve.

All computer models are limited in the accuracy of their predictions by the extent of scientific understanding, by the complexity that it is feasible to implement in a model, and by the data available to input into the modelling system. Uncertainties in computer model output are thus inevitable, and it is important to be able to assess accurately the uncertainty involved in the modelling process, whatever the purpose of a flood modelling study. These uncertainties should be expressed as a probability distribution, since what is acted upon is not generally the mean prediction, but one of its quantiles; an evacuation warning will be issued if the probability of flooding is greater than some threshold, and the probability distribution of a particular flood level being achieved is incorporated into a risk analysis.

1.1 Risk analysis for decision making in flood defence planning

Risk analysis permits a direct comparison between the expected damage and the expenditure involved in different flood protection strategies, thus providing a rational basis for flood management decision-making and allocation of resources. Indeed, both the UK government and the European Union require the drawing up of flood risk management plans for all areas of significant flood risk (European Union, 2007, Office of Public Sector Information, 2009, 2010), with a view to reducing the risk of adverse consequences according to appropriate objectives.

During a period of high river flow, flood defence failure may occur, either through over-

topping or through structural failure. Inundation of the floodplain will then result in damage to property and life, depending on the water depth and possibly the flow rate in the floodplain. In mathematical terms, if the water flow rate at some point in the river is denoted Q , the event that the flood defence fails is ϕ and financial estimate of the damage is denoted $c(h)$, dependent on floodplain water height h , then the expected damage is

$$E(c) = \int c(h) f(h|\phi, Q) p(\phi|Q) f(Q) dQ \quad (1.1)$$

where $f(h|\phi, Q)$ is the probability density of floodplain water height h , conditional both on flood defence failure ϕ and on river flow rate Q , $p(\phi|Q)$ is the probability of flood defence failure, conditional on Q , and $f(Q)$ is the probability density of Q . This distribution is to be determined from available information, which includes historical field observations and computer models.

The construction of a damage function $c(h)$ is a complex problem, bringing together, as it must, a wide range of consequences of inundation, including those for human health and life, the environment, cultural heritage, economic activity and infrastructure. In this study, the simplified approach is taken that the cost depends only on the exceedance of a particular water height h_0 :

$$\begin{aligned} c(h) &= 1 & h > h_0 \\ &= 0 & h \leq h_0 \end{aligned}$$

Similarly, the failure of flood defences involves detailed structural complexities and uncertainties, with weaknesses not able to be monitored, and deterioration mechanisms which are not fully understood. As with the damage function, in this study, the complexities of flood defence failure are by-passed, assuming that the conditional probability $p(\phi|Q) = 1$, and consequently that $f(h|\phi, Q)$ becomes $f(h|Q)$.

Taking into account these simplifications, the expected damage is reduced to a probability of inundation,

$$P(h > h_0) = \int I(h > h_0) f(h|Q) f(Q) dQ \quad (1.2)$$

where $I(\cdot)$ is the indicator function, taking a value of 1 for positive argument, and 0 otherwise.

The distribution $f(h|Q)$ is generally provided by a computer model. At its simplest, if Q is taken to refer to the same location as h , it is the inverse of the rating curve, which is usually expressed as a simple deterministic relationship, fitted to historical data. Such relationships are well known to involve uncertainties, but are nonetheless routinely used for estimating flow when only water height has been measured, for example. It is more usual to regard Q as the flow at some upstream location. Since, however, evaluation of the integral involves a knowledge of the distribution of the independent variable of integration, the integral could be formed in terms of upstream water height, or of catchment rainfall.

The probability distribution $f(h|Q)$ depends on a number of influences other than Q , and

correct evaluation of the integral may depend on identifying these, characterising their distributions and the dependence of h on them, and then integrating over their range. Besides Q , a part is played by the antecedent conditions, the capacity of the channel and the volume of inflow from lateral channels. In addition, it is necessary to estimate the probability distribution of the upstream condition Q from the available information, which includes both measured historical data and computer models. Both involve uncertainties, which need to be taken into account. Data may only exist for a small number of events; thus it is difficult to account for the full range of possible input conditions. In addition to limited coverage, data may be subject to measurement error. Many parts of the process are not directly observable, leading to substantial uncertainties in modelling and prediction.

Uncertainties can be classified as aleatory or epistemic. Aleatory uncertainty, or natural variability, is random, and is relatively straightforward to deal with; this can be handled by increasing the number of points sampled in the evaluation of the integral (1.2). By contrast, epistemic or knowledge uncertainty is much harder to detect and eliminate, as this requires identification of its source within the model used to formulate the integral for the probability of inundation. The problem is normally compounded in the calibration of the model; by simply minimising the errors in model fit to the historical data, model inadequacies are compensated by parameter bias, potentially weakening the predictive capacity of the model.

Hall et al. (2003) listed a hierarchy of flood risk assessment approaches, to support a range of flood risk management decisions. These range from a national flood risk assessment, based on the standard of protection and the flood frequency curve, such as the one outlined in their paper, to detailed local studies, using continuous hydraulic modelling. Naturally, the more detailed the study, the more opportunity exists to examine the modelling uncertainties involved. However, of those studies reported in the scientific literature (Dawson et al., 2005; Apel et al., 2004, 2008; Di Baldassari et al., 2009), none examine the uncertainty arising from the flood model itself, nor its calibration. Uncertainties exist nonetheless, both in the structure of the model itself, and in the values of any parameters used to tune the model for the river under examination.

1.2 Flood model calibration

Calibration is the process of finding the appropriate values of parameters for a computer model. In some cases, these parameters represent physical quantities, and can be found by independent measurements, but it often happens that the measurements cannot be made at the appropriate scale, and thus cannot be used, or that parameters do not represent physical quantities. Parameters must then be inferred by comparison of the model output with historical data.

Historically, the values of such parameters were found manually by running the model with

different parameters, and minimising some measure of fit of the model output to the data. Naturally, different measures of fit would lead to different parameter values. Automation of the process involves a numerical optimisation, but such techniques do not give rise to probability distributions required by Equation (1.2). In addition, it has frequently been found that a large number of parameterisations may lead to equally good model fit. This implies that the choice of a single parameterisation provided by optimisation of this fit is not the most appropriate means of model choice, as there is justification for considering a range of values for each parameter.

These last two issues are addressed by Bayesian calibration, since the parameters themselves are treated as variables with their own distributions, thus placing an emphasis on quantifying uncertainties, rather than maximising model fit. Bayesian analysis combines in a rigorous way the prior beliefs of the user, the statistical model for the data, and the data values themselves to lead to a distribution which may be used directly in Equation (1.2) .

The most widely used method for the parameterisation of flood models is undoubtedly the Generalised Likelihood Uncertainty Estimation (GLUE) method of Beven and Binley (1992). This is equivalent to the Approximate Bayesian Computation method which is currently receiving much interest from statistical and biological modellers (e.g. Beaumont et al., 2002; Gaggiotti, 2010). Starting from a prior joint parameter distribution, usually taken to be uniform uncorrelated univariate distributions on a limited range, the model is run a large number of times, randomly sampling from these parameter ranges. The model is evaluated by some convenient measure of fit, the generalised likelihood, and parameterisations leading to “non-behavioural” performance are discarded, according to a criterion determined by the user. A distribution for the model output is then generated by taking a weighted mean of the model outputs, the weights consisting of the measure of agreement found, according to the generalised likelihood, for each parameterisation. While the appropriate choice of likelihood the GLUE method can give rise to a rigorous parameterisation, such as that demonstrated by Romanowicz et al. (1996), the relaxation of the requirement to use such a carefully chosen likelihood, depending rigorously on a plausible statistical data model, has both ensured the popularity of the method, and attracted criticism from a number of authors (Montanari, 2005; Ewen et al, 2006; Mantovan and Todini, 2006; Stedinger et al., 2008), on the grounds of arbitrariness and inconsistency of the result.

It is important to consider the different error sources in the data model. Ewen et al. (2006) showed by comparing the output from a reduced model to that of a more detailed model, that oversimplifying some mechanisms leads to biased parameter outputs. Much of the literature on model calibration does not take into account different error sources, and is thus flawed. This shortcoming also applies to the GLUE method.

Other more rigorous calibration methods have been suggested to overcome these deficiencies, in particular for models with time-series input and output, with a view to forecasting. However, for risk analysis, and for modelling of flood extent using aerial or satellite images,

to date no adequate method has been developed.

In this context, a Bayesian calibration is required for flood inundation models, taking into account all the sources of possible error. The method should be statistically rigorous, should take into account the various sources of uncertainty in model output, and to be appropriate for risk analysis. In addition, it should be able to be used with data arising from river gauging, and from flood extent snapshots. This thesis concerns the development of such a method.

1.3 Bayesian statistics

Statistical inference is the process of drawing conclusions about a system from a data sample randomly drawn from a population of some quantity of the system under investigation. The data are generally analysed with reference to a statistical model of the assumed distribution of the population. The likelihood is defined as the distribution of the observed data z , under the assumption of the statistical model, dependent on parameters θ .

$$\mathcal{L} = f(z; \theta)$$

Classically, characterisation of a modelled population is performed by maximisation of the likelihood with respect to the model parameters θ . The Bayesian method is founded on the concept that the distribution of a variable representing the quantity under investigation is characterised not only by the available data, but also by the beliefs of the modeller about the parameters of the distribution, based on his prior experience. In this case, the parameters, instead of being assumed to have distinct values, are themselves considered to have a distribution.

Thus, if the user's belief is that the parameters θ of the system have a distribution $f(\theta)$ (the prior distribution), and the data z have a distribution $f(z|\theta)$, defined by the statistical model and conditional upon the parameters θ , then the conditional probability is given by

$$f(z|\theta) = \frac{f(z, \theta)}{f(\theta)}$$

Equally, the probability distribution of the parameters, conditional on the data, can be expressed as

$$f(\theta|z) = \frac{f(z, \theta)}{f(z)}$$

Bayes' theorem gives the probability of the parameters, conditional on the data, as follows:

$$\begin{aligned} f(\theta|z) &= \frac{f(z|\theta)f(\theta)}{f(z)} \\ &= \frac{f(z|\theta)f(\theta)}{\int (f(z|\theta)f(\theta))d\theta} \end{aligned} \tag{1.3}$$

The distribution $f(\theta|z)$ is known as the posterior distribution, and represents the synthesis between the user's prior beliefs $f(\theta)$, the statistical model $f(z|\theta)$, and the data. The use of Bayes' theorem provides a means of updating the user's beliefs about the system in the light of the data, and can be used recursively, as more data become available.

The distribution of the data has been described above by a model $f(z|\theta)$. However, the parameters θ may be dependent on secondary parameters, ϕ , with a conditional model $f(\theta|\phi)$. This is known as a hierarchical model. The conditionality could be extended, with ϕ dependent on further parameters, ψ . We then say that the distribution of parameters at a given level is conditional on those at the next lower level, and given that conditionality, is independent of parameters lower than that. We can thus specify the joint distribution of the data and parameters as:

$$f(z, \theta, \phi) = f(z|\theta) f(\theta|\phi) f(\phi)$$

This construct permits us to consider the dependence of our data on both model output, and different error sources; thus Huard and Mailhot (2006) described the sources of error in modelled streamflow data using a physically-based model M , by the diagram in Figure 1.1, and were able to construct a hierarchical model to represent the measured data. The data model shown in Figure 1.1 is described by the relationship

$$z = M(w + e, \theta) + \delta + \epsilon \quad (1.4)$$

where the observed data z are dependent on output of the model $M(w + e, \theta)$ at input $w + e$ with parameters θ , subject to model structural error δ and observation error ϵ . The research described in this thesis has been undertaken in the context of the model described in Equation (1.4).

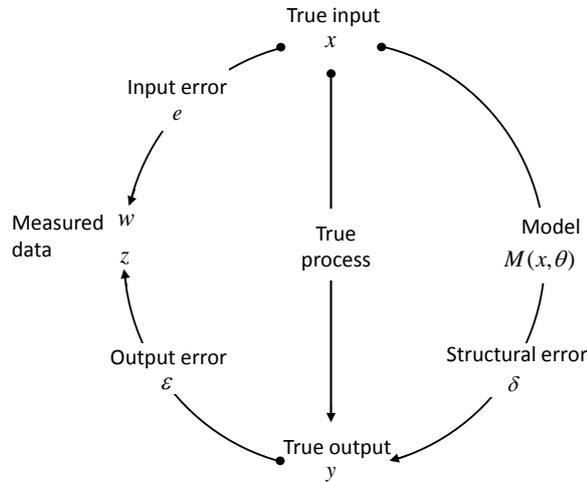


Figure 1.1: Data dependency model (after Huard and Mailhot, 2006)

Identification of the three unknown quantities, input error e , structural error δ and output error ϵ at each time step is a difficult problem, for which no solution has yet been achieved. It relies on the ability to define a statistical distinction between them, since without this

these unknown quantities cannot be individually identified. However, it is possible to achieve a partial solution to this problem by taking a simplified data model.

1.4 Thesis structure

This chapter has given an overview of the motivation, aims, and statistical context of the research undertaken. The research is described in further chapters, as follows.

Background to the research is provided in Chapter 2, which describes the models used in flood prediction, and the simplifications that are necessary to make their use practicable, followed by a review of the relevant literature on calibration of hydrodynamic models.

The methodology used is described in Chapter 3, followed by an example of its application to an algebraic example. Chapters 4, 5, and 6 describe the application of the methodology to examples of increasing complexity. These are a series of one-dimensional steady-state laboratory experiments (Chapter 4), a steady-state model of a flood extent on the Thames at Buscot, calibrated with satellite data (Chapter 5), and a dynamic model of the Severn in the region of Shrewsbury, calibrated with gauged river stage data (Chapter 6).

Chapter 7 illustrates the application of the methodology in the comparison of the impact of two hypothetical flood risk management schemes for the example on the Severn. The thesis concludes with Chapter 8, giving an indication of the success of the methodology with respect to the research aims, and suggesting avenues for extension of the investigation.

Chapter 2

Hydrological background

2.1 Introduction

The probability of inundation described in the Introduction (Equation (1.2)) is

$$P(h > h_0) = \int I(h > h_0) f(h|Q) f(Q) dQ .$$

As described in the Introduction, the probability integral may be formulated and evaluated in terms of upstream flow, water height or catchment rainfall. The term $f(h|Q)$, or its equivalent, according to the input conditions used, gives downstream water height conditional on upstream conditions, encapsulating the information to be derived both from hydrological and hydraulic models and from measured data. Its evaluation requires an understanding of the hydrological mechanisms being modelled, the types of models being used and their limitations, and the limitations of the data used, as well as a sound understanding of the statistical principles and techniques required. This chapter addresses not only the area of hydrological and hydraulic modelling, providing a background to the processes involved and the models used to describe them, but also the literature on hydrological model calibration, highlighting the issues to be considered.

Flood production from precipitation involves a number of processes. These are non-linear and their interaction complex. The dynamics and spatial distribution of runoff generation depend on the hydraulic properties and storage capacities of highly heterogeneous natural materials, including vegetation, soils and rocks. Runoff is driven primarily by gravity and precipitation, modified by the mechanisms of evapotranspiration and movement of water through porous media. Local surface or subsurface flow occurs when the volume of water exceeds the local storage capacity, and connectivity of areas of local flow can lead to larger-scale flow. Lower in the catchment, the flow dynamics within the channel may make a more significant contribution to total water transport than the runoff generating process.

The models describing the mechanisms of runoff production and channel flow are very

different. Runoff generation is described by rainfall-runoff, or hydrological models, while the flow dynamics are described by routing, or hydraulic models. It is often the case that the flow in the upper part of a catchment is described by a hydrological model, while the flow lower in the catchment is described by a hydraulic model, although hydrological models may account for the entire catchment.

While the work in this thesis concerns the calibration of hydraulic models, a large part of the literature on model calibration concerns hydrological models, so a brief description follows of the types of hydrological model. Following that, a description is given of the known behaviour of channel and floodplain flow, and an overview of the models used to describe these. The chapter concludes with a review of the relevant literature on calibration of hydrological and hydraulic models.

2.2 Rainfall-runoff models

Freeze and Harlan (1969) suggested a blueprint for a physically based model of hydrological catchment response to precipitation, including interception and evapotranspiration, infiltration and soil moisture flow, groundwater, overland and channel flow. While such models have been generated, they are costly to run, and require extensive data inputs. It is thus often not realistic to represent the movement of water through a catchment in such a detailed manner.

In spite of this blueprint, there is no consensus about the best way to achieve a simple and effective description of the flow of subsurface water. This is partly because there are many mechanisms at work, whose relative significance will differ according to the catchment or part catchment, and according to the flow regime. The other major difficulty in describing runoff formation is the lack of detailed information about soil structure and behaviour.

One of the simplest rainfall-runoff models is the Nash cascade model (Nash, 1959), which treats the catchment as a series of equally-sized reservoirs. The unknowns in the model are the number of reservoirs and their total capacity. Since the model is not really a physical description of the catchment behaviour, it may be extended to have a non-integer number of reservoirs.

A more sophisticated type of model is the lumped conceptual catchment model. With this type of model, the catchment is described as a single, aggregated entity, but different mechanisms of water storage are described as reservoirs; hence there may be reservoirs representing the water held in the canopy, the shallow soil, and deeper water storage. Deterministic differential equations are used to represent the transfer of water between these stores, as well as evaporation and runoff. While these relationships represent physical concepts, there is no attempt to relate them to spatial locations within the catchment, and the parameters used in the model do not represent actual physical quantities; thus values must be found by calibration.

Fiering developed a simple linear rainfall-runoff model for teaching purposes (Fiering, 1967); this has been used to demonstrate a number of calibration schemes, as it is a useful example of the type of equations used in lumped catchment models. Fiering's "abc" model is as follows:

$$\begin{aligned}Q_t &= (1 - a - b)r_t + cS_t \\S_{t+1} &= (1 - c)S_t + ar_t\end{aligned}$$

where r is the input rainfall volume, Q and S represent discharge and storage volume respectively with discrete time index t , and parameters a , b and c are the proportions of rainfall entering storage, and lost to evapotranspiration, and the proportion of water leaving storage. The parameters a , b , and c are to be determined by calibration, subject to the mass conservation condition $a + b \leq 1$.

There are many different conceptual rainfall-runoff models in use; modelling success has been achieved by the incorporation of nonlinearities resulting from the saturation of increasing area of the catchment, with this area being described probabilistically (Moore and Clarke, 1981) or deterministically (Wood et al., 1992; Zhao, 1992; Todini, 1996), treating the catchment as a collection of subcatchments, and solving simultaneously for the runoff formation in each subcatchment, as well as the channel flow.

Sivapalan et al. (2003) contrasted the "bottom-up" and "top-down" approaches to rainfall-runoff modelling. The "bottom-up" approach is that described above, where a concept of the physical processes, inferred from physical understanding and from observation of catchment behaviour, is translated into a system of mathematical equations, requiring fitting of a number of parameters. The "top-down" approach infers the model structure directly from the data, trying to identify the dominant processes at the catchment scale. Young (e.g. 2003) suggested that the "bottom-up" approach leads to a model requiring more parameters to be determined than the data quality can justify, and recommends a "data-based mechanistic" approach, based on a statistical analysis of the input and output data.

The most detailed rainfall-runoff models (Abbott et al., 1986; Ewen et al., 2000; Ciarapica and Todini, 2002) are those which correspond to the original blueprint of Freeze and Harlan (1969) to represent the spatial and temporal movement of surface and subsurface water on a grid covering the entire catchment. Such models can be considered to be truly physically based, but require large amounts of data to define the different characteristics of the catchment properties. While much of this data may be obtained from databases of material properties such as the HOST classification of soil types (Boorman et al., 1995), some parameters represent sub-grid-scale processes, and thus need to be estimated (e.g. Bathurst, 1986).

2.3 Hydraulic models

2.3.1 Description of mechanisms in floodplain flow

Under normal conditions, a river flows within its channel. When the flow is too great to be contained within the channel, it spills onto the floodplain, effectively using a new, broader and more complex channel.

Even simple steady flow in a uniform, prismatic channel may involve three-dimensional processes, as differential resistance leads to different velocities across the channel, and the setting up of secondary flows (Knight and Shiono, 1996). The more complex geometry of the floodplain, and resistance caused by vegetation and other obstacles leads to more complex flow processes, involving different flow velocities in the channel and the floodplain. In an effort to characterise these processes, many laboratory experiments have been conducted, involving measurements of the shear stresses and secondary circulation in steady-state flows in straight or meandering channels with uneven beds. Sellin (1964) described vortices with vertical axes along the edge of the main channel, accounting for the transfer of momentum between the channel and the slower-flowing floodplain. Ervine et al. (1993) described horizontal vortices in meandering channels, initiated immediately below the bend apex. Knight and Shiono (1996) pointed out that as the floodplain depth increases, it effectively becomes a larger channel, and the velocities tend to equalise; the impact of three-dimensional flow processes on the channel conveyance tends to be at its greatest for the ratio of floodplain depth to main channel depth in the range 0.1-0.3.

2.3.2 Equations representing fluvial flow

Treating a fluid as a continuum, and averaging out terms representing small-scale eddies, flow is described by the Navier Stokes equation of motion. Momentum balance for an incompressible fluid gives the following:

$$\rho \left(\frac{\partial u}{\partial t} + u \cdot \nabla u \right) = \rho F - \nabla p + \nabla \cdot T,$$

where u is the velocity at a point, ρ the density, p the pressure, F represents the external forces, and T is the deviatoric stress tensor, representing the internal stresses in the fluid arising from motion (Batchelor, 1967, p142). The terms on the left hand side of the equation describes the change in momentum at a point in space and time.

Conservation of mass gives the continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0$$

where clearly for an incompressible fluid the first term is zero.

These equations are difficult and costly to solve, not least because there is a discontinuity

at the free water surface. For most purposes, acceptable approximations can be made by assuming that the horizontal scale is much larger than the vertical scale, that the vertical velocity component is small, and there is no variation of flow with depth, and by integrating over the water depth. The resulting shallow water equations can be expressed in terms of the flow below a surface of constant pressure in the fluid, such as the free surface.

Making a further simplification, that the flow is predominantly in the longitudinal direction of the channel, so that the velocity is uniform and the water surface horizontal across any cross section perpendicular to the longitudinal axis, we arrive at the one-dimensional Saint Venant equations (e.g. Chow et al., 1988, p281), which are shown below. The continuity equation, representing mass conservation over a unit width is:

$$u \frac{\partial y}{\partial x} + y \frac{\partial u}{\partial x} + \frac{\partial y}{\partial t} = 0$$

where u is the longitudinal velocity, and y the channel depth.

The momentum equation is:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + g \frac{\partial y}{\partial x} - g(S_o - S_f) = 0$$

where in addition S_o is the slope of the channel bottom, S_f the friction slope, g the gravitational constant. A further relation is needed to describe S_f ; this provides a parameter to be determined.

These equations are known as the dynamic wave model, while ignoring the first two terms of the momentum equation, the acceleration terms, leads to the diffusion (or non inertial) wave model. Further simplification, ignoring the third term, the pressure force term, yields the kinematic wave model. Kinematic wave models can make use of empirical equations for steady flow, including the Manning, Chezy, and Darcy-Weisbach equations.

In practice, different simplifications of the Navier Stokes equations are needed to adequately describe the flow in different environments. The kinematic wave model is only suitable for rivers with significant bottom slope ($> 0.1\%$), and overland flow, while the diffusive wave model can include deceleration effects, but does not include a full description of backwater effects or reverse flows. Dynamic wave models are needed for rivers with mild bottom slopes, tidal rivers, and those with reservoirs.

The majority of commercial codes for solving hydraulic equations use a one-dimensional approximation for the channel flow. Two such codes are used in this study, which employ different adaptations of the Saint Venant equations to represent the different flow behaviour in the channel and the floodplain. These are described below.

2.3.2.1 LISFLOOD-FP

LISFLOOD-FP (Bates and de Roo, 2000) was developed to use as simple as possible a representation of flood flow, to enable modelling over an extended floodplain, or to enable sensitivity analysis. Flows in the channel and floodplain are represented by different sets of equations, which are coupled. Channel flow is represented by the one dimensional Saint Venant equations, this time expressed in terms of volumetric flow Q and cross-sectional area A ,

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q \quad (2.1)$$

where q is the flow into the channel from the floodplain or tributary channels, together with a momentum equation (Manning's equation),

$$\begin{aligned} S_o &= S_f \\ S_f - \frac{n^2 P^{4/3} Q^2}{A^{10/3}} &= 0 \end{aligned} \quad (2.2)$$

where as before S_o is the slope of the channel bottom, S_f the friction slope, and in addition n is the friction parameter, Manning's n , and P is the wetted perimeter of the flow.

A rectangular channel cross-section is assumed. Flow over the floodplain is represented by a network of two-dimensional rectangular storage cells, based on the DEM, and where the flow rates between the cells are calculated by the equations:

$$\begin{aligned} \frac{dh^{i,j}}{dt} &= \frac{Q_x^{i-1,j} - Q_x^{i,j} + Q_y^{i,j-1} - Q_y^{i,j}}{\Delta x \Delta y} \\ Q_x^{i,j} &= \frac{h_{flow}^{5/3}}{n} \left(\frac{h^{i-1,j} - h^{i,j}}{\Delta x} \right)^{1/2} \Delta y \end{aligned} \quad (2.3)$$

where

$h^{i,j}$ is the water free surface height at node (i, j)

Δx and Δy are the cell dimensions,

n is the effective grid scale friction parameter, Manning's n , for the floodplain,

Q_x and Q_y are the volumetric flow rates between floodplain cells, with Q_y defined analogously to Q_x , with superfixes $(\cdot)^{i,j}$ referring to the node (i, j) , and

h_{flow} , the flow depth, is the depth through which water can flow between two cells, and is defined as the difference between the highest water free surface in the two cells and the highest bed elevation.

Equations (2.3) embody the assumption that the flood spreading over the floodplain is a function of gravity and topography.

2.3.2.2 Hec-Ras

Hec-Ras is a commercially-used hydraulic modelling package written by the US Army Corps of Engineers (2002). The one dimensional Saint Venant equations of motion are adapted to allow for different conditions in the channel and floodplain as follows:

$$\begin{aligned} \frac{\partial A}{\partial t} + \frac{\partial(\Phi Q)}{\partial x_c} + \frac{\partial[(1 - \Phi)Q]}{\partial x_f} &= 0 \\ \frac{\partial Q}{\partial t} + \frac{\partial(\Phi^2 Q^2/A_c)}{\partial x_c} + \frac{\partial[(1 - \Phi)^2 Q^2/A_f]}{\partial x_f} + gA_c \left[\frac{\partial z}{\partial x_c} + S_{fc} \right] + gA_f \left[\frac{\partial z}{\partial x_f} + S_{ff} \right] &= 0 \end{aligned} \quad (2.4)$$

where

Q is the total flow, as before

Φ is the proportion of the total flow in the channel, $= \frac{K_c}{K_c + K_f}$

K_c is the channel conveyance, (defined as $\frac{Q}{S_{fc}^{1/2}} = \frac{A_c R^{2/3}}{n}$, incorporating Manning's equation)

K_f is the floodplain conveyance,

z is the water surface height, and

cross-sectional area, A , and friction slope, S_f , are distinguished for the channel and floodplain by suffices c and f respectively.

The river path is determined by plan geometry and cross-sections specified by the user across the channel and floodplain. An assumption is made that the water surface is horizontal across the entire cross-section, and the equations are discretised, calculating values at the cross-sections. Distance x_c is measured along the assumed middle line of the channel, under the assumption that the cross-sections are perpendicular to its direction, and x_f is the distance specified between the cross-sections. Cross-sections need to be specified adequately close together to ensure that the river flow is adequately represented by the one-dimensional model.

2.3.3 Data requirements of hydraulic models

Initial conditions are required for dynamic equations, as are boundary conditions at all boundaries. These can be in the form of water level or flow time series. A simplified relationship may be used for the downstream boundary condition, such as Manning's equation:

$$v = \frac{1}{n} R^{2/3} S_0^{1/2}$$

where v is the cross-sectionally averaged velocity, R is the hydraulic radius, normally taken to be the ratio between flow cross-sectional area and the wetted perimeter, S_0 is the channel slope, and n , Manning's roughness coefficient, is the constant to be determined. This boundary condition is an approximation, as Manning's equation is used as a description

of steady flow. It should thus be applied well below the reach where the model results are required.

In addition to the flow conditions, it is necessary to have information both on the river location and cross-sectional geometry, and on the floodplain elevation. Models differ in their precise requirements of geometric data; for example, Hec-Ras requires cross-section data extending over both the channel and the floodplain, while LISFLOOD assumes a rectangular cross-section for the channel, but requires the floodplain to be specified by a rasterised digital elevation map.

In general, the primary measurement of river discharge is stage, or water height, found by recording the level of a float in a stilling well adjacent to the river, by pressure sensing, or by reflection of an ultrasound signal from the river bed. It is much more difficult to determine flow, as this involves estimating the velocity throughout the river cross section, and integrating. Systematic measurement of flow is most easily undertaken at a weir, where the river is constrained, and the bed shape is such that the variation of velocity is well understood. In either case, measurements may be affected by weed growth, or by bypassing of the measurement station in high flow conditions.

Since flow is difficult to measure, and can require installation and maintenance of sophisticated equipment, it is common for flow records to be determined by rating curve, particularly at small measurement stations. This is a deterministic relationship between values of stage and flow. Such relationships are often based on a limited number of measurements, and may require extrapolation, particularly for modelling flood conditions. In addition, the relationship between flow and stage frequently exhibits hysteresis, which would be expected from the hydrodynamic equations, but which is not necessarily reflected in the rating curve.

Measurements of the river cross-section are laborious and time consuming. It is frequently the case that a hydraulic model is limited by the number and spacing of available cross-sections. These cross-sections should also include information about the floodplain, particularly where there are man-made obstacles such as bridges and culverts. It is well known that such structures significantly modify the upstream and downstream water levels, and often the case that the simplified flow equations may not be valid in their neighbourhood, but as their characteristics are often not well known, they are an additional source of model structural uncertainty.

Although many different types of data are used as input into a hydraulic model, in general for a one dimensional model the only parameter which is varied is the roughness parameter, n , although this may be allowed to vary spatially. This parameter is used to account for all the inadequacies of the data and model, as well as roughness of the river bed. It has been shown in laboratory experiments (Ervine et al., 1993) that a significant difference in value of Manning's n is required to account for sinuous channels in a one dimensional flow model.

2.4 Discussion

It is evident from the foregoing description of hydrological and hydraulic models that there are similarities as well as differences between the physical descriptions of rainfall-runoff and channel routing. The differences are most obvious for the greatest spatial aggregation represented in the rainfall runoff models, or indeed for the greatest detail in the representation of the hydraulic models.

One very significant difference is in the way in which the models are used; rainfall-runoff models tend to be used in continuous simulation over extended periods, whereas hydraulic models are usually limited to event-based modelling. Even this difference is not strict; Feyen et al. (2007) reported using a simple hydraulic model in continuous simulation; while large distributed models are limited by computing resources in the modelling period or spatial definition achievable.

In general the number of parameters used in rainfall-runoff models is greater than that used in hydraulic models, partly since the variety of mechanisms being represented is wider, but also because most of the inputs of a hydraulic model are considered fixed, being of a geometric nature. However, while it is in principle possible to measure all the necessary geometric data needed, this may not be realistically practicable, in particular as such characteristics may be subject to morphological change. For both hydrological and hydraulic models, the parameters may well not correspond to measurable entities; thus calibration may be required.

A significant source of error in both hydrological and hydraulic models is the essentially unknown information. In the case of hydrological models, this arises from the accuracy with which rainfall can be determined. The problem arises in hydraulic models, where flow measurements are dependent on a rating curve, and where gauging of lateral inflows is insufficiently accurate, or non-existent.

While both types of models correspond to variable spatial domains, information may not be available in sufficient spatial detail to properly describe the flow mechanisms. Thus, for rainfall-runoff models, it is likely that the availability of spatially varying rainfall measurements and of detailed understanding of spatial subsurface geometry is inadequate. In the case of hydraulic models, geometrical descriptions of the river bed may not be available in adequate detail.

The combination of inadequate data, an inability to model at a sufficiently broad range of scales to capture all the relevant processes, and a lack of clarity as to which mechanisms are dominant, must result in models which are in part inadequate to describe the processes at work. Thus a calibration method is needed which takes into account the fact that the modelling process is likely to be inadequate. In the review of literature on model calibration which follows, an emphasis has been placed on identification of the sources of model output error, and the implications of this for model calibration methodology.

2.5 Characterisation of uncertainty sources for flood models

Ewen et al. (2006) classified the different sources of error in computer code outputs by the stage in the modelling process at which they are introduced:

- model structural error, including conceptual and implementation errors,
- parameter error, including poor parameterisation, errors in parameterisation to compensate for structural deficiency and errors due to inaccurate calibration data, and
- runtime error, from errors in input data and in misuse of the model and misinterpretation of its results.

By contrast, Kennedy and O'Hagan (2001a) listed the different sources of uncertainty in computer code outputs in terms of where they may be introduced in a statistical model of the outputs. Their list is as follows:

- parameter uncertainty,
- model inadequacy, where the model may not perfectly specify the process under consideration,
- residual variability, the variability of the process starting from apparently identical conditions,
- parametric variability, where the model is insufficiently detailed to describe the process with a single parameter value,
- error in the observation of outputs for calibration purposes, and
- code uncertainty, because it may not be feasible to run the code for every possible combination of parameters and input data.

An additional uncertainty, which was largely ignored by Kennedy and O'Hagan, though not by Ewen et al., but is significant in hydrological modelling, is uncertainty in the measurement of input data. At the very least, for hydraulic models, input measurement errors may be expected to be of a similar order to the output measurement errors, having a similar measurement mechanism. However, for hydrological models, the input is rainfall, which is much more difficult to establish precisely over the entire catchment.

All parameterisations deal with output errors; while some studies have noted evidence for heteroscedacity or autocorrelation in the output errors, suggesting transformations or the incorporation of an autoregressive error model (e.g. Sorooshian and Dracup, 1980), the autocorrelation at least may in fact be caused by inattention to other types of error. A number of studies have been undertaken, concentrating on one or other error sources; a review of these is given below. It should be noted that calibration of hydrological and hydraulic models is undertaken for three reasons, to understand the underlying mechanisms of a process, to improve forecasting lead time and uncertainty, and to identify uncertainty

for performing risk analysis. While the purpose of this study is the third of these aims, the vast majority of the available literature reflects the first two. In undertaking a review of the literature on calibration of hydrodynamic models, then, it is necessary to examine studies which, while undertaken with a different purpose in mind, may have some relevance to the current aim.

2.5.1 Parametric variability

Boyle et al. (2000) imitated manual model fitting procedure for transient streamflow models, in fitting separately portions of the hydrograph they describe respectively as “driven”, “non-driven quick” and “non-driven slow” flows, using a multi-objective optimisation to achieve a Pareto surface for the parameters. They reconciled the different parameterisations required for the different flow regimes, by choosing the parameter set on the Pareto curve that gave the minimum bias in terms of overall mean flow. Wagener et al. (2003) extended the concept of parameterising different parts of the time series, by taking a moving window, and distinguishing well-identified parameter ranges as the window moves, showing that different parameters are better identified at different parts of the time series.

Many studies investigating parametric variability in dynamic models are related to the Kalman filter (Kalman, 1960). This will be described later in the thesis, along with other approaches designed specifically for time-varying models.

2.5.2 Structural uncertainty

A popular method of treating model structural uncertainty, which could be applied to temporal or spatial data, involves averaging the output of different models, thus increasing the number of mechanisms that can be treated. Georgakakos et al. (2004) demonstrated that taking an unweighted mean of the output of a number of models improved model fit compared with output from the individual models. This comparison was made within the Distributed Model Intercomparison Project (Smith et al., 2004), using 11 models and 6 catchments.

Bayesian model averaging (BMA) is a technique used to combine models, while taking into account their individual skills. Thus, if there are K models, M_k , ($k = 1, \dots, K$) with parameters θ_k , each predicting a quantity Δ , subject to data, D , the joint prediction is (Hoeting et al, 1999):

$$\begin{aligned}
 P(\Delta|D) &= \sum_{k=1}^K P(\Delta|M_k, D)P(M_k|D) \text{ where} \\
 P(M_k|D) &= \frac{P(D|M_k)P(M_k)}{\sum_{l=1}^K P(D|M_l)P(M_l)} \text{ and} \\
 P(D|M_k) &= \int P(D|\theta_k, M_k)P(\theta_k|M_k)d\theta_k
 \end{aligned}
 \tag{2.5}$$

The posterior probability distribution is effectively a mean, weighted by the skill of the individual models.

Neuman (2003) applied a simplification of this technique to a set of three pre-calibrated models, and showed improved prediction skill. Duan et al. (2007) again used three competing models, each previously calibrated using three different objective functions, the Nash-Sutcliffe criterion, which is 1- the ratio of the error variance to the data variance (Nash and Sutcliffe, 1970), a measure of absolute error, and a measure taking into account the heteroscedasticity of the data. Taking note that streamflow is heteroscedastic, they applied the methodology to data normalised using the Box-Cox transformation (Box and Cox, 1964). Having shown that the BMA algorithm demonstrated the superiority of different parameterisations at different parts of the time series, the authors went on to parameterise each model with different parts of the input data for each of three catchments, again demonstrating improved skill. Rojas et al. (2008) combined Bayesian model averaging with GLUE to undertake a full solution to Equation (2.5) for a groundwater flow model, again comparing the effect of three different objective functions. They demonstrated the feasibility of this technique, but noted that it is computationally expensive.

It should be noted that a model averaging technique is only as good as the individual models. If there is a common deficiency with all of the input models, no averaging technique can lead to a good resultant model.

2.5.3 Input errors

Kavetski et al. (2002) and Huard and Maillot (2006) have shown that model parameterisation without accounting for input errors leads to biased parameter estimates. In a spatially aggregated catchment rainfall-runoff model, there is substantial uncertainty in defining the input rainfall. This is because the monitoring of rainfall takes place at individual locations, while a whole-catchment model requires input representing rainfall throughout the catchment. Spatially aggregated rainfall is difficult to estimate in the face of a sparse rainfall gauge network, which may not even overlap the catchment; Linsley et al. (1988, p60) give an illustration of the estimation errors. For catchment rainfall-runoff modelling, this is in fact a structural as well as an input error; different spatial rainfall patterns will lead to different catchment response, which cannot be captured in a spatially aggregated model.

The issue of input errors is not confined to rainfall-runoff models. In hydraulic modelling, upstream model input is often in the form of flow, which may be subject to rating curve errors. In addition, lateral inflow to the river is often not recorded for smaller tributaries; when it is, it has to be introduced to the main model in the form of flow, which is again likely to have been derived using a rating curve.

Kuczera et al. (2006) pointed out that in calibration of spatially aggregated rainfall-runoff models, rainfall spatial distribution may vary from storm to storm, giving the effect of volumetric errors in input rainfall. They thus proposed a multiplicative error model

applied to the rainfall input, with different multipliers for each storm. The multipliers were estimated simultaneously with other model parameters in a Bayesian regression analysis (Kuczera et al., 2006, Kavetski et al., 2006).

Thyer et al. (2009) examined the influence of the time-scale used for the rainfall multipliers, comparing the use of daily and storm-length multipliers. The use of daily rainfall multipliers represents a significant increase in the number of parameters to be determined. The authors noted that the output was insensitive to a significant number of the daily rainfall multipliers, associated with days where there was low rainfall, and discarded these in a preliminary analysis stage. In spite of this, they were only able to use a 2-year calibration period for the daily input error model, while for the storm-length error model, they were able to use a 5-year calibration period. They demonstrated that their parameterisation of input errors represents a significant improvement in model fit, runoff distributional consistency and in parameter consistency over a formulation without input error model. They suggested that the daily rainfall input error model seemed more appropriate, but suggested that for the catchment they used, the Horton catchment in New South Wales, the evidence was not conclusive. However, it is possible, that for wetter catchments the division of the rainfall record into storms, which was done in advance of the regression, would be less obvious, and therefore more subjective.

Ajami et al. (2007) tried to combine a formulation of the input error problem with Bayesian model averaging to account for structural errors; however, their statistical reasoning was flawed, as pointed out by Renard et al. (2009), as they had tried to analyse the problem by solving only for the distributional parameters of the rainfall multipliers rather than their actual values, leading to an ill-posed problem. Indeed, Renard et al. (2010) examined the possibility of simultaneously modelling input and structural errors using the regression formulation of Kuczera et al. (2006) but concluded that without adequate prior information to distinguish the error sources, the problem is ill-posed.

The Bayesian Forecasting System of Krzysztofowicz is an alternative approach to the incorporation of both input errors and structural errors, in this case in a real-time flood forecasting system, described in a series of papers (Krzysztofowicz, 1999, 2002; Krzysztofowicz and Kelly, 2000; Krzysztofowicz and Herr, 2001). This is a Bayesian time series river height prediction system, involving a hierarchy of precipitation and rainfall-runoff models, and was formulated in a general manner, allowing for any plug-in precipitation input and single stage rainfall-runoff model, and was implemented as a short-range real-time forecasting system. The model structure allows for the hydrological model to be processed under the assumption that the only uncertainties are those of data input, but the parametric uncertainty of the hydrological model is incorporated further downstream in the hierarchy. The transformation of the data to a Normal distribution, using an empirical transformation based on historical data, besides ensuring that all assumptions are supported, permits analytical solution of the model equations at each step, alternating between data assimilation and stage prediction.

Price (2006) approached the problem of estimation of input errors in a hydraulic model.

Reformulating the Saint Venant equations as functions of the lateral inflow q , a kinematic wavespeed, $c_0(Q)$ and an attenuation parameter $a_0(Q)$, the latter two described in terms of the instantaneous channel inflow Q , the study found sub-optimal values for these quantities as follows. Taking approximate values for $c_0(Q)$ and $a_0(Q)$, based on a uniform cross-section river at uniform slope, the model was calibrated for a variable lateral inflow. After smoothing the calculated values of the lateral inflow with a low-pass filter for greater realism, the equations were then re-solved for the functions $c_0(Q)$ and $a_0(Q)$. The methodology used highlights the identifiability problems between the functions of the main channel flow and the lateral inflow.

2.6 Requirements of current research

It is clear from the foregoing, that although there has been much activity in the area of calibration of hydrological and hydraulic models, the problem is far from solved.

A parameterisation method is required which is statistically coherent, treats different error sources separately, is suitable for calibration with time series data, or with satellite images, or both, and is also suitable for risk analysis; in other words produces a calibration which is not conditional on the input time series.

A method which has been applied with some success in other fields is that proposed, separately, by Craig et al. (2001) and Goldstein and Rougier (2004, 2006, 2009), and by Kennedy and O'Hagan (2001a, 2001b). This method has been developed for computer models with stationary output which varies over a spatial domain, and takes as its starting point a description of model bias as a stochastic distribution, distributed as correlated Gaussian, conditional on the locations where data measurements are made. The two approaches differ in that the method of Kennedy and O'Hagan uses a fully Bayesian representation, fully specifying prior and posterior distributions, while the method of Craig et al. (2001) and Goldstein and Rougier (2004, 2006) uses a Bayes Linear representation (Goldstein and Wooff, 2007), which specifies only distribution means and variances, but is quicker and more stable in computation.

The additional feature of this method is that it can easily incorporate an emulator for the computer model being calibrated; this permits the undertaking of the model calibration, and calibrated prediction, using a limited number of calls to the model, which makes a significant saving in computational time for all but the simplest of models.

In the work which follows, the method of Kennedy and O'Hagan is applied to the calibration of hydraulic models for fluvial flooding, using both satellite image and time series measurements.

Chapter 3

Bayesian analysis of computer code output

The methodology under investigation is one to calibrate deterministic computer models in the presence of model structural inadequacy, and has been developed by Kennedy and O'Hagan (2001a, 2001b), drawing on previous work by Sacks et al. (1989).

The data model considered is

$$z = M(x, \theta) + \delta(x) + \epsilon$$

where the observed data z are related to the output of the model $M(x, \theta)$ at input x with parameters θ , subject to errors caused by model structural inadequacy δ , and observation ϵ .

Distinction is made between the contributions of model inadequacy and observation error to observed data by considering that the observation errors are uncorrelated, while the contribution of model structural error is not. Then, the parallel is drawn between observation errors, which can be described by a Normal distribution, and model inadequacy, which is described by a Gaussian process, which assumes that the variation at a point in physical space is Normal, but incorporates a covariance function to model the joint variation through space.

This chapter shows how Gaussian process models can be employed in Bayesian model calibration to allow for model inadequacy, in the following steps. First, Gaussian processes are described, and it is shown how they can be used for function interpolation and emulation. The equations are then developed for Bayesian model calibration. It is shown how a computationally expensive model can be replaced by a Gaussian process emulator, and the equations are developed for calibration and for calibrated prediction. The Markov chain Monte Carlo method is introduced for simulation from a probability distribution. Finally, the methodology is demonstrated with respect to a simple algebraic example, identifying potential pitfalls in its use for real problems.

3.1 The use of Gaussian processes for interpolation and emulation

A Gaussian process is a multivariate Normal distribution. Defined over a d -dimensional space, a Gaussian process can be described in a hierarchical fashion as

$$y \sim N(m(x), V(x, x')) \quad (3.1)$$

where the mean, $m(x)$, and the covariance function, $V(x, x')$ are defined separately. For a stationary process, it may be assumed that the covariance between the two points x and x' is dependent on their separation; thus, the covariance could be described as $\sigma^2 r(x - x')$, where r is a correlation function, and $r(0) = 1$. As an example, one possible function to describe the covariance between points evaluated at two locations x and x' is an autoregressive relationship,

$$V(x, x') = \sigma^2 |x - x'|^\rho. \quad (3.2)$$

Another is the negative squared exponential function,

$$V(x, x') = \sigma^2 \exp(-(x - x')^T \Omega (x - x')) \quad (3.3)$$

where Ω is a positive semidefinite matrix, often taken to be diagonal, under the assumption that the different dimensions of the data are separable. In the above, the hyperparameters σ^2 representing variance, and the autoregressive coefficient ρ , in the case of Equation (3.2), or in the case of Equation (3.3), the coefficients ω of Ω , representing rate of variation (“roughness”, as described by Kennedy and O’Hagan) of the Gaussian process correction term, are to be determined. This last form (Equation (3.3)) is useful to describe smoothly varying functions, and has the advantage that all orders of derivatives exist and are continuous.

The properties of conditional Gaussian distributions are then used to describe the distribution at other locations, dependent on the measured or modelled values, since if x, y are partitioned into (x_1, x_2) and (y_1, y_2) ,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N \left(\begin{pmatrix} m(x_1) \\ m(x_2) \end{pmatrix}, \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \right)$$

where

$$\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \begin{pmatrix} V(x_1, x'_1) & V(x_1, x'_2) \\ V(x_2, x'_1) & V(x_2, x'_2) \end{pmatrix}$$

then if y_2 are known to have a value \tilde{y} , the conditional distribution is given by (e.g. Anderson, 1958, p36)

$$(y_1 | y_2 = \tilde{y}) \sim N(m(x_1) + V_{12} V_{22}^{-1} (\tilde{y} - m(x_2)), V_{11} - V_{12} V_{22}^{-1} V_{21})$$

Figure 3.1 illustrates a Gaussian process, with mean $\sin(2\pi x)$, and covariance a one-dimensional version of Equation (3.3), $V(x, x') = \sigma^2 \exp(-\omega(x - x')^2)$, conditional on specific values of the mean function, under different assumptions for known values of the “roughness” coefficient, ω . The values have been chosen for illustration; it can be seen that as ω increases, the rate of variation of the individual Gaussian process draws increases; while for large ω , the width of the 95% probability intervals away from known data points is dependent on the standard deviation σ , for small ω the width is limited by the high correlation between points in the x -direction.

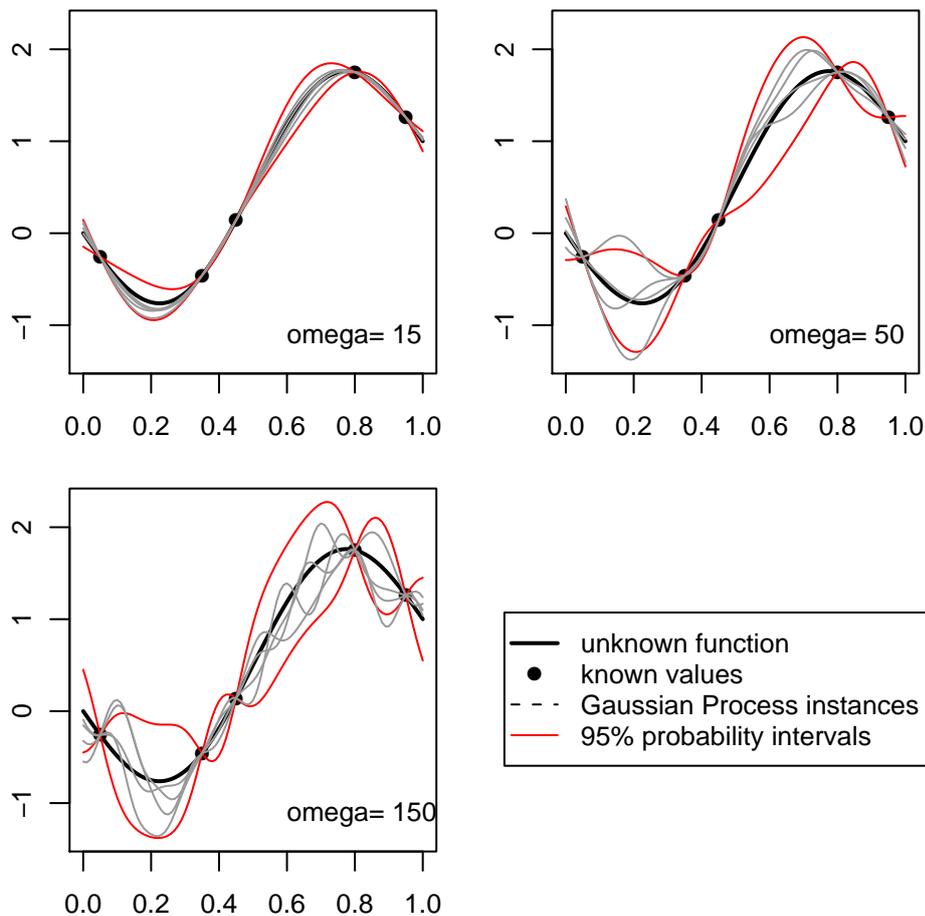


Figure 3.1: Illustration of a Gaussian process, with given values of ω and σ^2 , and known mean function, conditional on given data points.

The mean function in Figure 3.1 was taken to be the known generating function for the data points, to better illustrate the influence of the parameters ω and σ^2 . In practice, the generating function is not known, so has to be estimated. This can be achieved by describing it as a regression, $m(\cdot) = h(\cdot)^T \beta$, where $h(\cdot) = (h_1(\cdot), h_2(\cdot), \dots, h_p(\cdot))^T$ is some suitable basis, and $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ are regression coefficients to be determined. Figure 3.2 shows a more realistic illustration of the Gaussian process, based on an analytic solution for the conditional distribution subject to a maximum likelihood estimate for the “roughness” coefficient. The calculations have been performed using the BACCO

computer package (Hankin, 2005), and the equations are detailed in the Appendix at the end of this chapter. In this case, the mean function of the Gaussian process is a regression on the basis $(1, x)$. The first frame in Figure 3.2 shows that the Gaussian process is not well able to estimate the region of higher curvature away from the data points, but the addition of a single extra point makes it possible to approximate the original data curve sufficiently closely that the range of the Gaussian process is not visually evident.

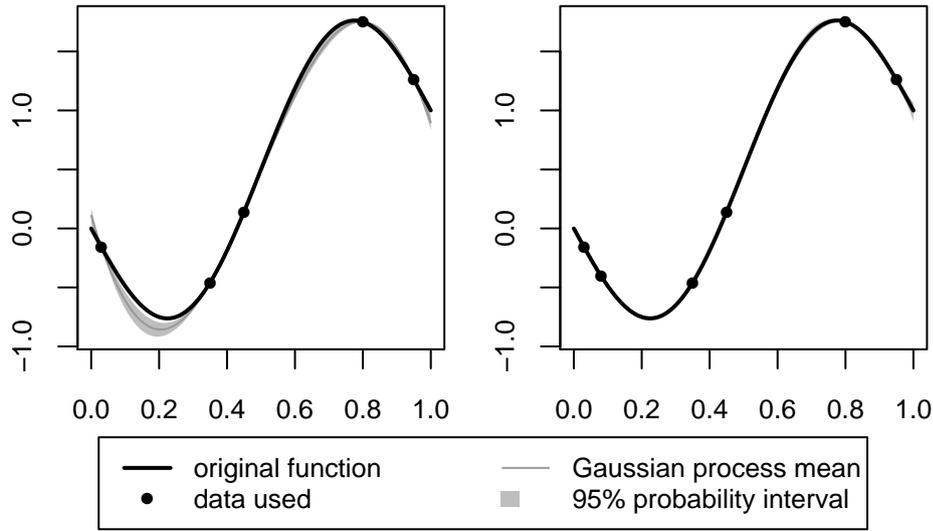


Figure 3.2: Estimated Gaussian approximation to given data points

3.2 The use of Gaussian processes for calibration

Suppose that a physical process $\zeta(x)$ is to be described, dependent on inputs x comprising locations x_1, \dots, x_n , and using a model $M(x, \theta)$, which invokes parameters θ requiring determination. It should be noted that θ does not necessarily include all the parameters of model M ; some may be sufficiently well determined by other means that they are included for the purposes of this analysis in x or even in M . The relationship between the model and the physical process it represents can then be described at a location x_i by

$$\zeta(x_i) = M(x_i, \theta) + \delta(x_i)$$

where $\delta(x)$ represents the model inadequacy, dependent on the inputs x .

Ignoring for the time being any observation error in the inputs x_i , the observed output z_i of the process ζ at x_i can be described by

$$z_i = M(x_i, \tilde{\theta}) + \delta(x_i) + \epsilon_i \quad (3.4)$$

where $\tilde{\theta}$ is the “best estimate” value of the parameters θ , and ϵ represents observation

error, assumed to be Gaussian, with zero mean and variance σ_ϵ^2 .

Kennedy and O'Hagan suggest that $\delta(x)$ and its prior distribution are conveniently described as Gaussian processes. As in expression (3.1), we can say

$$\delta(\cdot) \sim N(h_\delta(\cdot)^T \beta_\delta, V_\delta(\cdot, \cdot))$$

where as before, the mean model inadequacy is described as a regression $h_\delta(x)^T \beta_\delta$ on some suitable basis functions $h_\delta(\cdot)$, and $V_\delta(\cdot, \cdot)$ is a covariance function, described by hyperparameters ψ_δ .

Let $x = (x_1^T, x_2^T, \dots, x_n^T)^T$ be the vector of locations where observations have been made. Then the likelihood of the observed data is

$$\mathcal{L}\left((z_i - M(x_i, \theta)) \mid \theta, \beta_\delta, \psi_\delta, \sigma_\epsilon^2\right) \sim N(H_\delta(x_i, \theta)\beta_\delta, \Sigma)$$

where

$H_\delta(x)$ is a matrix representing the regression basis describing the Gaussian process mean, whose i^{th} row is $h_\delta(x_i)^T$

β_δ are the regression parameters describing the Gaussian process mean

ψ_δ represents the parameters describing the Gaussian process covariance

σ_ϵ^2 is the observation noise variance, and

$\Sigma = V_\delta((x, \theta), (x', \theta')) + \sigma_\epsilon^2 I$ is the covariance matrix, comprising the sum of the covariance representing the model inadequacy function and that of the noise

Thus, subject to prior distributions $f(\theta, \beta_\delta, \psi_\delta, \sigma_\epsilon^2)$ on θ , the regression parameters β_δ , the hyperparameters ψ_δ and the observation noise σ_ϵ^2 , the posterior distribution is given by

$$f(z_i \mid \theta, \beta_\delta, \psi_\delta, \sigma_\epsilon^2) \propto f(\theta, \beta_\delta, \psi_\delta, \sigma_\epsilon^2) \Sigma^{-1/2} \exp\left(\frac{1}{2} (z_i - M(x_i, \theta) - H_\delta(x_i)\beta_\delta)^T \Sigma^{-1} (z_i - M(x_i, \theta) - H_\delta(x_i)\beta_\delta)\right) \quad (3.5)$$

3.2.1 Model specification when the computer model is expensive to evaluate

The solution of this equation requires repeated calls to the computer model, to evaluate it with different parameter estimates during an iterative numerical solution process. If the model is of any computational complexity, solution rapidly becomes prohibitive. Kennedy and O'Hagan suggest that the answer to this is to replace the model in equation (3.4) by a further Gaussian process, defined conditionally on the computer model output at a finite

set of locations in the (input, parameter) space where it has been run; thus

$$\begin{aligned}\eta(x, \theta) &\sim N(m(x, \theta), V(x, x')) \text{ and} \\ \eta(x^*, \theta) &= M(x^*, \theta)\end{aligned}$$

at locations x^* where the computer model has been run. The suggestion that the output of a deterministic computer program can be described as a stochastic process is due to McKay et al. (1979). Clearly, this is not the case, as repeated runs of a deterministic program will yield the same output. However, before the program is run for a specific input configuration the output is not known, but may be approximated given the output of previous runs, under the assumption of smoothly varying output at locations x .

If the vector of computer program output is denoted y , and the observed data z , then these can be described by

$$\begin{aligned}y &= \eta(x^*, \theta) \\ z &= \eta(x, \tilde{\theta}) + \delta(x) + \epsilon\end{aligned}\tag{3.6}$$

where $\tilde{\theta}$ is the unknown true value of θ . Distinguishing between the Gaussian processes

$$\begin{aligned}\eta(x^*, \theta) &\sim N\left(H_1(x^*, \theta)\beta_1, V_1((x^*, \theta), (x^*, \theta'))\right) \quad \text{and} \\ \delta(x) &\sim N\left(H_2(x)\beta_2, V_2(x, x')\right)\end{aligned}$$

where H_1 and H_2 are defined as the matrices whose i^{th} rows are the regression bases $h_1(x_i, \theta_i)^T$ and $h_2(x_i)^T$ respectively, then Equation (3.6) becomes

$$\begin{aligned}y &\sim N\left(H_1(x^*, \theta)\beta_1, V_1((x^*, \theta), (x^*, \theta'))\right) \\ z &\sim N\left(H_1(x, \tilde{\theta})\beta_1, V_1((x, \tilde{\theta}), (x, \tilde{\theta}))\right) + N\left(H_2(x)\beta_2, V_2(x, x')\right) + N\left(0, \sigma_\epsilon^2\right)\end{aligned}$$

Combining these two, define

$$d = \begin{pmatrix} y \\ z \end{pmatrix}$$

then

$$d \sim N(H\beta, V)$$

where

$$\begin{aligned}H &= \begin{pmatrix} H_1(x^*, \theta) & 0 \\ H_1(x, \tilde{\theta}) & H_2(x) \end{pmatrix} \\ \beta &= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}\end{aligned}$$

and

$$V = \begin{pmatrix} V_1((x^*, \theta), (x^*, \theta)) & V_1((x^*, \theta), (x', \tilde{\theta})) \\ V_1((x, \tilde{\theta}), (x^*, \theta)) & V_1((x, \tilde{\theta}), (x', \tilde{\theta})) + V_2(x, x') + \sigma_\epsilon^2 I \end{pmatrix}$$

and H_1 and H_2 are defined above, and V_1 and V_2 are characterised by hyperparameter sets ψ_1 and ψ_2 respectively.

Thus the likelihood is

$$\mathcal{L}(d|\theta, \beta, \psi_1, \psi_2, \sigma_\epsilon^2) \propto V^{-1/2} \exp\left(\frac{1}{2}(d - H(x)\beta)^T V^{-1}(d - H(x)\beta)\right) \quad (3.7)$$

3.2.2 Integrating out the linear regression

Solution of Equation (3.5) or Equation (3.7) requires integrating out the parameters β_δ and the hyperparameters ψ_δ , and σ_ϵ^2 . Solution is facilitated by removal of the regression coefficients β_δ since these are often highly correlated with each other (if more than one) and with the θ parameters. Higdon et al. (2004) scale the problem to remove the β parameters, effectively taking point estimates from pre-analysis, and allowing the Gaussian processes to absorb the distributional uncertainty. The approach taken by Kennedy and O'Hagan (2001a, 2001b) is to integrate out the β parameters under the assumption of an improper uniform prior distribution, and to solve the problem conditionally on this assumption. Taking expression (3.5) and recognising that

$$(z - H(x)\beta)^T \Sigma^{-1}(z - H(x)\beta) = (\beta - \hat{\beta})^T H(x)^T \Sigma^{-1} H(x)(\beta - \hat{\beta}) + \text{constant terms}$$

where $\hat{\beta} = (H(x)^T \Sigma^{-1} H(x))^{-1} H(x)^T \Sigma^{-1} z$ is the classical least squares solution to the linear equation $z - H(x)\beta \sim N(0, \Sigma)$ (e.g. O'Hagan and Forster, 2004), expression (3.5) can then be integrated with respect to β_δ leading to

$$\mathcal{L}(z|\theta, \psi_\delta, \sigma_\epsilon^2) \propto \frac{\mathcal{W}^{1/2}}{\Sigma^{1/2}} \exp\left(\frac{1}{2}(z - M(x, \theta) - H_\delta(x)\hat{\beta}_\delta)^T \mathcal{W}^{-1}(z - M(x, \theta) - H_\delta(x)\hat{\beta}_\delta)\right) \quad (3.8)$$

where

$$\begin{aligned} \beta_\delta &\propto N(\hat{\beta}_\delta, \mathcal{W}), & \text{with} \\ \hat{\beta}_\delta &= \mathcal{W} H_\delta^T \Sigma^{-1} z, & \text{and} \\ \mathcal{W} &= (H_\delta^T \Sigma^{-1} H_\delta)^{-1} \end{aligned}$$

A similar treatment of expression (3.7) leads to

$$\mathcal{L}(d|\theta, \psi_1, \psi_2, \sigma_\epsilon^2) \propto \frac{W^{1/2}}{\Sigma^{1/2}} \exp\left(\frac{1}{2}(d - H(x)\hat{\beta})^T W^{-1}(d - H(x)\hat{\beta})\right) \quad (3.9)$$

where

$$\begin{aligned}\beta &\propto N(\hat{\beta}, W), && \text{with} \\ \hat{\beta} &= WH^T V^{-1}d, && \text{and} \\ W &= (H^T V^{-1}H)^{-1}\end{aligned}$$

3.2.3 Specification of priors

Other than the prior on the regression parameters β , the remaining priors needed to specify the problem are the priors on the computer model parameters θ , the hyperparameters ψ of the model inadequacy and the emulator if included, and of the observation error variance, σ_ϵ^2 . Under the assumption that the observation error is a function only of the data recording mechanism, and that error values are uncorrelated and independent of the data values, its variance should be fairly well known from the observation mechanism. It is to be expected that the prior on the parameters are independent of the priors on the model inadequacy, and both of these are expected to be independent of the priors on the emulator.

3.2.4 Solution of equations

Kennedy and O'Hagan point out that analytical solution of Equation (3.9) is not feasible, and instead find the hyperparameters by optimisation. Although they derive an expression for the distribution of the computer model parameters θ conditional on these optimal values, they do not evaluate it, concentrating instead on the calibrated prediction of the output. This solution method has been implemented in R (R Development Core Team, 2009) by Hankin (2005) for a smooth covariance function (Equation (3.3)), using Markov chain Monte Carlo (described in Section 3.3 below) to evaluate the posterior distribution of the computer model parameters θ .

The alternative is to simulate from the complete posterior distribution (Equation (3.8) or Equation (3.9)) by MCMC. This is the approach primarily taken in this study. The Markov chain is constructed with reference to the logs of the hyperparameters, firstly because the transformation ensures positive values of the hyperparameters, and secondly because the hyperparameter values can vary by several orders of magnitude.

3.2.5 Calibrated prediction

Once the conditional posterior distribution of the data has been found by simulation, output of the process can be estimated at any input location x^\dagger , using the form of the conditional Gaussian distribution. Thus, for the formulation with direct calls to the com-

puter model (Equations (3.8)),

$$\begin{aligned} E\left(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2\right) &= M(x^\dagger, \theta) + h_\delta(x^\dagger, \theta)^T \hat{\beta}_\delta + \tau(x^\dagger, \theta)^T \Sigma^{-1} \left(z - M(x, \theta) - H_\delta(x, \theta) \hat{\beta}_\delta\right) \\ \text{Var}\left(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2\right) &= V_\delta(x^\dagger, x^\dagger) - \tau(x^\dagger, \theta)^T \Sigma^{-1} \tau(x^\dagger, \theta) + \Lambda^T \mathcal{W} \Lambda \end{aligned} \quad (3.10)$$

where in addition to variables previously defined

$$\begin{aligned} \Lambda &= (h_\delta(x^\dagger, \theta) - H_\delta(x, \theta)^T \Sigma^{-1} \tau(x^\dagger, \theta)), \text{ and} \\ \tau(x^\dagger, \theta) &= V_\delta((x^\dagger, \theta), (x, \theta)) \end{aligned}$$

The formulation where the computer model is replaced by an emulator (Equations (3.9)) proceeds similarly:

$$\begin{aligned} E\left(\zeta(x^\dagger)|d, \theta, \psi, \sigma_\epsilon^2\right) &= h(x^\dagger, \theta)^T \hat{\beta} + t(x^\dagger, \theta)^T V^{-1} (d - H(x, \theta) \hat{\beta}) \\ \text{Var}\left(\zeta(x^\dagger)|d, \theta, \psi, \sigma_\epsilon^2\right) &= V_1((x^\dagger, \theta), (x^\dagger, \theta)) + V_2(x^\dagger, x^\dagger) - t(x^\dagger, \theta)^T V^{-1} t(x^\dagger, \theta) + L^T W L \end{aligned} \quad (3.11)$$

where

$$\begin{aligned} L &= (h(x^\dagger, \theta) - H(x, \theta)^T V^{-1} t(x^\dagger, \theta)), \\ h(x^\dagger, \theta) &= \begin{pmatrix} h_1(x^\dagger, \theta) \\ h_2(x^\dagger) \end{pmatrix}, \text{ and} \\ t(x^\dagger, \theta) &= \begin{pmatrix} V_1((x^\dagger, \theta), (x, \theta)) \\ V_1((x^\dagger, \theta), (x^*, \theta)) + V_2((x^\dagger, x^*)) \end{pmatrix}, \end{aligned}$$

and the other variables are as defined previously.

The conditional posterior expectation and variance are combined to give the unconditional expectation and variance of the required quantity. The computation is implemented using Markov chain Monte Carlo, which is described below.

3.3 Markov chain Monte Carlo

Many problems formulated in the Bayesian paradigm, as this one, are not amenable to analytical description of the posterior distribution. Under these circumstances, the posterior distribution must be characterised by simulation. This can be done by Monte Carlo methods, by random draws from the prior distribution, accepting these with probability proportional to the likelihood. However, the most efficient way of simulating from a distribution is by taking a Markov process, whose equilibrium distribution is that of the posterior distribution we wish to simulate, and drawing from this process. A Markov process is a sequence, with the property that any member is dependent only on the im-

mediately preceding member of the sequence:

$$P(X_i|X_1, X_2, \dots, X_{i-1}) = P(X_i|X_{i-1})$$

Transition from one member of the chain to the next is characterised by a transition kernel, thus:

$$P(X_{i+1}|X_i) = P(X_i|X_{i-1})$$

If a Markov chain is both irreducible, and aperiodic, that is, if it is possible for the chain to reach every part of the parameter space in a finite number of transition steps from any starting point, then it is ergodic; in other words, it has a limiting stationary distribution. This property enables a Markov chain to be used to simulate from a target distribution; once the chain has achieved the stationary distribution, all further members will belong to the distribution, permitting simulation from it. A sufficient condition for the chain to have a specified invariant distribution $f(\cdot)$ is for it to possess the property of detailed balance; that is that

$$f(\theta)P(\phi|\theta) = f(\phi)P(\theta|\phi) \text{ for all } \theta, \phi \text{ in the parameter space } \Theta \quad (3.12)$$

3.3.1 Metropolis Hastings algorithm

The Metropolis Hastings algorithm (Metropolis et al., 1953; Hastings 1970) provides a simple way to ensure that a Markov chain possesses the detailed balance property. Given the current state of the chain, X_i , a proposal X^* is generated from a proposal distribution $q(X_i, X^*)$. This proposal is accepted with probability

$$\alpha\{X_i, X^*\} = \min\left(1, \frac{f(X^*)q(X_i|X^*)}{f(X_i)q(X^*|X_i)}\right), \quad (3.13)$$

and becomes the next member of the chain, X_{i+1} ; otherwise, X_{i+1} is set to X_i . The transition kernel density is $K(X_i, X^*) = q(X_i, X^*)\alpha(X_i, X^*)$. It can easily be seen that this transition kernel satisfies the detailed balance equation (3.12), with f as an invariant distribution; thus convergence is guaranteed. In addition, it is not necessary to compute the denominator in Bayes equation (1.3), as if the detailed balance property holds, this cancels out from the acceptance relationship, Equation (3.13).

Different proposal distributions $q(\cdot|\cdot)$ can be used. A common choice is to take $X^* = X_i + \epsilon$, where ϵ is a random increment independent of X_i . This is known as random walk Metropolis-Hastings algorithm, and the distribution of ϵ is commonly taken as uncorrelated multivariate Normal.

Although convergence to the target distribution is guaranteed by the Metropolis Hastings scheme, this does not imply that convergence will occur in a reasonable number of iterations. The variance of the proposal increment ϵ needs to be prespecified, and too small a step leads to inadequate exploration of the parameter space, while too large a stepsize results in a small acceptance rate at Equation (3.13). Common practice is to run short

pilot chains, to ensure that the stepsize is appropriate. Although Roberts et al. (1997) suggested that for multivariate Normal distributions, the optimal acceptance rate to permit adequate exploration of the parameter space and reasonable convergence rates, should be approximately 23%, in reality, for more complex distributions, a lower rate has to be accepted. For a multivariate distribution, different step sizes may well be required in each dimension. For highly correlated or skewed variable distributions, it is best to apply the algorithm to transformed variables.

3.3.2 MCMC diagnostics

One difficulty with using MCMC methods, is to know how long the chain has to be. Although convergence to a stationary distribution is guaranteed by the Metropolis Hastings scheme, there is no theoretical indication as to how long this will take. In addition, when proposal distributions are based on the current state of the chain, it can be expected that there is significant autocorrelation in the chain, affecting the length required for an effective estimate of the parameter distributions. Apart from visual inspection of the chains, their cumulative sums and their correlations, a number of methods have been suggested to identify the length of the “burn-in” period, and to estimate the sample size required for effective estimation of distributional properties (e.g. Cowles and Carlin, 1996). The CODA package (Best, Cowles and Vines, 1995) incorporates several of these for analysis of the output chain.

A test for stationarity of the chain was proposed by Geweke (1992), who used spectral methods to compare the means of the first 10% and the last 50% of the chain under the assumption that the two parts of the chain are asymptotically independent. This method is applied to a single variable at a time.

Gelman and Rubin (1992) pointed out that it may not be possible to detect very slow convergence from a single chain, while by taking independent chains starting at a sample of points, convergence to a single distribution gives greater confidence. They proposed a diagnostic based on the comparison of the covariances of individual chains and the covariances between chains.

Raftery and Lewis (1996) proposed an estimator for the length of chain necessary to estimate a quantile of the posterior distribution with the required confidence. Taking the q^{th} quantile u of a distribution U of a function of the parameters in the chain, they form a new chain $Z : z_t = I(U \leq u)$, where $I(\cdot)$ is the indicator function. While Z_t is not a Markov chain, it can be expected to behave like one if thinned to every k^{th} value, for large enough k ; the Raftery and Lewis algorithm *gibbsit* identifies the lowest value of k required, and uses this to estimate both the length of the burn-in period, and the chain length for estimation of the required quantile. It should be noted that the estimate may change if the chain is extended, so new estimates must be calculated, until they are shorter than the actual chain length.

R-CODA (Plummer et al., 2006) implements a modified version of the method of Heidel-

berger and Welch (1981) to examine the autocorrelation behaviour of the chain, fitting a generalised linear model to the lower part of the spectrum, and making it possible to estimate the effective length of the chain.

It is generally considered that no method for analysing MCMC performance is foolproof, and it is considered prudent to use a number of methods in parallel, to ensure the validity of the MCMC output.

An example of the problems which can occur in MCMC is demonstrated in Figures 3.3, showing the chains and 3.4, showing their correlations. This is in fact a solution to the toy example described in the next section, but with deliberately poor step sizes for the proposal distributions, and without integrating out the regression parameter. Thus, the variables are the parameter θ , the regression parameter β , the log model inadequacy “roughness” parameter ω_δ , and the log variances for the model inadequacy σ_δ^2 and the observation noise σ_ϵ^2 , with priors taken as uniform on a limited range (see Table 3.1), to be able to see the posterior distributions. The fact that this is a comparatively straightforward distribution can be seen in that the variable values settle down almost immediately into distributions seen throughout the rest of the chain.

Table 3.1: Prior ranges for MCMC example

variable	range
θ	[-1, 5]
β_δ	[-6, 6]
$\log(\omega_\delta)$	[-2, 6]
$\log(\sigma_\delta^2)$	[-18, 5]
$\log(\sigma_\epsilon^2)$	[-18, -2]

The first two variables, θ and β can be seen to be highly correlated, both from the correlation plots, and also the symmetry between their chains. This results in a low acceptance rate of the proposal, giving rise to the step-like behaviour seen in the first two chains, and the irregular appearance of the corresponding distributions shown in the the diagonal plots in the correlation figure. The third and fourth variables $\log(\omega_\delta)$ and $\log(\sigma_\delta^2)$ are poorly identified; the posterior distributions can be seen to be strongly influenced by the priors. The proposal step size given to $\log(\omega_\delta)$ is larger than that given to $\log(\sigma_\delta^2)$; this is responsible for the very different appearance of the chains. By contrast with the other variables, the chain for $\log(\sigma_\epsilon^2)$ is fairly well behaved. Lastly, there are positions in the chain where all of the variables stop moving together, resulting in local isolated peaks in the correlation plots, and while at about 8500 this appears to be associated with extreme values of θ and β , this is not always the case. This last behaviour occurred for calibration problems in a number of applications of the methodology, and while the length of these flat patches in the current example is not large, for more difficult problems it can become extended.

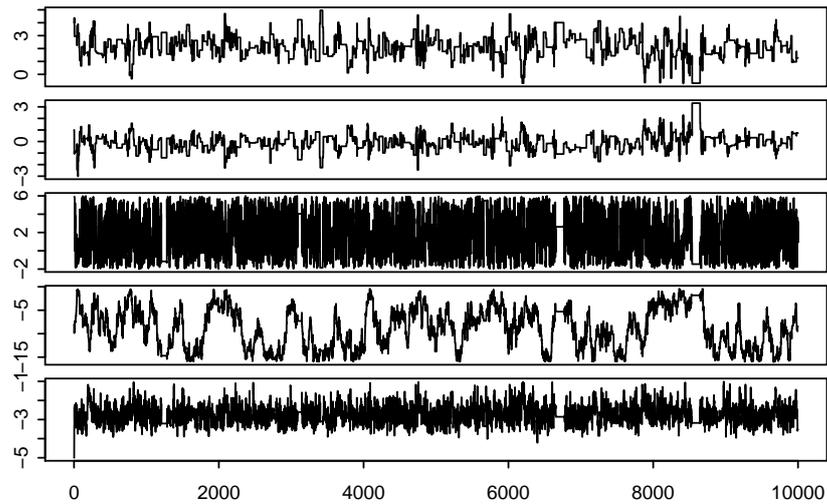


Figure 3.3: Example chains to show MCMC problems

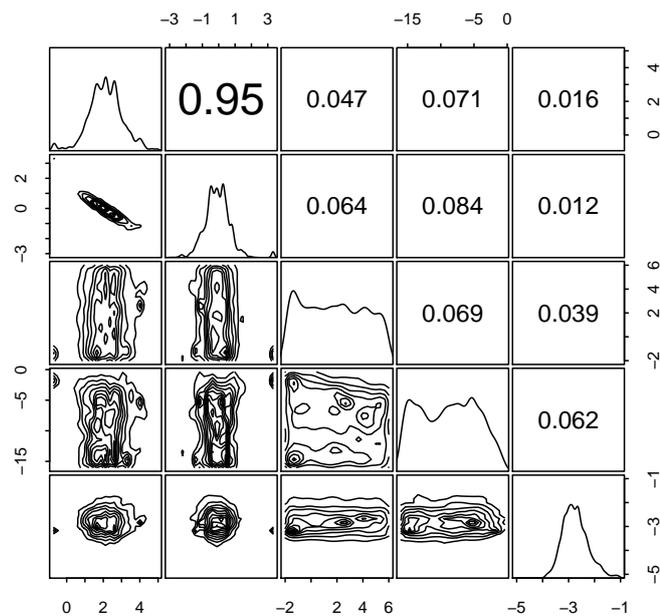


Figure 3.4: Example correlations corresponding to the chains in Figure 3.3: Upper right, pairwise correlation; Diagonal, individual variable density; Lower left, 2-dimensional density plots

3.3.3 Refinements of the Metropolis-Hastings algorithm

A refinement to the simple random walk Metropolis Hastings algorithm described above was suggested by Haario et al. (2001). This algorithm uses the empirical covariance of the entire chain, up to the current point, to inform the (still multivariate Normal) distribution of the increment ϵ for the proposal. Thus the increment for each variable is appropriate to the scale of variation of that variable. It is necessary to use the entire chain to generate the proposal distribution to maintain ergodicity; Andrieu and Thoms (2008) pointed out that

a necessary condition for this is that the adaptations to the proposal distribution vanish as the step number $i \rightarrow \infty$. Marshall et al. (2004) observed that the adaptive algorithm of Haario et al. (2001) provides substantial improvement in convergence properties over a simple random walk Metropolis Hastings algorithm in the calibration of a rainfall-runoff model.

Vrugt and co-workers have implemented two MCMC methods, designed to simulate from difficult distributions, both employing parallel and interacting chains; each of these incorporates a number of performance-improving characteristics. The first of these (Vrugt et al., 2003) draws inspiration both from the algorithm of Haario et al. (1999, 2001), and from the successful shuffled complex evolution optimisation method of Duan et al. (1992), running a number of parallel chains. In a reflection of the work of Duan et al., the algorithm maintains a “complex” of candidate points for each chain, which are used to choose the next point in each of the parallel chains. The method of choosing a new point reflects the early algorithm of Haario et al. (1999) in using the mean and covariance last m members of the complex to base the multivariate Normal candidate distribution for the new point. As in the work of Duan et al., the complexes are periodically shuffled, so that the chains are not independent. The algorithm appears fast and efficient, but is flawed, since it uses the covariance of the last m members of the chain to calculate the new candidate point. Haario et al. (2001) pointed out this flaw in their earlier work, proving that it contravenes the conditions necessary to ergodicity of the chain.

A second algorithm, DREAM (Vrugt et al., 2008), was introduced to tackle larger parameter sets encountered when solving explicitly for input errors in hydrological models. It incorporates a number of separate methods to improve the robustness and convergence speed of the Metropolis-Hastings algorithm. As with the previous method, the inspiration for the proposal distribution came from the optimisation literature. The differential evolution method (ter Braak, 2006) involves running parallel chains, but updating each one using the difference between a randomly chosen pair from the others. DREAM goes one step further, using the mean of the differences between a number of randomly chosen pairs of separate chains. In the case where the proposal is rejected, DREAM employs the delayed rejection algorithm of Haario et al. (2006), rejecting the update only if a second, less stringent acceptance test is failed. In addition, Vrugt et al. note that this method admits the possibility of a single chain which is substantially different from the rest of the population, and replace such chains with others which are more similar. The method appears fast and efficient.

In the work described in this thesis, both of the algorithms described above by Vrugt have been used at different times, in model development. It proved helpful to have access to more robust and efficient codes at the stage where the models were not completely specified. However, all the results in shown the thesis have been achieved with a simple random walk Metropolis-Hastings code, with fixed time step, since, correctly specified, the problems here do not present such a difficult task for a Markov chain Monte Carlo code.

3.4 Application: simple algebraic example

The methodology is illustrated with respect to a simple algebraic example. Data were generated by the process

$$y = e^x - 1$$

in the range $x \in [0, 1]$. The computer program used to describe the data follows the relationship

$$y = \theta x^2$$

where θ is the parameter to be determined. Observations were measured at each of 5 points, (0.1, 0.3, 0.5, 0.7, 0.9), subject to additive Gaussian noise. Initially, the simulator was used directly in the calculations, solving the model described in Equations (3.5), (3.8) and (3.10). The regression basis for the model inadequacy function was taken to be (x) , and a smooth covariance was chosen, as in Equation (3.3). The equations were solved, using MCMC, for the the parameter θ and the logs of the variables ω , σ_δ^2 and σ_ϵ^2 , and to estimate the calibrated prediction of the generating process.

Prior distributions are summarised in Table 3.2, and were chosen as follows: for the parameter θ , it was assumed that the final value would be approximated by the value found in a classical least squares solution to the equation $y = \theta x^2$, using the data; thus $\theta \sim N(1.94, 1)$. It was assumed that the data standard deviation was in the region of 0.05, and the model inadequacy standard deviation was assumed to be approximately twice this; thus $\log(\sigma_\delta^2) \sim N(-4.6, 2)$, and $\log(\sigma_\epsilon^2) \sim N(-6, 2)$. In order to make an initial judgement of the “roughness” parameter ω_δ , a range was taken by comparison with representations of zero-mean Gaussian processes (Figure 3.5), following the suggestion of Oakley (2002); thus $\log(\omega_\delta) \sim N(2.3, 2)$.

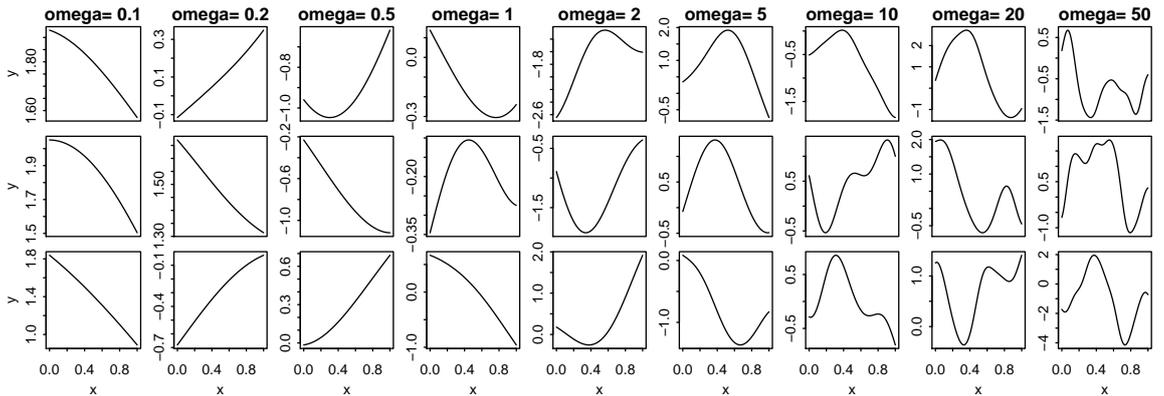


Figure 3.5: Reference Gaussian processes for use in determining prior distributions for “roughness” values: three realisations of each.

Calibration was undertaken with two measurements at each data point, subject to noise of standard deviation 0.05. Posterior distributions were found by simulating from dis-

Table 3.2: Prior and posterior distributions for calibration

variable	prior distribution	posterior mean	posterior standard deviation	
θ	$N(1.94, 1)$	1.31	0.48	
β_δ		0.50	0.17	
$\log(\sigma_\delta^2)$	$N(-4.6, 2)$	-5.43	1.26	corresponds to $\hat{\sigma}_\delta = 0.066$
$\log(\omega_\delta)$	$N(2.3, 2)$	2.00	2.58	
$\log(\sigma_\epsilon^2)$	$N(-6, 2)$	-4.81	0.65	corresponds to $\hat{\sigma}_\epsilon = 0.090$

tribution given by Equation (3.8), using a chain of length 20000. None of the problems demonstrated in Figures 3.3 and 3.4 were experienced to any significant extent; chains and correlations in this case are shown in Figures 3.6 and 3.7. Note that there are here only four variables represented, since the regression coefficient β_δ has been integrated out of Equation (3.8).

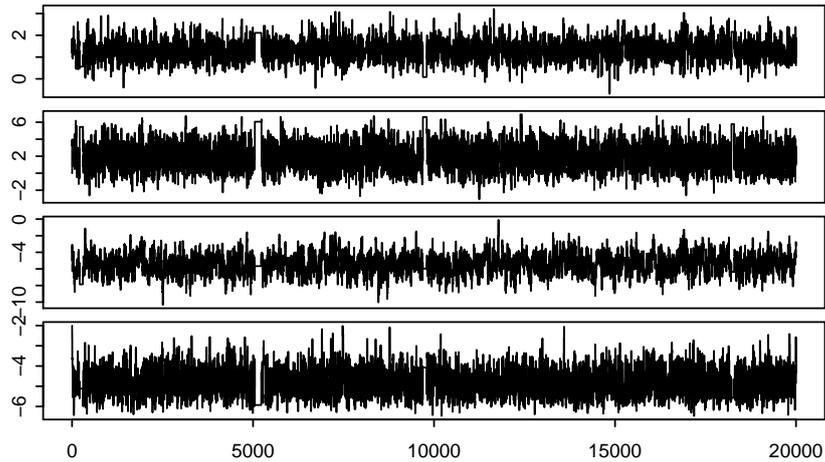


Figure 3.6: Chains for toy example

Posterior distributions are summarised in Table 3.2. The posterior distribution of the parameter θ has mean 1.31 and standard deviation 0.48 (Figure 3.8), and is very different from the estimate by classical methods, which corresponds with the prior mean.

Calibrated prediction was performed by conservatively discarding the first quarter of each chain, and for each remaining member, calculating its contribution to the calibrated prediction, using Equation (3.10). It is clear that the calibrated prediction follows the data closely (Figure 3.9a), with the mean width of the 95% prediction interval 0.065.

For comparison, the analysis was repeated with 4 measurements at each data point (Figure 3.9b). In this case, the posterior mean of θ was lower, at 1.11, and standard deviation was 0.37. The estimated noise standard deviation was somewhat reduced, at $\hat{\sigma}_\epsilon = 0.068$, with the mean width of the 95% prediction interval 0.041.

It should be noted that the interpretation of the prediction interval in Bayesian statistics

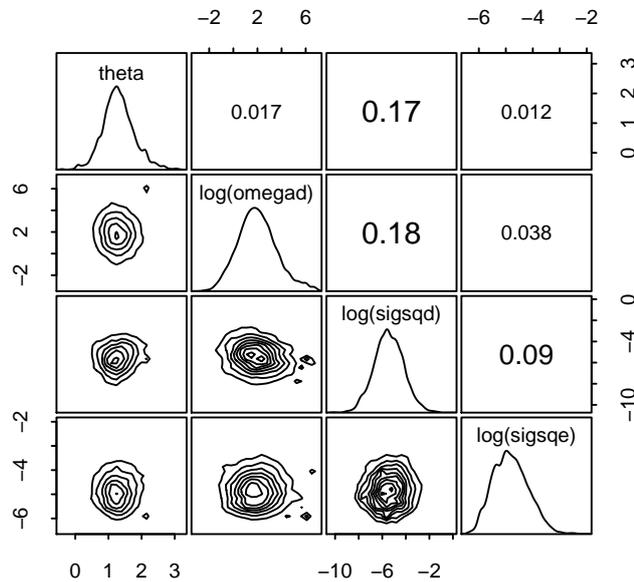


Figure 3.7: Correlations corresponding to the chains in Figure 3.6: Upper right, pairwise correlation; Diagonal, individual variable density; Lower left, 2-dimensional density plots

is different to that used in classical statistics. Since the solution of the Bayes' scheme leads to a probability distribution for the required variables, this is most easily comprehended in relation to the original research question by specifying the mean of the distribution, and an interval within which the variable lies, with a specified probability p . By contrast, the confidence interval of classical statistics is the interval within which the mean lies, with probability p .

3.4.1 Identifiability of the solution with sparse data

The above solutions to the calibration problem were undertaken with replicated data, and demonstrated that replication provides sufficient data to identify adequately the noise variance. A more realistic test of the type of situation which may be encountered in practice in calibration of a hydraulic model, is an example where only one measurement is able to be made at each data point. Solving the calibration problem in this case is much more difficult, illustrating identifiability issues in the formulation.

Poor identifiability occurs when the data do not support full identification of the parameters. In the case of the Kennedy and O'Hagan formulation, there may be difficulties in distinguishing between the model inadequacy function and the observation error on the measured output data, as originally suggested by Wynn (2001). Even when a parameter is poorly identified by the data, the problem is not necessarily ill-posed; provided that the prior distributions are adequately precisely defined, it is still possible to achieve a solution.

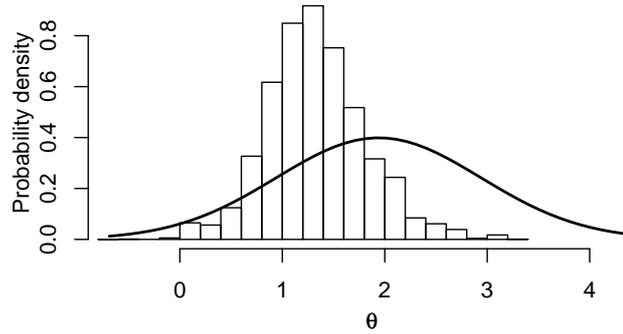
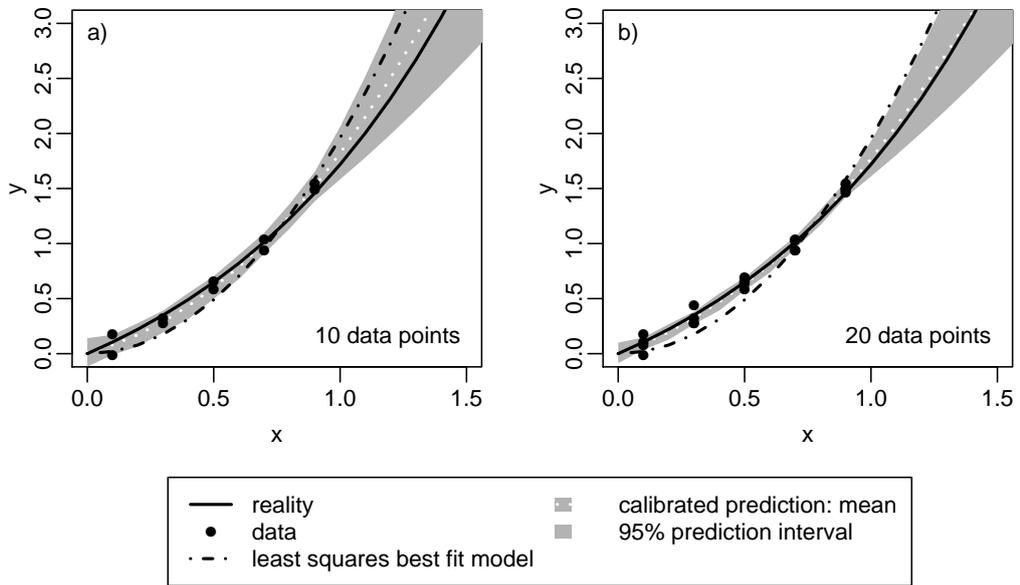
Figure 3.8: Prior and posterior distributions for parameter θ 

Figure 3.9: Calibrated prediction of algebraic example, for a) 10 data points and b) 20 data points

In order to try and find out what influences the identifiability of the hyperparameters, the likelihood (Equation (3.8)) was calculated on a grid of values of $\log(\omega_\delta)$, $\log(\sigma_\delta^2)$ and $\log(\sigma_\epsilon^2)$, taking values of β and θ corresponding to the classical least squares solution. Figure 3.10 represents two series of slices through the $(\log(\omega_\delta), \log(\sigma_\delta^2), \log(\sigma_\epsilon^2))$ -space, the first row parallel to the $\log(\sigma_\epsilon^2)$ -plane, and the second to the $\log(\sigma_\delta^2)$ -plane for a given realisation. It can be seen that there appear to be two competing solutions, for $\log(\sigma_\delta^2) \approx -6$, $\log(\omega_\delta) \approx 2.5$, and $\log(\sigma_\epsilon^2) < -10$, and the second for $\log(\sigma_\delta^2) \approx -6$, $\log(\omega_\delta) \approx 1.5$, and $\log(\sigma_\epsilon^2) \approx -7.5$, interpreting the residual with different proportions of model inadequacy and noise.

Further investigation of the likelihoods in this way revealed that for a number of realisations, one or other of the hyperparameters appeared to be unidentifiable. This appears to happen when one of the σ^2 variables is at least 2 orders of magnitude greater than the other. In this case, the smaller σ^2 value makes little difference to the total variance. If

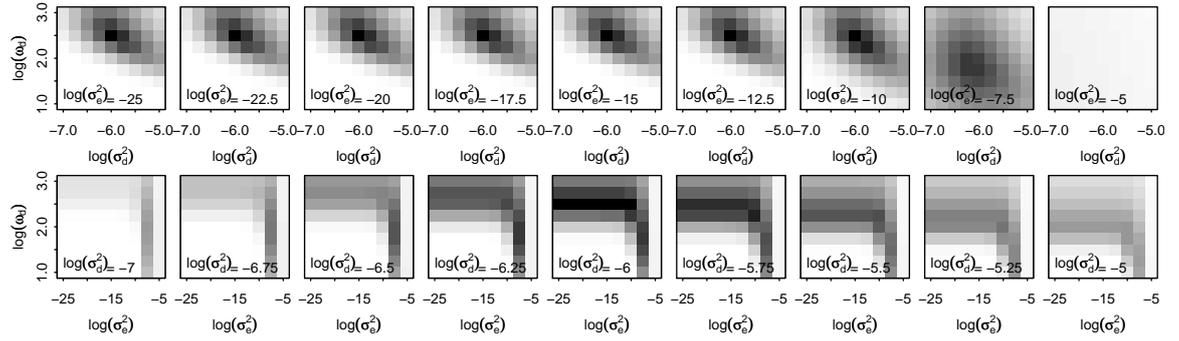


Figure 3.10: Greyscale plot of the likelihood over a 3-dimensional grid of hyperparameter values

the smaller value is σ_δ^2 , then ω_δ is also unidentifiable.

There are two other causes for the non-identifiability of ω_δ . The first is that with a large enough value of ω_δ , a one dimensional inadequacy function is indistinguishable from the noise. Thus, for a spacing of 0.1 between data points, there is little effective difference between uncorrelated noise, and correlated noise with a “roughness” of $> O\left(\frac{1}{0.1^2}\right) = O(100)$; that is $(\log(\omega_\delta) > 5)$. In addition, it should be noted that using a log parameterisation for the hyperparameters may have the consequence that $\log(\omega_\delta)$ is not well defined for small ω_δ , for example.

It has already been noted that where there are identifiability problems in the specification of a Bayesian problem, it can still be well-posed if the priors are adequately specified. However, the solution will follow the information in the prior, more or less closely, according to how much information is in the likelihood. This was demonstrated in the third and fourth variables of the MCMC example in Figures 3.3 and 3.4.

In the full solution of the problem, with a single data point at each of 5 equally spaced locations (noise variance 0.05), Normal priors, variance 100 for the parameters and 10 for the log hyperparameters, the posterior distributions of the parameter θ , the discrepancy β , and the noise variance σ_ϵ^2 were reasonably well determined, and not much influenced by the prior. However, the hyperparameters of the inadequacy function, σ_δ^2 and ω_δ were quite strongly influenced by the prior means. In addition, the priors had considerable influence on the posterior prediction intervals, as shown in Table 3.3 below:

Table 3.3: Parameters affecting the posterior credible interval

increase in posterior credible interval	no increase in posterior credible interval
σ_δ^2 mean	θ variance
σ_δ^2 variance	β variance
σ_ϵ^2 mean	ω_δ mean
	ω_δ variance
	σ_ϵ^2 variance

3.4.2 Use of an emulator to represent the computer program

For the purposes of comparison, an emulator was used to represent the “computer program” - in this case, the algebraic example, $y = \theta x^2$. The “computer program” was run for selected values of x and θ , and a regression was performed on the output values. Kennedy and O’Hagan recommend using the simplest possible regression basis that is supported by the data, suggesting that in many cases it is sufficient to use the mean computer output value; that is, the regression basis is simply (1). This is in contrast to the approach of Rougier et al. (2009), who point out that the better the regression basis, the less the effect of the choice of an inappropriate covariance function. In this case, the basis $(1, x, \theta)$ has been taken.

It is advisable to test the accuracy of the emulator away from the points at which the computer program has been run. In this case, since there are two independent variables, it is not so simple to visualise the emulator error over the entire domain. Instead, a selection of points has been chosen in the domain, and the emulator error has been calculated for those. As for Figure 3.2, the calculations have been performed using the BACCO package.

The first frame of Figure 3.11 shows a contour map of the emulator mean response surface, identifying points where the computer program has been run, and where the emulator error has been tested. The second frame shows the 95% error range compared with the exact function at a number of validation points that were not used for training, marked in Figure 3.11a by circles. It can be seen that both the mean error and the uncertainty in the emulator are low, but are higher further away from data points, particularly close to the edge of the domain where the prediction intervals expand rapidly.

3.4.3 Experimental design

When characterising computer code output to construct an emulator, it is necessary to decide at which input values the code should be run in order to ensure that the emulator faithfully represents the computer code output throughout the input domain. This question was discussed by Sacks et al (1989), although no solution was offered in the context of emulator formulation.

The issue of deriving statistics from output of computer code sample runs was discussed by McKay et al. (1979), who devised the Latin hypercube sampling strategy. This concerns the choice of a sample of size N from a K dimensional input domain $[0, 1]^K$. For each dimension $k \in \{1, \dots, K\}$, the value of the i^{th} sample ($i \in \{1, \dots, N\}$) is taken to be at

$$\frac{\pi_k(i) - U(0, 1)}{N}$$

where $U(\cdot, \cdot)$ refers to the uniform distribution, and $\pi_k(i)$ refers to the i^{th} member of a permutation of the N intervals, the suffix emphasizing that a different permutation is taken for each dimension k . McKay et al. showed that a Latin hypercube sample provides

accuracy. The concept of space-filling sequences was introduced by Halton (1960), while Sobol (1998) showed that use of such non-random sequences is a more efficient base for Monte Carlo simulations than random draws, but noted that the number of sampling points should be increased by doubling. An alternative approach to increasing the size of a sample was suggested by Sallaberry et al. (2008), who proposed a method of increasing the size of a Latin hypercube sample to a multiple of the original sample size by generating a new Latin hypercube sample, and subdividing the interval size so that the samples do not interfere.

3.4.4 Effect of the choice of regression basis on solution behaviour

An investigation has been made of the dependence of the calibrated prediction on the choice of regression bases for the emulator and the model inadequacy function. In interpolation, provided that there are no excessively large gaps between the data points (either code outputs in emulator generation, or measured data points), the Gaussian process follows the data, and its mean is unaffected by the regression basis, while the variance can be reduced by increasing the data. However, in extrapolation, the Gaussian process follows the regression, deviating from the trend of the data at a rate depending on the “roughness” coefficient, ω , and the solution adequacy thus depends on the regression basis, for both emulation and calibration.

Figure 3.12 shows calibrated prediction for four cases, using the same data and priors (for the calibration) as the example in Section 3.4. The first frame shows the same solution as the first frame of Figure 3.9, extending the extrapolation further for illustration. In this case, the calibrated mean eventually follows the path $y = E(\theta)x^2 + E(\beta)x$ where $E(\theta) = 1.31$ and $E(\beta) = 0.50$. The rate at which the calibrated mean approaches this asymptote is determined by $E(\frac{1}{\sqrt{\omega}})$, the length scale of the model inadequacy Gaussian process, in this case, 0.50. The second frame shows the equivalent solution when an emulator is used to represent the computer program, solving equations (3.9), taking a regression basis $(1, x, \theta)$ for the model inadequacy, and priors $\omega_1 \sim N(\cdot)$, $\sigma_1^2 \sim N(\cdot)$. It can be seen that the asymptote for the calibrated mean is linear, following $y = E(\beta_{1,1}) + E(\beta_{1,\theta})\theta + (E(\beta_{1,x}) + E(\beta_2))x$,

where

$$E(\beta_{1,1}) = -0.68$$

$$E(\beta_{1,\theta}) = 0.69$$

$$E(\beta_{1,x}) = 2.28$$

$$E(\beta_2) = 0.60$$

This behaviour is even more marked in the third frame, where the emulator regression basis is taken as 1 instead of $(1, x, \theta)$. The asymptote for the calibrated mean is $y = E(\beta_{1,1}) + E(\beta_2)x$,

where

$$E(\beta_{1,1}) = 1.59$$

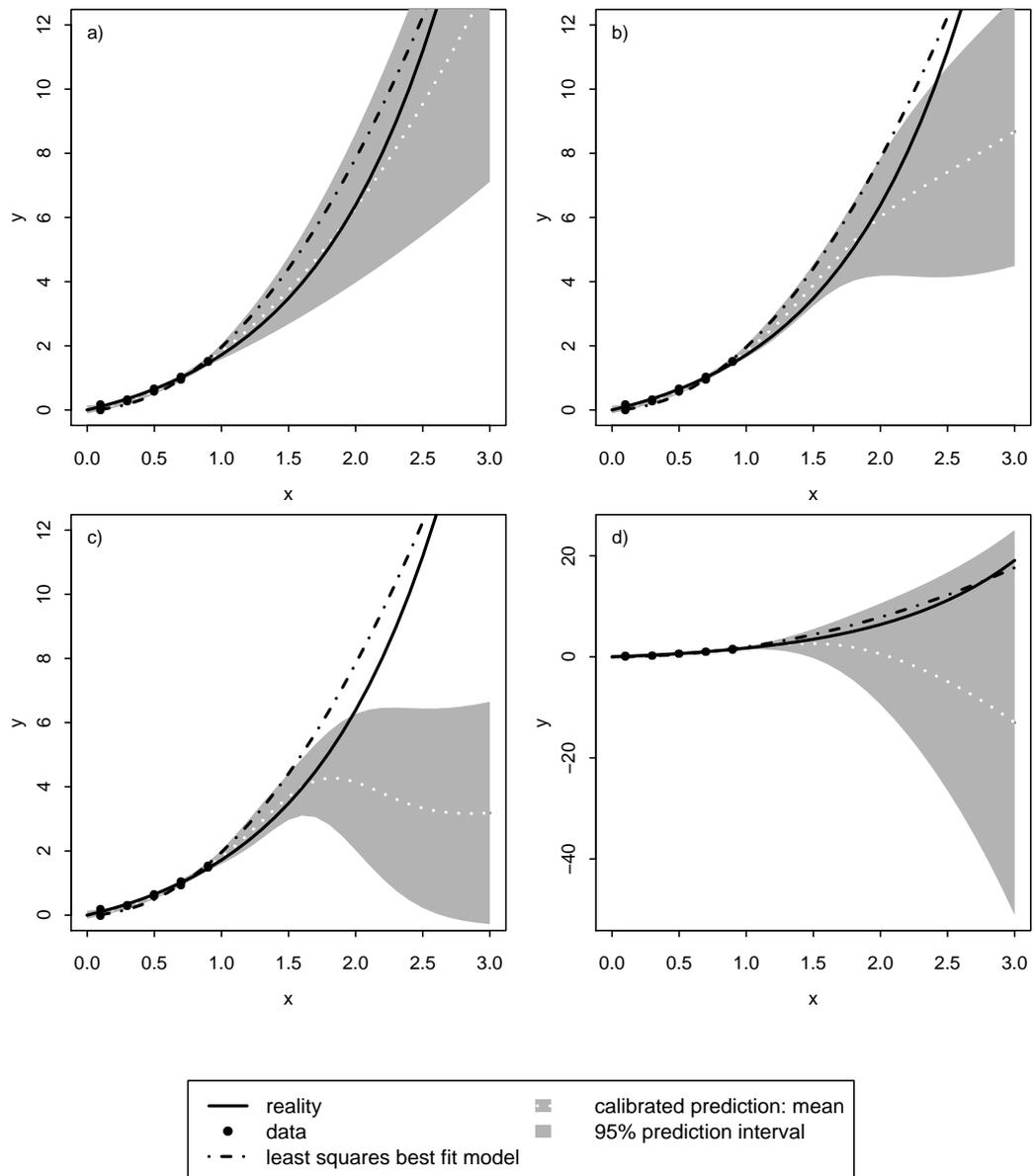


Figure 3.12: Effect of different emulator and model inadequacy regression bases on the calibrated prediction in extrapolation: a) calibration only, model inadequacy basis (x) ; b) emulator basis $(1, x, \theta)$, model inadequacy basis (x) ; c) emulator basis (1) , model inadequacy basis (x) ; d) emulator basis $(1, x, \theta)$, model inadequacy basis (x^3)

$$E(\beta_2) = 0.58.$$

It should be noted that the calibrated mean is just starting to approach the direction of the asymptote at the right hand side of the frame. The rate of approach depends on the size of the hyperparameters ω_1 and ω_2 ; the smaller these are, the more slowly the calibrated prediction will approach its asymptote. In this case $E(\omega_1) = 1.83$ and $E(\omega_2) = 12.61$; Figure 3.5 indicates that the corresponding variability of the Gaussian process is not large.

The fourth frame in Figure 3.12 shows the effect of an inappropriate regression basis for the model inadequacy. Although the calibrated prediction is almost identical to that of the other three cases presented here within the range of the data, in extrapolation the calibrated mean departs from reality, while the prediction limits increase rapidly.

3.5 Summary

The methodology of Kennedy and O’Hagan (2001a, 2001b) has been presented, for the emulation and calibration of computer models using Gaussian process description of the emulator and model inadequacy. The equations have been set out for computer model calibration both with and without use of an emulator. The method has been demonstrated in terms of a simple algebraic example, firstly without emulator, and some practical issues have been raised over the identifiability of the model inadequacy function and observation errors. Finally, an indication has been given of the practical issues involved in the modelling choices involved in the additional use of an emulator. In the next three chapters, the methodology will be applied to three practical examples of flood models, demonstrating further the capabilities and difficulties of the method.

3.6 Appendix: Use of BACCO to estimate a Gaussian process emulator

The method used in the BACCO code (Hankin, 2005) to estimate a Gaussian process emulator is due to Oakley and O’Hagan (2002). They take as a model for the computer output d , observed at n input values $x = (x_1^T, \dots, x_n^T)^T$, a Gaussian process, where

$$\begin{aligned} \eta(x) &\sim N(m(x), V(x, x')) \text{ with} \\ m(x) &= H^T \beta, \\ H &= \begin{pmatrix} h(x_1) \\ \vdots \\ h(x_n) \end{pmatrix} \\ V(x, x') &= \sigma^2 c(x, x'), \text{ and} \\ c(x, x') &= \exp(-(x - x')^T \Omega (x - x')), \end{aligned}$$

where Ω is a positive semidefinite matrix, and other variables were defined earlier in the chapter. Further, taking a prior on β and σ^2 as

$$p(\beta, \sigma^2) \propto \sigma^{-\frac{1}{2}(r+q+2)} \exp\left(-\frac{(\beta - b)^T B^{-1}(\beta - b) + a}{2\sigma^2}\right),$$

where q is the dimensionality of β and a, r, b and B are to be determined by expert elicitation, they show that at an unobserved input value x , the computer model output

has distribution

$$\frac{\eta(x) - m^*(x)}{\hat{\sigma} \sqrt{c^*(x, x')}} |d, \Omega \sim t_{r+n},$$

a multivariate t-distribution where

$$m^*(x) = h(x)^T \hat{\beta} + t(x)^T A^{-1} (d - H \hat{\beta})$$

$$c^*(x, x') = c(x, x') - t(x)^T A^{-1} t(x') + \{h(x)^T - t(x)^T A^{-1} H\} B^* \{h(x)^T - t(x)^T A^{-1} H\}^T$$

$$t(x)^T = (c(x, x_1), \dots, c(x, x_n)), \quad H^T = (h^T(x_1), \dots, h^T(x_n)),$$

$$A = \begin{pmatrix} 1 & c(x_1, x_2) & \cdots & c(x_1, x_n) \\ c(x_2, x_1) & 1 & & \vdots \\ \vdots & & \ddots & \\ c(x_n, x_1) & \cdots & & 1 \end{pmatrix},$$

$$\hat{\beta} = B^* (B^{-1} z + H^T A^{-1} d), \quad \hat{\sigma}^2 = \frac{a + b^T B^{-1} b + d^T A^{-1} d - \hat{\beta} (B^*)^{-1} \hat{\beta}}{n + r - 2}$$

$$B^* = (B^{-1} z + H^T A^{-1} H)^{-1}, \quad d^T = (\eta(x_1), \dots, \eta(x_n))$$

The specification then depends on an estimate for the matrix of “roughness” coefficients, Ω . Assuming a single scale, ω , so that $\Omega = \omega I_q$, this can be found by maximising the likelihood,

$$\mathcal{L}(\omega|y) = \hat{\sigma}^{-\frac{(n-q)}{2}} |A|^{-\frac{1}{2}} |H^T A^{-1} H|^{-\frac{1}{2}}$$

The standard weak prior $p(\beta, \sigma^2) \propto \sigma^{-2}$ (Jeffreys, 1961), is a special case of the above, taking $B = 0$, $a = 0$ and $r = -q$. This is used in BACCO. However, in the case where the above prior on β and σ^2 is not appropriate, or the problem cannot be specified so that the variance is described by a single scale over different dimensions, the above semi-analytical description is not applicable, and the parameters of the emulator are found by Markov chain Monte Carlo.

Chapter 4

Calibration of steady state laboratory experiments

4.1 Introduction

The previous chapter outlined a methodology for Bayesian calibration of computer models, and demonstrated this in the context of a simple algebraic example. This chapter and the following two demonstrate the application of the methodology to hydraulic models. The work described in this chapter concerns the calibration of a steady state flow model using data obtained in large scale laboratory experiments. The model used is Manning's equation, which is a simple algebraic relationship, so that the use of an emulator is not necessary for calibration. In tightly controlled laboratory conditions, measurement error is small, but since the model does not adequately describe the data, a model inadequacy function is required.

The experiments examined the effect of a number of variables on the relationship between flow and depth. These have been introduced successively into the formulation of the model inadequacy, requiring the development of increasingly complex regression basis functions.

4.2 Experimental data

The Flood Channel Facility (Knight and Sellin, 1987) was established during the 1980s and 1990s at Wallingford as a cooperative venture between a number of University departments, jointly funded by SERC and Hydraulics Research Ltd at Wallingford, with a view to improving the understanding of the hydraulic processes involved in out-of-bank river flows. Three series of experiments were undertaken; on rigid straight channels, on rigid meandering channels, and on straight channels with mobile boundaries. This study is concerned with modelling the first of these experimental series.

The experimental flume was constructed in concrete, 56m long, and with total width 10m (Figure 4.1, downloaded from <http://www.flowdata.bham.ac.uk/fcfa-photos.shtml>). The central channel was trapezoidal, with a depth of 0.15m, and a fixed gradient of nominally 1 in 1000. Horizontal floodplains were allowed on both sides of the channel, and floodplain widths, and channel and floodplain sideslopes, were able to be varied. In addition, it was possible to vary the roughness of the floodplain by the insertion of obstacles. The facility was instrumented to measure water levels, discharge, boundary shear stress, velocity profiles and turbulence.

A single test series was undertaken of in-bank flow, consisting of measurements for flows at a number of different levels. For out-of-bank flow, several test series were measured. In this context, the most convenient to use were series involving variation of the floodplain width, and the slope of the channel sides.



Figure 4.1: View of the Flood Channel Facility experimental setup.

4.3 Hydraulic model

The model used in this analysis is Manning's equation:

$$v = \frac{1}{n} R^{2/3} S^{1/2} \quad (4.1)$$

where

v is the cross-sectionally averaged velocity

R is the hydraulic radius, which is normally taken to be the ratio between flow cross-sectional area and the wetted perimeter

as a function of water height, or stage, there is a discontinuity at bankfull height, $y = h$, where a small increase Δy in y beyond $y = h$ gives an instantaneous increase in flow cross-sectional area of $\Delta y(b_w - b - h s)$. This corresponds to the discontinuity in channel width, in addition to the unrealistic assumption that the depth over the floodplain can be infinitesimally small. However, while results of experiments (Figure 4.3a) do show a discontinuity, it is considerably smaller than that predicted by the model, indicating that the model is inadequate to describe this behaviour. The error between the measured data and the model output, fitted by least squares fit to the in-bank data, is shown in Figure 4.3b, illustrating that the greatest discrepancy between the model and data occurs when the flow is just out of bank, as might be expected since at low depths, floodplain resistance has the greatest impact.

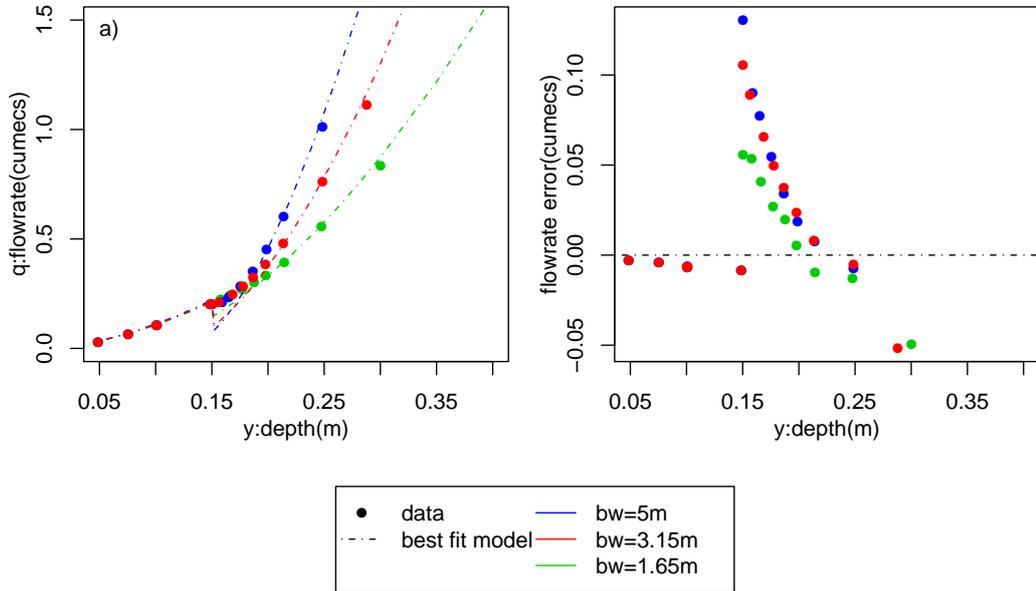


Figure 4.3: Comparison of measured data with model predictions: a) least-squares best fit of Manning’s equation for different floodplain widths b) error in measured data, by comparison with least squares fit to inbank data

4.4.1 Variation of flow with depth

Recall that the data model being used is

$$z = M(x, \theta) + \delta(x) + \epsilon$$

where the observed data z are related to the output of the model $M(x, \theta)$ at input x with parameters θ , subject to errors caused by model structural inadequacy δ , and observation ϵ .

Recall that the model inadequacy $\delta(x)$ is described as a Gaussian process,

$$N(h_\delta(x)^T \beta_\delta, V_\delta(x, x'))$$

Referring to Figure 4.3b, it seems reasonable to describe a model inadequacy function separately for the in-bank and out-of-bank flows. The data were initially examined for a single floodplain width ($b_w=3.15\text{m}$). Apart from the data points close to depth 0.25m (for which no explanation is available from the descriptions of the experimental results), it seems reasonable to take a linear regression model for the out-of-bank flows; thus $(1, y-h)$ was taken for the regression basis, defined for $y > h$ only, effectively making the assumption that the model is unbiased for in-bank flow. In addition, a smooth covariance was assumed for the flow values as a function of depth; $V_\delta(y, y') = \sigma_\delta^2 \exp(-\omega_{\delta_y} |y - y'|^2)$. It was assumed that the in-bank and out-of-bank flows were uncorrelated; however, σ_δ^2 and ω_{δ_y} were assumed to have the same value for both in-bank and out-of-bank flows.

Calibration was performed for a single floodplain width ($b_w=3.15\text{m}$), using Equations (3.8), that is using the model directly without an emulator. Calculation was performed using the logs of all variables, including the parameter θ (Manning's n), and distributional priors used are given in Table 4.1. Note that the regression parameters β_1 and β_2 have improper uniform priors. The table also shows the moments of the posterior distributions. It should be noted that in this analysis, the depth range was not scaled to $[0, 1]$. Thus, the mean of the hyperparameter ω_{δ_y} should be scaled by a factor of approximately 0.06 to be compatible with the reference curves in Figure 3.5.

Table 4.1: Prior and posterior distributions for calibration with respect to depth

variable	prior distribution	posterior mean	posterior standard deviation
$\log(\theta)$	$N(-4.6, 1.5^2)$	-4.7	0.12
β_1	$U(-\infty, \infty)$	0.12	0.04
β_2	$U(-\infty, \infty)$	-1.6	0.89
$\log(\omega_{\delta_y})$	$N(5, 2^2)$	5.6	0.66
$\log(\sigma_\delta^2)$	$N(-7, 2^2)$	-7.3	1.0
$\log(\sigma_\epsilon^2)$	$N(-12, 3^2)$	-13.2	1.0

The posterior distribution of the parameter Manning's n is symmetric, with mean 0.0094, and standard deviation 0.0011. Calibrated prediction of the input data is shown in Figure 4.4, where frame (a) shows the calibrated prediction, and frame (b) shows the difference between the calibrated prediction and the output of Manning's equation, using the value of Manning's n found by least squares fit to the in-bank data, permitting a more detailed scrutiny of the prediction. Like the data, the calibrated mean shows a small apparent discontinuity at the bankfull level; however, the prediction interval, which is elsewhere except in extrapolation extremely narrow, reflecting low errors in laboratory experiment output, shows a local increase at the bankfull level, indicating some uncertainty in the correction for the step model. This is clearly related to the large posterior "roughness" hyperparameter. The calibrated mean in extrapolation continues the trend of the data, but the prediction interval increases rapidly, reflecting similar behaviour seen in the toy example in Figure 3.12.

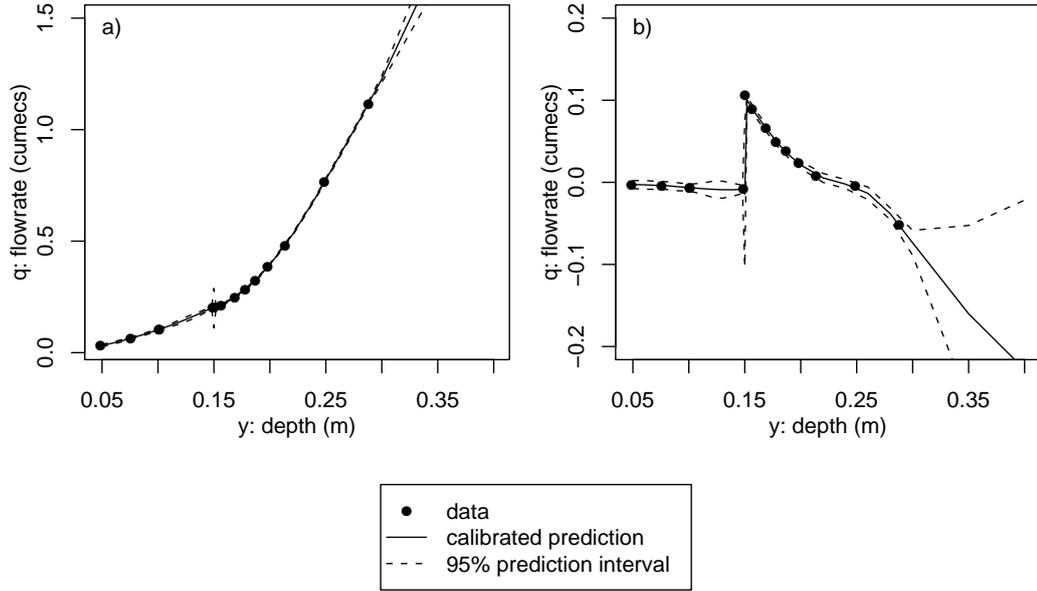


Figure 4.4: Bayesian calibrated prediction using data for a single floodplain width using a model inadequacy: a) calibrated prediction; b) difference between calibrated prediction and Manning's equation, fitted to the in-bank data.

4.4.2 Variation of flow with depth and floodplain half-width

If the discharge q is taken as a function of both y (height) and b_w (full channel half width), it is possible to use all of the data in Figure 4.3a. Calibration was performed using two data sets, in order to predict the third. This is a potentially problematic formulation, as the b_w -dependence is estimated on only two values.

As in the case for a single floodplain width, the model inadequacy regression basis was defined only for out-of-bank depths. Three basis functions were taken, $(1, (y - h), b_w)$, defined for $y > h$. It is not clear whether this is an appropriate regression basis, since Figure 4.3b does not give the impression that the dependence on b_w is linear, and regression analysis bears this out. However, with so few data series, it is difficult to choose an appropriate measure for the dependency.

As in the previous example, the covariance was defined separately for in-bank and out-of-bank flows. Two covariance relationships were needed;

$$V_{\delta}((y, b_w), (y', b'_w)) = \sigma_{\delta}^2 \exp(-\omega_{\delta_y}(y - y')^2), \quad \text{for in-bank flows, and} \quad (4.2a)$$

$$V_{\delta}((y, b_w), (y', b'_w)) = \sigma_{\delta}^2 \exp(-\omega_{\delta_y}(y - y')^2 - \omega_{\delta_{b_w}}(b_w - b'_w)^2), \quad \text{for out-of-bank flows.} \quad (4.2b)$$

As in the previous example, the covariance between in-bank and out-of-bank flows was assumed to be 0.

Two calibrations were performed, in each case using two of the out-of-bank datasets, to predict the results of the third dataset. In the first case, calibration data corresponded to

flood plain half-width of 1.65m and 5m, and calibrated prediction was made for $b_w=3.15$ m. In the second case, calibration undertaken with data measured at $b_w=1.65$ m and $b_w=3.15$ m was used to predict the relationship $q(h)$ at $b_w=5$ m.

Prior and posterior distributions are given in Table 4.2. The same prior distributions were taken as in the single dimensional case, with the addition of a prior on $\log(\omega_{\delta_{b_w}})$; however, it was found that while a reasonably vague prior $\log(\omega_{\delta_{b_w}}) \sim N(5, 2^2)$ was adequate in the case where the validation data corresponded to $b_w=5$ m (case (b) in Table 4.2), a tighter prior was required where there was a larger interval between the floodplain width calibration data (case(a)), to avoid a bimodal posterior distribution, a possible outcome for a large enough posterior “roughness” in the b_w -dimension.

Table 4.2: Prior and posterior distributions for calibration with respect to depth and floodplain half-width

variable	prior distribution case(a)/case(b)	posterior mean case (a)	posterior st. deviation case (a)	posterior mean case (b)	posterior st. deviation case (b)
$\log(\theta)$	$N(-4.6, 1.5^2)$	-4.6	0.07	-4.5	0.04
β_1	$U(-\infty, \infty)$	0.06	0.04	0.04	0.02
β_2	$U(-\infty, \infty)$	0.02	0.009	0.02	0.006
β_3	$U(-\infty, \infty)$	-0.72	0.53	-0.22	0.27
$\log(\omega_{\delta_y})$	$N(5, 1^2)/N(5, 2^2)$	5.5	0.64	6.1	0.68
$\log(\omega_{\delta_{b_w}})$	$N(-3, 2^2)$	-5.8	0.84	-3.0	0.80
$\log(\sigma_{\delta}^2)$	$N(-7, 2^2)$	-7.1	0.83	-8.0	0.75
$\log(\sigma_{\epsilon}^2)$	$N(-12, 3^2)$	-10.8	0.55	-112.9	0.55

In both cases the posterior distribution of θ is fairly symmetric, with a mean of 0.01, and small standard deviation. Calibrated prediction is shown in Figure 4.5. In both cases, there is good agreement between the calibrated prediction for the unseen floodplain width and the validation data (coloured points). The results of the calibrated prediction show a slightly wider prediction interval than in the single-dimensional case, and a bias in the extrapolation case (Figure 4.5b) of the mean prediction, just above the bankfull level. Interestingly, the spike in the prediction uncertainty at bankfull is smaller than the one-dimensional example, while the uncertainty in in-bank prediction is slightly larger.

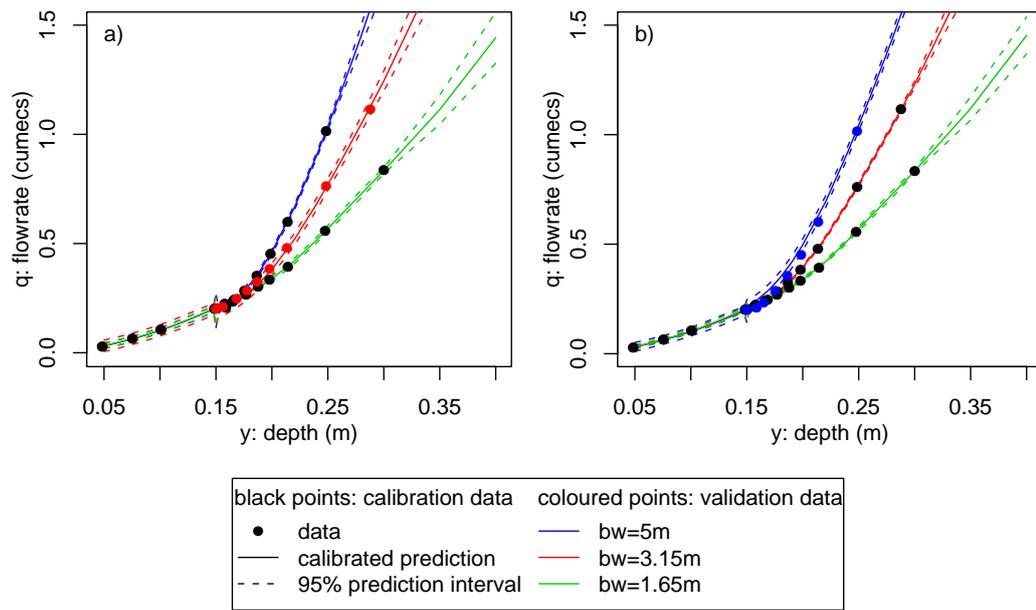


Figure 4.5: Calibrated prediction: calibration undertaken using data from two floodplain widths, and comparing the calibrated prediction with the third. Validation dataset a) $b_w=3.15\text{m}$ b) $b_w=5\text{m}$.

4.4.3 Variation of flow with depth, floodplain half-width, and channel side slope

A further extension to the calibration can be made with inclusion of experiments undertaken varying channel sideslopes. There is a problem in doing this, caused by the design of the original experiments. The design is shown in Figure 4.6. While the design permits assessment of variation of flow with respect to changes in floodplain width, or of channel bank slope, it is not possible to assess their effect simultaneously. Thus, while it is possible to undertake the calibration with respect to both variables, nothing can be said about any interaction which may exist between these variables.

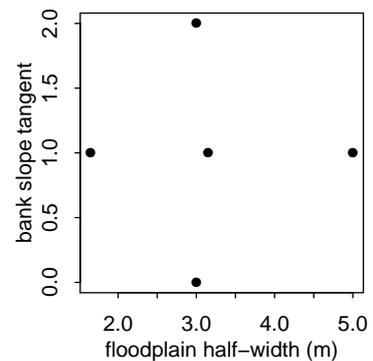


Figure 4.6: FCF straight channel experimental design

In addition, while it may be expected that variation of the river channel sideslope may alter the inbank flow as well as the out-of-bank flow, measurements were not taken of the inbank flow for different sideslope values. The solution for this exercise, was to take the inbank measurements for the middle sideslope value, and attribute them to all sideslope values.

The model inadequacy regression basis was again defined for out-of-bank flow only, as

$(1, y - h, b_w, s)$, where s is the tangent to the sideslope. As before, the covariance structure for the model inadequacy was considered to have two parts, relating to in-bank and out-of-bank flow; these parts of the model inadequacy were assumed to be uncorrelated. The covariance structure for out of bank flow was taken to be the smooth structure

$$V_\delta((y, b_w, s), (y', b'_w, s')) = \sigma_\delta^2 \exp(-\omega_{\delta_y}(y - y')^2 - \omega_{\delta_{b_w}}(b_w - b'_w)^2 - \omega_{\delta_s}(s - s')^2)$$

Two possible covariance structures were considered for inbank flow:

$$V_\delta(y, b_w, s), (y', b'_w, s')) = \sigma_\delta^2 \exp(-\omega_{\delta_y}(y - y')^2), \quad \text{and} \quad (4.3a)$$

$$V(\delta(y, b_w, s), (y', b'_w, s')) = \sigma_\delta^2 \exp(-\omega_{\delta_y}(y - y')^2 - \omega_{\delta_s}(s - s')^2) \quad (4.3b)$$

Calibration of the Manning equation was performed for both of these covariance structures, in order to assess their effect on the results. The prior distributions (Table 4.3) were the same as those used in the previous case, with the tighter prior distribution for $\omega_{\delta_{b_w}}$ and a similar prior distribution for ω_{δ_s} .

Table 4.3: Prior and posterior distributions for calibration with respect to depth, floodplain half-width and channel side slope tangent

variable	prior distribution	posterior mean case (a)	posterior st. deviation case (a)	posterior mean case (b)	posterior st. deviation case (b)
$\log(\theta)$	$N(-4.6, 1.5^2)$	-4.6	0.08	-4.6	0.08
β_1	$U(-\infty, \infty)$	0.011	0.058	0.012	0.056
β_2	$U(-\infty, \infty)$	0.021	0.014	0.020	0.013
β_3	$U(-\infty, \infty)$	-0.15	0.65	-0.19	0.64
β_4	$U(-\infty, \infty)$	0.015	0.016	0.015	0.015
$\log(\omega_{\delta_y})$	$N(6, 2^2)$	5.5	0.71	5.5	0.69
$\log(\omega_{\delta_{b_w}})$	$N(-3, 1^2)$	-2.8	0.55	-2.8	0.57
$\log(\omega_{\delta_s})$	$N(-3, 1^2)$	-2.2	0.59	-2.3	0.72
$\log(\sigma_\delta^2)$	$N(-7, 2^2)$	-6.1	0.84	-6.2	0.78
$\log(\sigma_\epsilon^2)$	$N(-12, 3^2)$	-11.6	0.56	-11.6	0.58

It can be seen from Table 4.3 that the calibration of these two models appears identical. The posterior distributions of θ are similarly almost identical, with mean 0.01 and standard deviation 0.0009. It is not surprising, then, that in general the calibrated prediction of the input data for both models is at first sight identical. Indeed, the calibrated prediction for the entire range has only been shown for the first case, for floodplain halfwidth $b_w=3\text{m}$, in Figure 4.7a_i). The effect of the different floodplain channel slopes on the out-of-bank flow is merely a translational effect of the relationship.

However, close inspection of the calibrated prediction of the input data for the two models shows a substantial difference in the prediction interval in the inbank flow (Figure 4.7a_{ii} and b), with the larger prediction interval coinciding with a dependency in the covariance structure on the additional variable, s . The difference in prediction interval cannot be resolved without additional data.

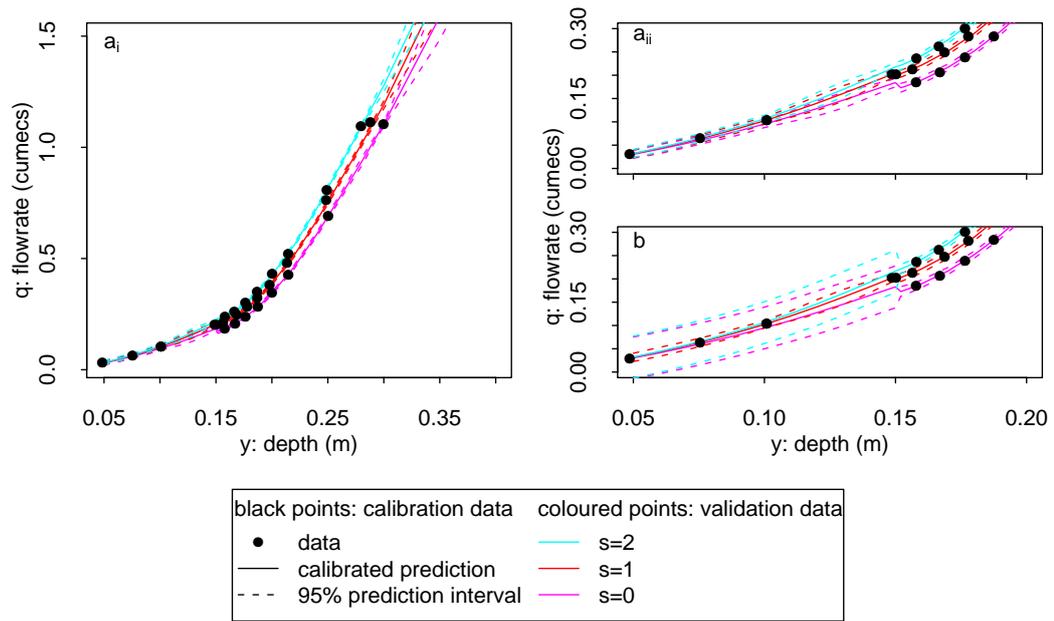


Figure 4.7: Calibrated prediction for floodplain half-width $b_w=3\text{m.}$, comparing the effect of in-bank covariance structures in Equations: a_i) and a_{ii}) (4.3a) and b) (4.3b)

4.4.4 Comment

During the development of these models, the β parameters were not integrated out of the equations, and the solution was found by integrating the full equations (3.5) including the regression coefficients explicitly. The posterior distributions were much more difficult to simulate from, and in particular were sensitive to redundancy in the regression basis for the model inadequacy mean, and errors in the specification of the covariance structure. For example, in the case where the flow q was modelled as a function of depth y and floodplain half-width b_w , regression analysis of the out-of-bank measurements also indicated dependence on the product, $(y-h)b_w$, as well as on the single variates $(1, (y-h), b_w)$. However, inclusion of this term in the basis led to non-identifiability of the parameter Manning's n , whose posterior distribution then became dependent entirely on the prior distribution. Further investigation of the input data locations in the (y, b_w) plane showed that the product $b_w(y-h)$ is highly correlated with flow depth $(y-h)$ (correlation 0.92), indicating that this fourth term in the regression basis is redundant.

4.5 Taking into account experimental errors

4.5.1 Observation errors in the experimental programme

The foregoing analysis has been undertaken without reference to the experimental errors reported in the measurement of the data. Although the experimental programme was more complex, the two quantities used in this analysis are water height above the channel base,

and total discharge. Other measurements implicit in this analysis are those of geometry of the experimental rig, and their errors are assumed to be small.

Myers and Brennan (1990) reported that the water depth was measured by digital gauges over stilling wells connected to tapping points in the channel bed, and reading to the nearest 0.01mm. In addition, they reported that discharge was measured by orifice plate meters installed in the mains supplying the upstream end of the model, although they did not give a value for the accuracy of this measurement. It is however clear from the reports of these experiments, both by Myers and Brennan, and by other authors, that the primary adjustable quantity in the experiments was the flowrate, and that other quantities were measured and validated with respect to this quantity. This understanding led to the definition of a very small prior for the variance of the flowrate data in the above analysis.

4.5.2 Including input errors in the calibration

While in practice the input errors in the experimental program analysed in this chapter are small, the theory is discussed below for the inclusion of these errors in the Bayesian formulation.

The formulation used in this analysis is to take discharge as a function of water depth, where the main measurement errors are in the independent variable, water depth. The model of Kennedy and O'Hagan (2001a), and indeed of Craig et al. (2001) and Goldstein and Rougier (2004, 2006) do not allow for errors in the independent variable in the calibration. This is a simplification of reality. Clearly, no input errors are required in emulator formulation. However, in calibration both the input and output data are measured, and are thus subject to measurement errors. Given calibration under the assumption of no input errors, Kennedy and O'Hagan (2001b) allow for uncertainty in the input data on the calibrated prediction, but focus on the ability to achieve an analytical solution for the mean and variance of properties of the distribution of $z(x)$.

However, for the purposes of finding a probability of inundation, the full distribution of the process is required. Expressions were given in Equations (3.10) for the expectation $E(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2)$ and variance $\text{Var}(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2)$ of the process, which are Normal, conditional on the parameter value. If however further calibrated output is required, for input whose measured value is x^\dagger , but whose actual value is uncertain, then the required distribution, rather than being $\zeta(x^\dagger)$, is in fact $\zeta(X|x^\dagger)$, where X is a random variable, with distribution $\zeta(X|x^\dagger)$. Then, assuming conditional Normality of the calibrated prediction so that the distribution $f(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2)$ is determined by Equation (3.10), the output value is found by integrating this conditional distribution with respect to $f(X|x^\dagger)$, before summing over the Markov chain to arrive at a posterior distribution.

$$E(\zeta(X|x^\dagger, z)) = \iiint_{\theta, \psi_\delta, \sigma_\epsilon^2} \left\{ \int_X f(\zeta(X|z, \theta, \psi_\delta, \sigma_\epsilon^2)) f(X|x^\dagger) dX \right\} f(\theta, \psi_\delta, \sigma_\epsilon^2) d\theta d\psi_\delta d\sigma_\epsilon^2$$

It should be noted, however, that if the input errors are significant, they should be taken into account in the calibration. This is the situation which has been addressed in the BATEA project (Kuczera et al., 2006, Kavetski et al., 2006, Thyer et al., 2009, Renard et al., 2010), giving rise to a large numbers of unknowns to be solved for, as this involves an unknown variable for each input measurement used. As has been noted previously, Renard et al. (2010) have found that the input and structural errors are unidentifiable when both are specified in a rainfall-runoff model.

Calibration in presence of input errors is more complicated than under the assumption of no input errors. Starting from Equation (1.4),

$$z = M(w + e, \theta) + \delta(w + e) + \epsilon$$

the data model now contains additional unknowns to be estimated. Besides the parameters θ , the hyperparameters for the model inadequacy function δ and the output observation noise ϵ , there are also not only hyperparameters for the input error distribution, e , but the input error values themselves at each measurement location w . The expression for the Gaussian process for the model inadequacy becomes more complicated;

$$\delta(w + e) \sim N\left(h(w + e)\beta, V((w + e), (w' + e'))\right)$$

If it is assumed that the errors on the input are small relative to the differences between the input values, then the covariance could be approximated by $V(w, w')$. It would not be feasible to integrate out the regression parameters β , necessitating their estimation by MCMC simultaneously with the other variables.

4.6 Summary

The work described in this chapter concerns the calibration of simple steady state flow models using laboratory data, as a simple application of the Kennedy and O'Hagan calibration methodology. The nature of the model inadequacy necessitated a correction function defined over part of the input domain. By including the output of successively larger numbers of experiments, it has been possible to develop increasingly complex regression bases.

The treatment of input errors has been discussed, although not demonstrated.

The following two chapters concern the calibration of flood models, using historical flood data, where it can be expected that the data contain greater errors, and the inadequacies of the model structure are larger and more complex.

Chapter 5

Calibration of a steady state flood model

5.1 Introduction

The work in this and the following chapter concerns the calibration of fluvial flood models using historical flood data. The models used are the hydraulic models described in Chapter 2, which solve the partial differential equations representing flow in the channel and floodplain. Although the models used in this thesis are simple, and thus inadequate to describe the flow in geometrically complex channels, they take sufficiently long to execute that emulators are required to make the calibration manageable. The data too, measured under flood conditions, can be expected to contain significant errors.

This chapter concerns the calibration of a steady state model of a flood on the river Thames at Buscot, in December 1992. The flood coincided with an overpass of the ERS-1 remote sensing satellite, so calibration can be undertaken with reference to the resulting SAR image, which gives a map of the flood extent. The flood has been modelled, using steady state simulations, with the LISFLOOD-FP package, giving depths across the floodplain which can be compared with the satellite image.

The results obtained in this chapter were found using the MCMC formulation described before. However, as in Chapter 4, demonstration of the emulator has been obtained using the BACCO computer package (Hankin, 2005), which has implemented the emulator formulation of Oakley and O'Hagan (2002) in R (R Development Core Team, 2009); these equations were detailed in Section 3.6. Much of the work in this chapter has been reported in Hall et al. (2011), although the calibration results there were achieved using BACCO. It is thus possible to compare the results of the full calibration using MCMC, with those achieved using optimisation to find the hyperparameters, as described by Kennedy and O'Hagan (2001b) and implemented in BACCO.

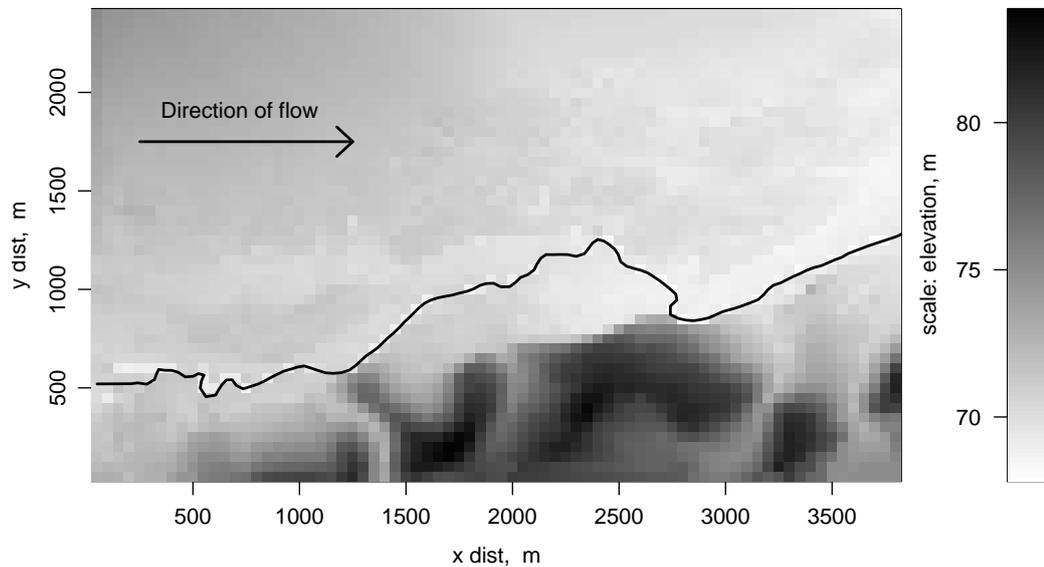


Figure 5.1: Digital elevation map of the study region, with river course superimposed; direction of flow, left to right

5.2 Flood model

The LISFLOOD-FP package, used to model the flood, was described in Chapter 2. The model has been applied to a region of 3.8 x 2.4 km, with a low-lying floodplain on the left bank. The river reach is bounded upstream by a gauged weir, and reasonably well-contained at the downstream end. Thus, upstream discharge was taken as the value corresponding to the gauged flow at the time of the satellite overpass. Downstream outflow is given by the weir equation, based on the slope of the given bed dimensions.

A 50km resolution DEM was used (Figure 5.1) to describe the floodplain topography, with a vertical accuracy of 0.25m. The channel depth and width are assumed uniform in the model; these and other geometrical values were provided by the Environment Agency. The model was set up by Aronica et al. (2002), who point out that a dynamic simulation was unnecessary, in view of the short reach and broad hydrograph involved.

The parameters to be determined through calibration are the roughness coefficients of the channel and floodplain. Werner et al. (2005), using GLUE, demonstrated a lack of sensitivity to using spatially varying roughness parameters in a floodplain with mixed vegetation, while Hall et al. (2005) used global sensitivity analysis to show that this model is not sensitive to floodplain roughness coefficient. Consequently, calibration has been undertaken with regard to a uniform channel roughness coefficient.

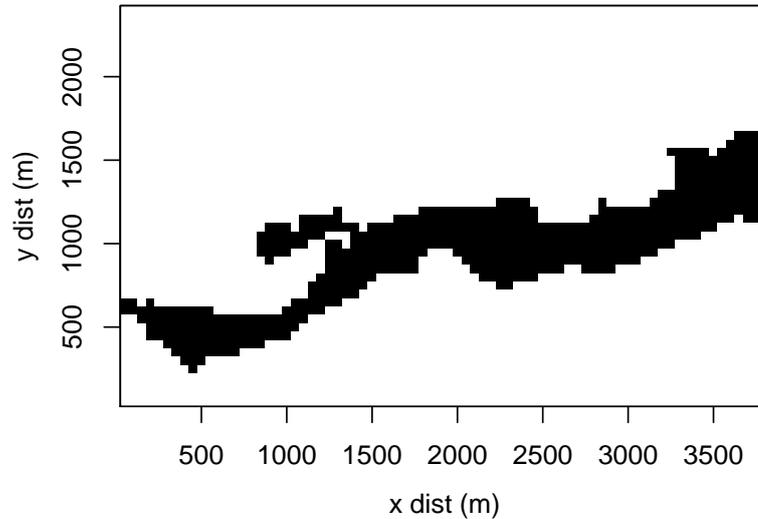


Figure 5.2: Flood extent, obtained from processed SAR data

5.3 Calibration data

The calibration data were provided from radar imagery, and processed by Horritt et al. (2001) to form a binary image of inundated and non-inundated areas (Figure 5.2). Misclassification is possible in some green fields which give a similar low backscatter to the free water surface. Equally, it is possible to misclassify flooded area as non-inundated, where the water surface is wind-roughened, although this was thought to be a lesser problem in this instance.

Given that the output of LISFLOOD-FP is a map of water surface height, there are two ways in which this can be related to the processed SAR data. Aronica et al. (2002) used the binary radar image to calibrate the LISFLOOD-FP model using GLUE, by classifying the model output array as inundated and non-inundated, and devising a measure of goodness of fit of these binary images. This measure is

$$F = \frac{\sum_{i,j} \{(d_{i,j} = 1) \wedge (m_{i,j} = 1)\}}{\sum_{i,j} \{(d_{i,j} = 1) \vee (m_{i,j} = 1)\}}$$

where $(d_{i,j} = 1)$ and $(m_{i,j} = 1)$ are the events that pixel (i, j) is inundated, according to the data and model output respectively. Woodhead (2007) undertook a statistically coherent calibration of the model, using Bayesian techniques and the same binary data image, investigating a number of binary data models, but without considering model structural uncertainty.

An alternative is to identify from the data the water surface height at the flood shoreline, and compare this with the water surface heights predicted by LISFLOOD-FP.

The accuracy with which water heights can be recovered from the superposition of a flood outline onto a DEM depends not only on the resolution and accuracy, but also on

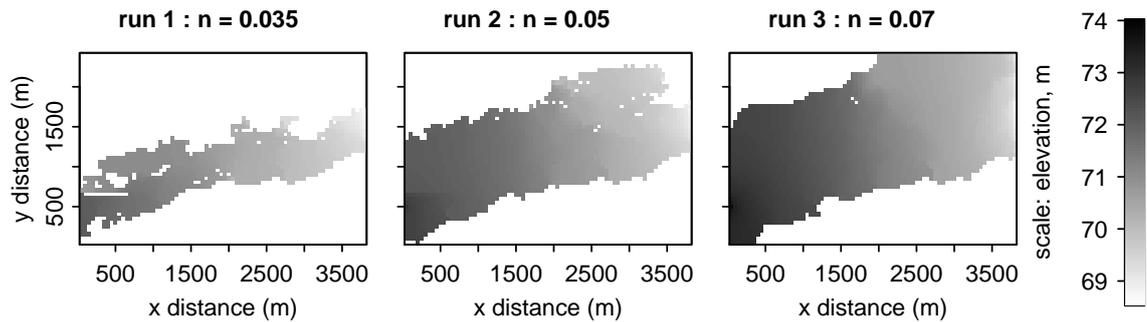


Figure 5.3: Modelled flood extent for different values of Manning's n

the slope of the DEM; in the limit, the bounds to the water height at the shoreline are the height of the DEM at the pixel at the water's edge, and the height of the lowest adjacent non-inundated pixel. The uncertainty on the water height at the shoreline has been reduced (Hall et al., 2011) by incorporating information from an ensemble of 638 model runs with randomly chosen values of Manning's n in the interval $[0.01, 0.1]$. For each shoreline location of the data, those runs in the ensemble were taken whose output wet/dry classification in the nine immediately adjacent pixels matched those of the data. The water height taken for that shoreline point was the mean value from those runs. This technique allows the model to interpolate between regions where the shoreline water height is more and less well-defined. The resulting flood elevations, defined at locations illustrated in Figure 5.4b, are shown in Figure 5.4a.

5.4 Calibration

5.4.1 Emulator construction

Running the model with different values of the roughness parameter gives flood extents both greater and less than those found from the satellite data (Figure 5.3). A small number of the ensemble of 638 model runs was taken, ensuring that the range of their predicted flood extent spans the inundation range found from the satellite data. Figure 5.4c illustrates the water surface profiles at the left bank obtained from these selected runs, which span reasonably well the observations shown in Figure 5.4a.

In order to construct an emulator in (x, y, n) space, a random sample of 40 water surface elevations was taken from these selected runs, at points on the boundaries of the model output flood outline. The projection of these points onto the spatial plane is shown in Figure 5.4d. Note that these points are concentrated near the boundaries of the flooded area rather than being uniformly distributed over the plane. This has been done, recognising that emulator errors increase away from the design points, as it is here that the emulator is required to be most accurate in order to make best use of the observation data and also to generate accurate flood predictions. Note too, that the right bank is steeper than the left bank, so the right shoreline changes little with varying roughness parameter,

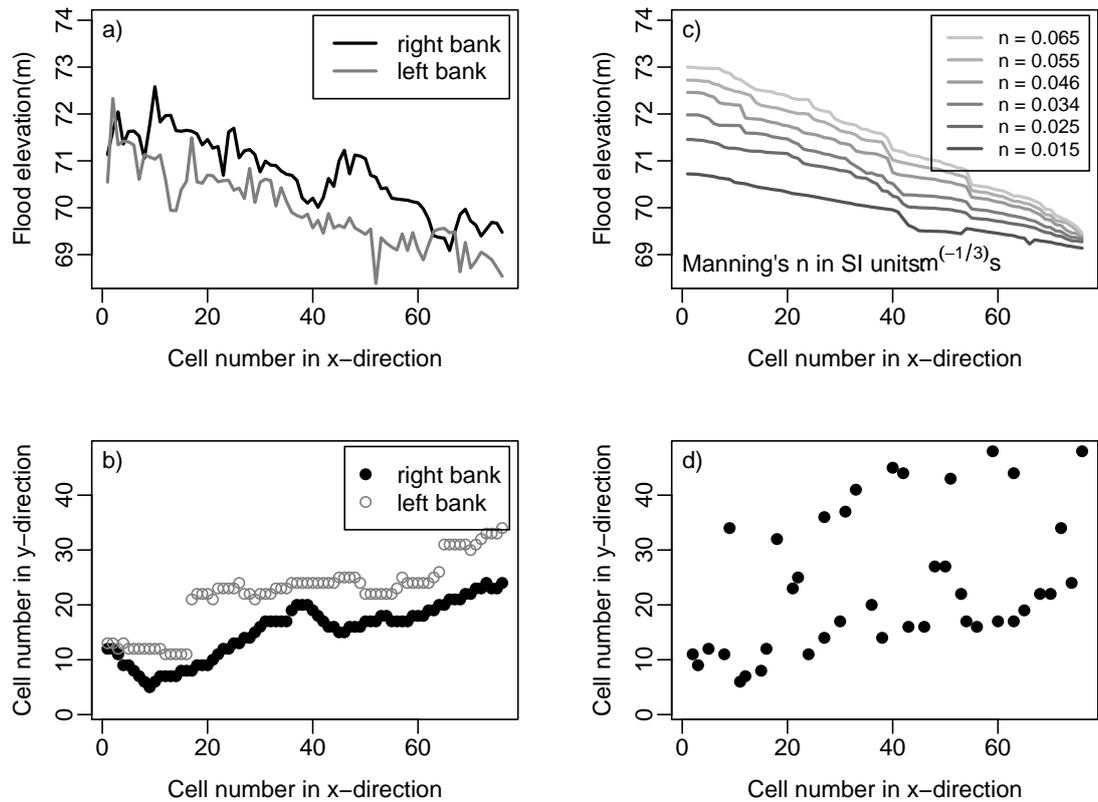


Figure 5.4: LISFLOOD calibration and emulation data a) Flood water elevation inferred from the SAR flood outline image b) Points where flood elevation values were extracted from SAR observations c) Typical LISFLOOD water surface profiles, at different values of Manning's n d) Points used to construct the LISFLOOD emulator

and thus emulator design points appear to be spatially closer together on the right bank. An alternative method for emulator design would have been to have taken a Latin hypercube of training points over the (x, y, n) -plane, extrapolating the water surface, using distance-weighted averaging, from the shoreline to the edge of the domain. This has been tried, but since the water surface is very smooth, has made little difference. The number of design points is limited by the stability of the covariance matrix. If the design points are taken too close together, then the covariance matrix becomes singular. The use of pivoted Choleski decomposition or singular value decomposition permits the number of design points to be increased, firstly since these are robust decomposition methods, and secondly, as an approximation can be made by discarding the smallest eigenvalues. Care should be taken however, as the exclusion of small eigenvalues in the covariance matrix smoothes the solution, which may be undesirable.

Regression analysis of the model output at the emulator design points suggests linear dependence on x , and on the parameter, Manning's n . Dependence on y is not significant. A smooth covariance function has been assumed, reflecting our belief about the water surface.

For illustration, the emulator has been produced using BACCO. Improper vague priors

were used for all variables, and the calculated regression coefficients were $(0.58, -0.78, 0.46)$ in $(1, x, n)$, the optimised emulator variance was 0.011, while the $\log(\text{“roughness”})$ coefficients were $(-5.62, 19.2, -3.94)$ in (x, y, n) -space. Emulator output has been compared with LISFLOOD runs not used in the emulator construction. This is difficult to visualise, as the emulator error is defined as a distribution for every point in the three dimensional (x, y, n) -space. Figure 5.5 shows the absolute error in mean emulator prediction, and the standard deviation of the predicted emulator uncertainty plotted on the spatial plane, compared with output from three LISFLOOD runs not used to train the emulator, whose values of Manning’s n are close to, midway between, and outside the range of the training runs. The scale of the mean error is large, but investigation has shown that this is not due to the choice of training points, but rather to the nature of the model output. While it might be expected that the water surface varies gradually and smoothly along the reach, Figure 5.6 shows the deviation of the model output from this assumption, which is considerable at some locations close to the shore line, or to the edge of the domain. Figure 5.5 also shows that the uncertainty in the emulator prediction increases away from the training runs.

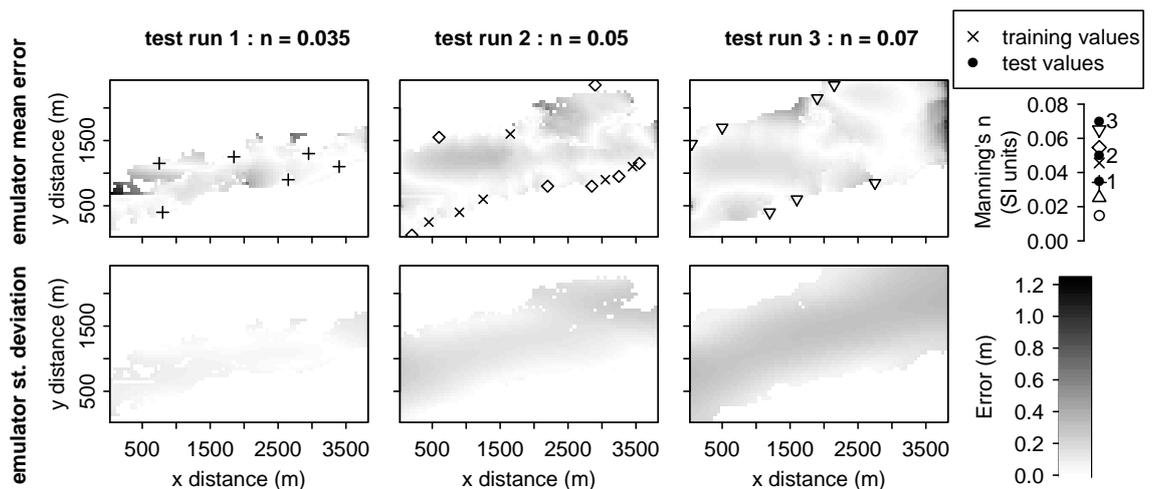


Figure 5.5: Absolute error in mean emulator prediction, and standard deviation of the predicted emulator uncertainty for three different values of Manning’s n : close to, in between, and away from the training runs. Values of test runs and training runs are illustrated

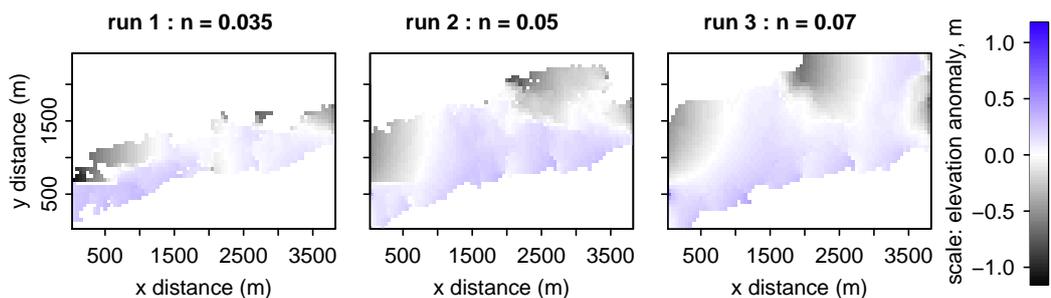


Figure 5.6: Deviation of model output from linear slope

Bastos and O’Hagan (2009) discuss the validation of Gaussian process emulators. Given the correlated nature of the emulator output, they suggest calculating the Mahablanobis distance between the emulator output with the computer model output. Oakley and O’Hagan (2002) note that the emulator errors are distributed as a multivariate t -distribution, with mean $m^*(x)$ and variance $\hat{\sigma}^2 c^*(x, x')$ given in Section 3.6. For a multivariate emulator,

$$\frac{\sqrt{q-p}}{\hat{\sigma}_1 \sqrt{q-p-2}} Q^{-T} (m^*(x, y, n) - M(x, y, n)) \sim t_{q-p} \quad (5.1)$$

where q is the number of points, p is the number of regression parameters, and Q is a matrix such that $QQ^T = C^*$, the correlation matrix defined by $C_{i,j}^* = c^*(x_i, x_j)$, with $c^*(x, x')$ defined in Section 3.6. The simplest formulation for Q is generally the Choleski decomposition of C^* . However, if the points represented in the correlation matrix C^* are too close together, this matrix becomes numerically unstable, so an alternative decomposition was used, determining the eigendecomposition of the matrix C^* using singular value decomposition.

The distribution was found as follows. Taking the ensemble of 638 model runs, a number of draws were made in the (x, y, n) space, in a Latin hypercube, and spaced as far as possible from each other and from the training locations. Those points where the water surface was not above the DEM were then discarded, and the expression on the left of Equation (5.1) was calculated.

In view of the numerical difficulties found in determining Q , it was infeasible to apply this process for more than approximately 100 points at a time, so in order to have a reliable estimate of the distribution, the calculation was repeated until the total sample was 2037 points. The degrees of freedom should, however be related to the number of runs used to generate the emulator, giving a theoretical t distribution with 5 degrees of freedom. The transformed differences have mean 0.051, standard deviation 1.877 corresponding to a t distribution with 3 degrees of freedom (Figure 5.7).

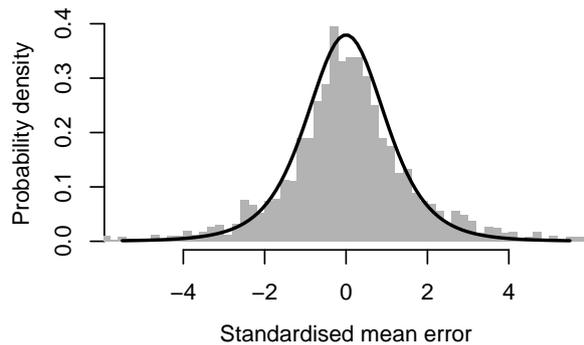


Figure 5.7: Emulator error distribution; with superimposed t distribution with 5 degrees of freedom.

5.4.2 Calibration

In the absence of any evidence pointing towards greater complexity, the chosen regression basis for the model inadequacy, as for the emulator, is a linear spatial and parameter dependence, this time on (x, y, n) , and the covariance is assumed smooth. Note that while the computer model output is that of a water level, when this is above the land surface,

both the emulator and the calibration spread across the entire spatial domain.

Calibration has been undertaken using a subset of the data; initially 26 data points, spread along the shoreline at both sides of the flooded river, and for comparison, with 61 data points. The prior distributions were as shown in Table 5.1. Specification of prior distributions for the noise variance is not straightforward. The Bayesian methodology requires that the prior distributions represent the best knowledge of the technical expert. However, in this case the observation data have been derived by a complex process; first the interpretation of satellite images into binary images, followed by translation into water heights. It was felt that the standard deviation of the error could be considered to be somewhere between 10cm and 25cm. If, however, the assumption were taken, that the water level should vary linearly with distance along the x -direction, the residual variance of the data would imply an error standard deviation of approximately 35cm. Accordingly, four different prior distributions were taken for the observation error, Normal distributions with means corresponding to each of these values, but small standard deviations, and a more vague distribution, which encompassed all possibilities. Bearing in mind the sensitivity found for the posterior prediction interval of the toy example to prior distribution on model inadequacy variance (Table 3.3), this was also investigated, but had little effect on the calibration and calibrated prediction.

Table 5.1: Prior and posterior distributions for calibration

variable	prior distribution	comment	posterior mean and (st. deviation)	
			26 data points	61 data points
n	$N(.0265, .0224^2)$		0.027 (0.012)	0.028 (0.013)
β_1		emulator (const)	0.62 (0.07)	0.61 (0.07)
β_2		emulator (x)	-0.77 (0.09)	-0.78 (0.10)
β_3		emulator (n)	0.44 (0.07)	0.44 (0.07)
β_4		inadequacy (x)	0.33 (0.18)	0.29 (0.23)
β_5		inadequacy (y)	-0.68 (0.32)	-0.63 (0.42)
$\log(\omega_{1x})$	$N(3.5, 2^2)$		2.57 (0.35)	2.44 (0.35)
$\log(\omega_{1y})$	$N(2.5, 2^2)$		2.84 (0.58)	2.88 (0.58)
$\log(\omega_{1\theta})$	$N(2, 2^2)$		1.59 (0.47)	1.46 (0.46)
$\log(\omega_{2x})$	$N(2, 1.4^2)$		2.07 (1.19)	4.13 (0.66)
$\log(\omega_{2y})$	$N(2, 1.4^2)$		2.52 (1.38)	5.11 (0.76)
$\log(\sigma_1^2)$	$N(-4.5, 2^2)$		-4.71 (0.37)	-4.62 (0.40)
$\log(\sigma_2^2)$	$N(-4.5, 1.4^2)$		-5.32 (1.10)	-4.05 (0.49)
$\log(\sigma_\epsilon^2)$	$N(-4.5, 0.1^2)$	mean $\leftrightarrow \sigma=0.35\text{m}$	-6.15 (0.44)	-5.72 (0.42)
	$N(-5.23, 0.1^2)$	mean $\leftrightarrow \sigma=0.25\text{m}$		
	$N(-7.06, 0.1^2)$	mean $\leftrightarrow \sigma=0.10\text{m}$		
	$N(-5.23, 2^2)$	mean $\leftrightarrow \sigma=0.25\text{m}$		

Distributional means and standard deviations for posterior distributions are given in Table 5.1 for the cases with vague prior distributions. Emulation and calibration for both datasets leads to posterior emulator parameterisation very similar to that found by BACCO. The differences in posterior parameterisation of the model inadequacy function

are largest when looking at the “roughness” coefficients, $\log(\omega_{2x})$ and $\log(\omega_{2y})$; these are somewhat larger for the cases with 61 data points than those with 26 data points. In addition, the model inadequacy variance is slightly larger for the cases with 61 data points. It is also noticeable that the standard deviations for the model inadequacy regression parameters are large, indicating that it may have been better to have calibrated with a zero model inadequacy mean in this case. The posterior mean observation error variance corresponds to an observation error standard deviation of 16cm in the case with 26 data points, and 20cm in the case with 61 data points.

As regards the effect of prior observation error distribution, the distribution of the posterior observation error variance follows the prior closely when the prior is specified with a small variance. However, the model inadequacy variance and “roughness” in both x and y directions increase with decreasing prior observation error variance (although the range is rather less than the difference between the cases with different data sets), indicating that the variability is interpreted as model inadequacy when the solution is constrained.

The calibrated prediction of the input data is shown, for each of the eight cases, in Figure 5.8. From the first column of the output water profiles, it can be seen that the mean profiles appear somewhat uneven, reflecting the uneven path of the shorelines. In particular, there is an apparent step in elevation at approximately 800m along the left bank. This corresponds to the end of the branching area of inundation seen in Figure 5.2, at a low-lying area on the DEM; the shoreline has been defined (Figure 5.4a) to be at each position on the x -axis, the extrema of the inundated area on the y -axis, thus giving rise to the step seen in Figure 5.4a and Figure 5.8. With hindsight, it may have been more appropriate to define the axis along the centre-line of the river, as would be natural for a one-dimensional model. This would certainly have been the case for a reach which did not lie conveniently along the x -axis, but the benefits of the additional notational complexity were less obvious in this case.

The second issue to appear from the first column of output, is that the width of the prediction interval reduces with reducing prior mean observation error; indeed, the third and fourth results (Figure 5.8c_i and d_i), appear to be plausible in predicting a profile which could have given rise to the observed data.

Looking at the second column of results, however, the picture is not so clear. These runs have been done with more data points, showing greater scatter, and in particular there are two apparently low values on the left bank at approximately 700m along the reach, and a series of three apparently high values on the right bank, at approximately 2500m. It can be seen in the first three frames in the second column (Figure 5.8a_{ii}, b_{ii} and c_{ii}) that the predicted mean water height profile follows the data increasingly closely with decreasing prior observation error variance. The fourth case (d_{ii}) looks fairly close to the second (b_{ii}), unsurprising since the posterior observation error estimates are similar.

What is not clear, is whether the fourth profile (Figure 5.8d_{ii}) follows the data implausibly closely, indicating that for this problem there may be an issue of identifiability between

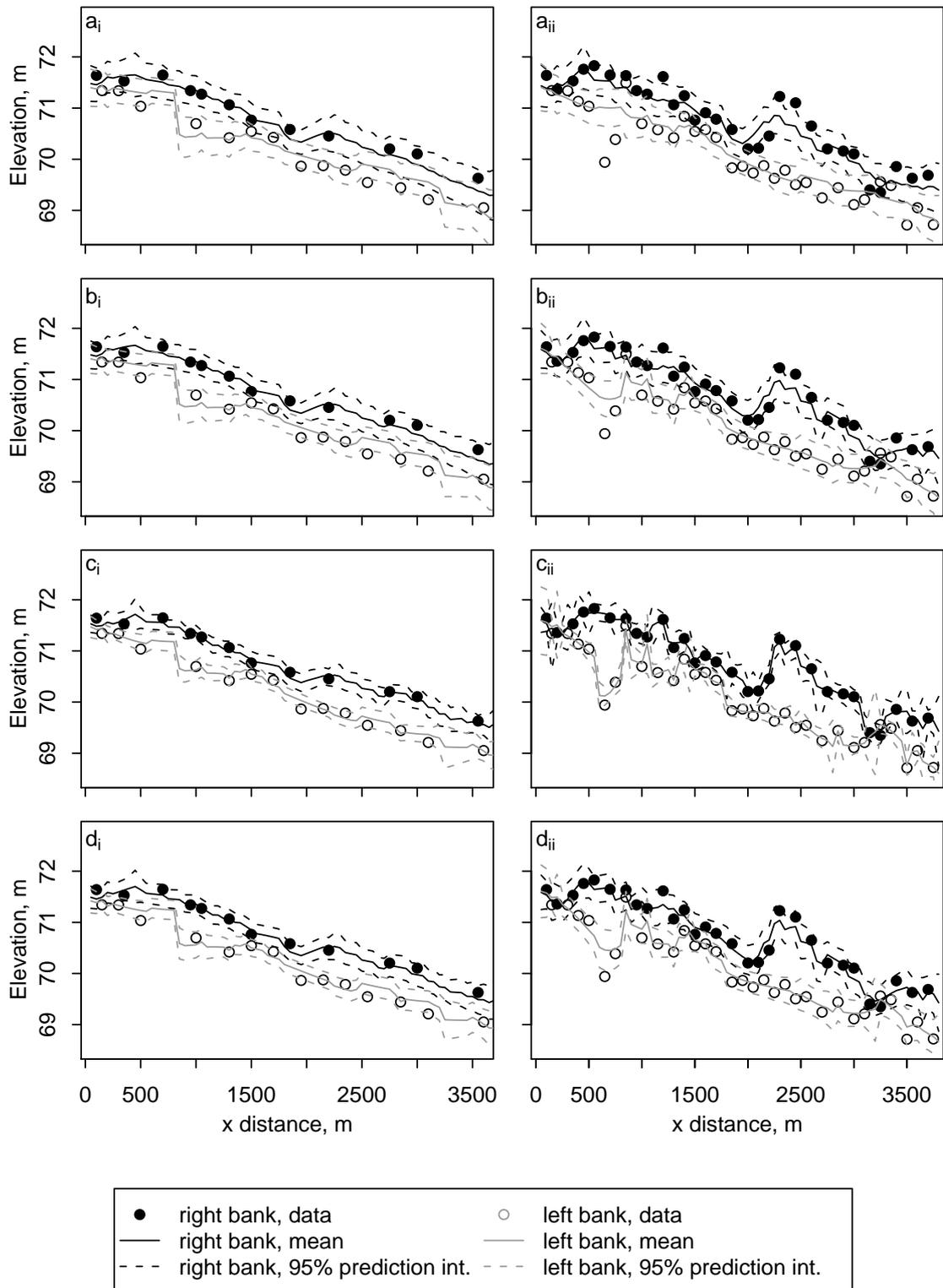


Figure 5.8: Prediction of water levels along the measured shorelines, with different prior assumptions about measurement error, and numbers of data points: a) mean error 35cm, precise distribution, b) mean error 25cm, precise distribution, c) mean error 10cm, precise distribution, d) mean error 25cm, vague distribution, i) 26 data points, ii) 61 data points

the observation error and the model inadequacy function (Wynn, 2001), which can only be resolved by a tight specification of the observation error prior mean. An alternative

interpretation would be that the excursions from a smooth profile are real; the low values on the left bank near 700m correspond with the jump in the shoreline location, while the high values on the right bank occur close to the steep bank section (Figure 5.1, where the data processing may be susceptible to greater errors in the translation from binary to height data. In any case, since this methodology distinguishes between model inadequacy and observation error on the grounds of correlation, it may be unreasonable to expect a strict separation between the two where there are sequences of points deviating from the mean line.

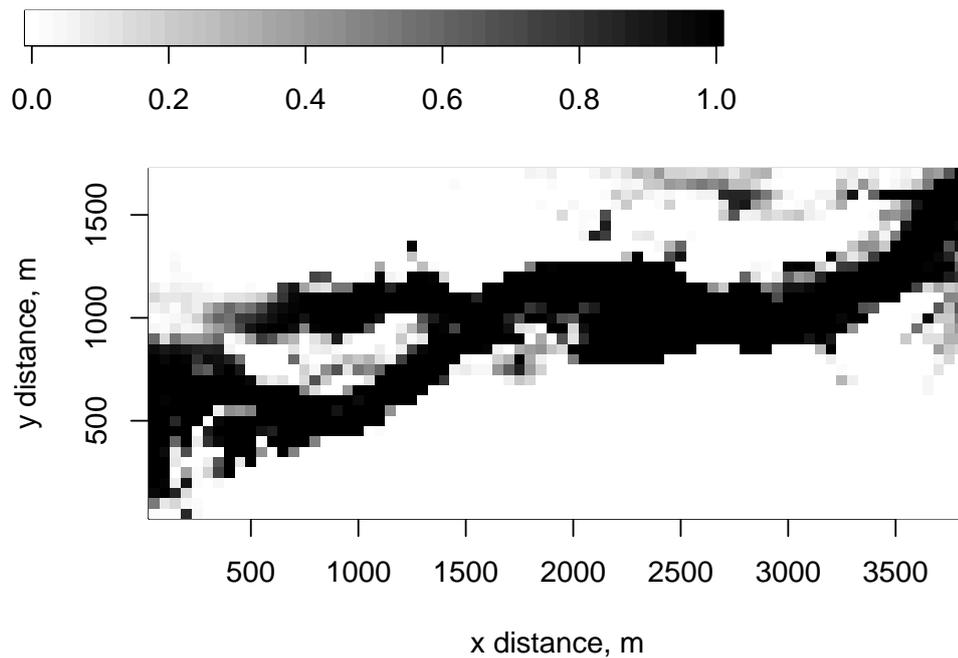


Figure 5.9: Probability of inundation

5.5 Comparison of results with previous work

By combining the predicted water profile from the case c_i with the DEM, it is possible to produce a map of probability of inundation across the entire domain (Figure 5.9). This can be compared with the flood likelihood map for the same event generated by Aronica et al. (2002) who used GLUE (Figure 5.10). The images are similar, although some difference may be expected from the difference in preparation of the data: the Aronica et al. (2002) analysis predicted binary flood inundation on a 50m grid, whereas the current analysis has been performed in terms of depths, making it more likely to show small areas of isolated inundation.

Aronica et al. (2002) noted the sensitivity of their prediction to the choice of threshold for discarding “non-behavioural” runs. This corresponds with the analysis of GLUE by Montanari (2005), who undertook a systematic examination of the dependency of GLUE model parameterisation on the assumptions commonly taken in its application. The re-

sults in the previous section illustrate that the Bayesian predictions are equally sensitive to modelling assumptions, in this case the prior distribution of observation error. However, it is more natural to express a prior belief in the distribution of observation error, which is a physical consequence of the quality of the observation process, than in the behavioural threshold used in GLUE, which does not correspond to a physical quantity. Moreover, the results generated here can strictly be referred to as predictive probabilities of flooding, whilst the results presented by Aronica et al. (2002) are relative measures of agreement which do not correspond to a carefully constructed statistical model. Clearly, the comparison between the two methods is based on a single example, moreover one where the existence of model inadequacy is not obvious from the hydraulic model output, but the current Bayesian method has the advantage of a coherent statistical analysis, and the potential of handling more substantial model bias.

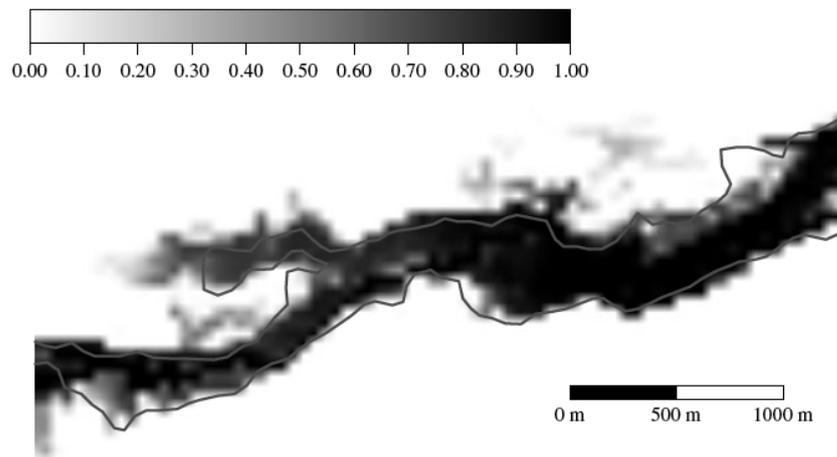


Figure 5.10: Probability of inundation obtained by Aronica et al. (2002)

5.6 Alternative emulator formulations for multivariate model output

The problem of the flood extent model could be classed as one of multivariate output. A number of methods have been explored to deal with multivariate output in the context of the formulation of Kennedy and O’Hagan; the chief issue being that the dimensionality of the calibration problem can easily become unwieldy. Rougier (2008) has classified a number of methods for dealing with emulators for multivariate output. Of these two are particularly suitable for the flood extent model. One is the approach used here. The other is the use of principal components analysis, suggested independently by Higdon et al. (2008a), and by McNeall (2008).

Principal components analysis is a technique involving linear transformation, which can be applied to a correlated set of zero mean vectors, to yield an orthogonal vector set. The transformation is unique (to scaling) and invertible. The orthogonal set, known

as eigenvectors, can be ordered by the eigenvalues, each of which contains the variance of the set in the direction of the related eigenvector. This provides the key to the use of the technique for dimensionality reduction, when the variance in some directions can be considered small enough to be ignored. The technique effectively involves projection of the data matrices onto the hyperplane which will explain the significant part of the variation.

Higdon et al. (2008a) and McNeall (2008) have used this technique to formulate empirical emulators for their high-dimensional model output. It would be possible to transform the data in a similar way, and conduct the entire analysis in the transformed space, but instead the data is introduced untransformed. This ensures straightforward treatment of uncorrelated noise, but raises the issue of the cross-covariance between transformed emulator runs and untransformed data. Higdon et al. sidestepped this issue by assuming zero cross-covariance without comment.

The use of principal components analysis has been explored for the generation of an emulator for the Buscot flood model, although not for calibration. To simplify the problem, computer model output was preprocessed to extend the specified water surface beyond the point at which it intersects the DEM to the edge of the domain. A stratified sample of runs was then drawn from the range of Manning's n , and the entire extended water surface was used for that parameter value, comprising $48 \times 76 = 3648$ data points for each parameter value. A complete transformation reduces the dimensionality of the problem from 3648 to the number of parameter values chosen. Discarding dimensions with insignificant variance will reduce the size of the problem still further. Thus, taking 6 runs with a spread of input parameter values, and applying principal components analysis, Table 5.6 shows that 98.7% of the variance is explained by the first principal component. An emulator was formulated with a single principal component, yielding results similar to those already shown.

The potential advantage of this emulator formulation is in condensing information from an extensive flood domain to a few principal components, easing the eventual calibration study. While the method is particularly useful in fields where the model output values contain underlying structure, such as the global spatial distribution of climate variables of McNeall's study (2008), it could equally be used in the case of an extensive flood domain, where it might not be appropriate to describe the model output by a few individual points, as has been done in this comparatively small calibration study.

Table 5.2: Variance explained by different numbers of principal components for a flood model emulator

Number of principal components	1	2	3	4	5	6
Percentage of total variance	98.69	0.70	0.55	0.06	0.00	0.00
Cumulative percentage	98.69	99.38	99.94	100.00	100.00	100.00

5.7 Summary

The application has been presented of the calibration methodology of Kennedy and O'Hagan to a steady state flood model, using satellite data. This has been demonstrated in the case of a model, using LISFLOOD-FP, of a reach of the river Thames, near Buscot. The application necessitated the construction of an emulator for the model, to speed up the calibration. A number of methodological choices have been discussed and demonstrated, and sensitivity of the calibration to prior distributions was examined. It was particularly noted that attention must be paid to the prior estimate of observation error variance, as this does affect the prediction interval, and hence, any inundation probability based on it.

In spite of this weakness, this calibration of a steady-state flood model on flood extent data illustrates a significant step forward, since the only other studies reported in the literature of flood model calibration based on flood extent data use a rather unsatisfactory method, leaving the uncertainties embodied in a poorly justified probability of inundation. With the increasing availability of satellite images, the calibration of flood models using such information may be expected to become more prevalent, and the introduction of an effective calibration method is thus important.

Chapter 6

Calibration of a dynamic flood model

6.1 Introduction

This chapter concerns the calibration of a dynamic flood model using historical data. Most flood models are dynamic, reflecting the evolution of the flood wave. Time series of gauged river stage or flow data are far more freely available than satellite images, such as that used for calibration in the previous chapter. Aerial images, similarly, are rarely available until well after the peak of the flood. It is thus important to be able to calibrate dynamic models.

The model calibration methodology based on Gaussian process representation of emulator and model inadequacy function was conceived for static computer models, or computer models where the output can be captured as a snapshot. A number of extensions to the Gaussian process emulator have been proposed for time-varying output, but none are so far sufficiently promising to be attractive as a starting point for calibration of output from a dynamic model, using dynamic data. The chapter begins with a brief review of these extensions in the context of time series methods used in the calibration of dynamic models, followed by a description of an emulator approach transforming the time-varying problem to one which can be treated in an analogous manner to the static model.

The approach to dynamic emulation is applied to a dynamic model of a reach of the river Severn, near Shrewsbury. Flows on this river reach have been previously modelled, with both gauged records (Romanowicz et al., 2008) and satellite images of flood extent (Bates et al., 2004) used for calibration. The work presented here is the first development of a Bayesian calibration method for dynamic flood models that explicitly incorporates model structural inadequacy. Part of this work has been reported in Manning and Hall (2010).

6.2 Calibration of dynamic models

Most calibration methods for spatially varying models can be adapted for calibration of dynamic models, by treating time as another spatial dimension. However, this leads to a high-dimensional problem. Recursive methods, where the estimates are updated one step at a time, are often more suitable for forecasting, permitting assimilation of new data as it becomes available. These methods lend themselves to the identification of models whose parameter values vary with time, although this is not a necessity. Moreover, recursive methods can be particularly efficient, using the current time series value as a starting point for estimation of the value at the next time point.

A popular and efficient method for recursive time-series estimation is the Kalman filter, which has also inspired related methods. The Kalman filter (Kalman, 1960) is a recursive algorithm designed to solve the linear time series problem:

$$\begin{aligned}x_t &= Fx_{t-1} + Du_{t-1} + w_{t-1} \\y_t &= Mx_t + \epsilon_t\end{aligned}\tag{6.1}$$

where u_t is the input, x_t the state variable vector, and y_t the observed value at time t , with system noise w_t and observation noise ϵ_t . The assumed system model is represented by matrices F and D , while M represents the relationship between the system variables x_t and the observed variables y_t . Given the covariance of the observation errors, the Kalman filter alternates between predicting the state at the next time step, and its covariance, and updating the prediction, using the deviation of the data from that forecast, under the assumption of multivariate Normal errors. The Kalman filter is most commonly thought of as a technique for data assimilation in forecasting, but can also be applied as a smoother, to provide the optimal estimate of the state x_t at all values of t in the observed series $t \in [0, T]$, under the assumption of uncorrelated Gaussian observation noise.

An alternative derivation is the Dynamic Linear Model of Harrison and Stevens (1976), who started from a Bayesian standpoint, but effectively used the Kalman filter to solve their models. The context of their work (West and Harrison, 1989), is in data models which do not necessarily have exogenous time series input, while the Kalman filter was derived expressly to interpret systems with time series input and output.

The Kalman filter represents a linear model with known parameters, but an extension, the ensemble Kalman filter (Evensen, 2003), and its associated smoother (Evensen and van Leeuwen, 2000), allow for model nonlinearity by using an ensemble of model states to estimate the state covariance. This permits simultaneous estimation of the system states and the model parameters, which need not be constant, and has been successfully applied in the assimilation of river flow data (Moradkhani et al., 2005a, Todini, 2008).

Drécourt et al. (2006) described a version of the ensemble Kalman filter, designed to allow for bias in model predictions, while Kollat et al. (2008) combined such a bias-aware ensemble Kalman filter with sequential Gaussian simulation (Deutsch and Journel,

1992), which assumes a spatial Gaussian field, forming a spatio-temporal model, and demonstrated the spatio-temporal modelling method with respect to the assimilation of monitoring data in a three-dimensional groundwater experiment.

The ensemble Kalman filter still makes an approximation to a nonlinear system model, only taking into account the first two moments of the distribution. The particle filter provides a closer approximation to such a system model, enabling greater accuracy and stability (Moradkhani et al., 2005b). The model solved by the particle filter can be generalised from Equations (6.1) to

$$\begin{aligned}x_t &= f(x_{t-1}) + d(u_{t-1}) + w_{t-1} \\y_t &= m(x_t) + \epsilon_t ,\end{aligned}$$

where the system and observation models are now allowed to be nonlinear, and it is not necessary to make the assumption that the noise series w_t and ϵ_t are Gaussian. Bayesian solution of the model equations results in an intractable posterior distribution; simulation from this distribution is achieved by tracking the behaviour of particles, or points sampled from the state space.

It could be said that the identification of model parameters which vary with time, often the outcome of recursive time series estimation, points to a structural inadequacy of the specified model to describe the system under investigation. A number of studies have similar aims in investigating the structural uncertainty in dynamic models; by identifying parallel time-dependent structures, and relating them to the original constant parameter model, it is possible to detect structural weaknesses in the original physics-based model.

The data-based mechanistic (DBM) methodology of Young (2003) has Kalman filters at its core. Many of Young's transfer function models involve the use of a nonlinear state-dependent transformation of the input data, followed by a linear transfer function model. This is a flexible and efficient model, which has been demonstrated in a number of flood catchments (e.g. Young, 2003, Romanowicz et al., 2008). It is also possible to use Young's nonlinear transfer function to replace the original, physics-based model, effectively providing an emulator for it (Beven et al., 2009).

Stigter and Beck (2004) developed a recursive prediction error algorithm, as an alternative to the extended Kalman filter, with the additional flexibility that it can be used for nonlinear as well as linear models, and used this to estimate parameters for a water-quality model for the river Calder. Lin and Beck (2007a, 2007b) synthesized the algorithm with the data-driven techniques of Young, using the parallels between the approaches to identify model structures for environmental processes.

Reichert and Mieleitner (2009) proposed a method to identify which model parameters should be considered variable, suggesting that single variables, or groups of variables be replaced by a time-dependent, stochastic parameter. The stochasticity is described by an Ornstein-Uhlenbeck process, which permits short-term variation from the parameter mean value, depending on a characteristic correlation time and an asymptotic variance. Setting

the parameters for the Ornstein-Uhlenbeck process to appropriate values, the model parameters are estimated by Bayesian regression, with a significant reduction in errors taken to imply that the data justify the replacement of the parameter under examination with a stochastic parameter in the original model. The analysis can be repeated for all variables in the model, to find out which, if any, should be considered time-dependent.

6.3 Extensions of the Kennedy and O’Hagan methodology to time-varying problems

A number of different approaches have been suggested to extend the Kennedy and O’Hagan approach to time-varying problems. Not all of these have been demonstrated in the context of calibration, as some emulation formulations are quite cumbersome.

The first approach applies to those problems where the input can be characterised by a set of scalar parameters. In this case it is possible to treat the time domain as an additional spatial dimension, and to emulate and calibrate in the same way as the foregoing examples. It can be seen that this approach may limit the length of the output time series which can be treated. An extension of this approach is thus to transform the model output into a more parsimonious domain. Higdon et al. (2008b) used Principal Components Analysis to reduce the dimensionality of a heat conduction problem in the time domain, in a similar approach to that of Paragraph 5.6 in the previous chapter. McNeall (2008) and Wilkinson (2010) both used Principal Components Analysis to reduce the dimensionality of time varying output from earth system models of intermediate complexity. Bayarri et al. (2007) used wavelet transformations to describe the complex time-series output from a model for vehicle crashworthiness. In both cases, the transformation was undertaken off-line, and the analysis proceeded on the basis of this transformation. In the case of the study of Bayarri et al., the analysis was done in the domain of the transformed output. In neither case, however, was the input also high-dimensional, such as a time series.

A second approach to emulation of time-varying models, which has not been demonstrated in the context of calibration, is to train an emulator to describe a single step of the computer model, and to emulate a time-varying output by simulation through repeated application of the single-step emulator. This approach has been used independently by Bhattacharya (2007), and by Conti et al. (2009), with slightly different implementations concerning the locations where the emulator is to be evaluated - Bhattacharya defined a grid, and Conti et al. chose points as they arose - and which simulated emulator values are used to condition subsequent evaluations. Both implementations are rather cumbersome. Conti et al. claim that their emulator is more efficient than that of Bhattacharya, but demonstrated it for only 25 time-steps of a simple rainfall-runoff model.

Little et al. (2004) used Bayes Linear modelling, while Liu and West (2009) used fully Bayes analysis, to combine the Gaussian process methodology with a time-varying autoregressive Dynamic Linear Model (West and Harrison, 1989), to form a temporo-spatial

calibration. These approaches offer greater speed than that of Bhattacharya and of Conti et al., but replaces the Gaussian process representation in the time domain with one which is expressly designed for time series analysis. However, as with the models of Higdon et al. (2008b), McNeall (2008), Wilkinson (2010) and Bayarri et al. (2007), neither formulation allowed for time series input. For hydrological applications, the autoregressive model would need to be replaced by an autoregressive model with exogenous inputs, to account for forcing. In this case, this approach would resemble that of Kollat et al. (2008), who combined a bias-aware ensemble Kalman filter with a spatial model based on a Gaussian process.

The approach taken in this study is determined by the application, involving the evaluation of Equation (1.2) in the Introduction, where the probability of downstream inundation is integrated over all possible upstream input conditions. In particular, having calibrated the hydraulic model for a given input hydrograph, the calibrated prediction is needed for other input hydrographs. The most parsimonious method of evaluation of this integral would be to couch the effect of the catchment on the upstream hydrograph as a transfer function. This is the approach taken in this study, and calibration has been performed in the transformed space, in a similar way to the study of Bayarri et al. (2007).

6.4 Flood model

The dynamic calibration methodology is developed here in the context of a dynamic flood model of the river Severn, in the region around Shrewsbury. The Severn and its chief tributary in its upper reaches, the Vyrnwy, rise in the Welsh mountains in an area of high rainfall and steep slopes. Below their confluence, the profile is much shallower, and flooding is a frequent issue in the centres of population along its course.

The area under study concerns a 22 km reach of the river Severn below the confluence with the Vyrnwy and above Shrewsbury (Figure 6.1¹). At its upstream point, Montford, long-term records exist of 15-minute measurements of both stage and discharge, although the Environment Agency (2010) note that the discharge measurements are not reliable at high flows, and recommend use of discharge values obtained from the stage using a rating curve. Similarly, records of river stage have been kept at Welsh Bridge in the centre of Shrewsbury.

The model of the river Severn used in this study was originally set up by the Environment Agency, using Hec-Ras, a commercially-used hydraulic modelling package developed by the US Army Corps of Engineers (2002), and described in Section 2.3.2.2. This was applied to a 60km reach from Montford to well below Shrewsbury. Upstream boundary conditions are provided by hourly gauged stage at Montford, while initial flow conditions along the reach were specified as the upstream gauged flow at the beginning of the modelled time frame. Downstream boundary condition was provided by the normal depth equation (Manning's

¹Data obtained from EDINA (Edinburgh University Data Library).

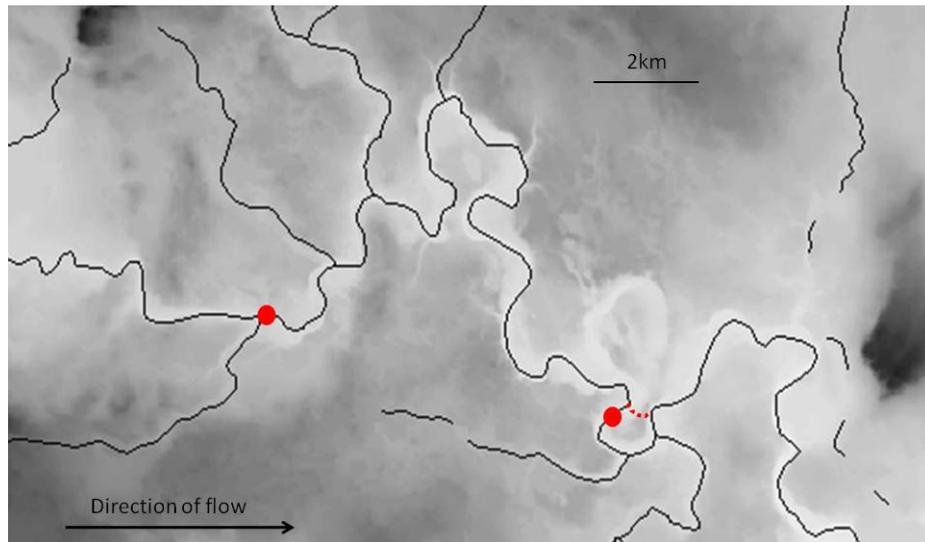


Figure 6.1: Digital elevation map of the Severn catchment in the area around Shrewsbury, with the river course superimposed, and gauging stations marked in red. Modelled flood relief channel in Shrewsbury marked with dotted line.

equation), relating flow to slope.

Taking the historical record at Montford for the period of high flow from 15th January to 7th March 2002, Hec-Ras has been run for a number of (constant) values of Manning's n for the channel. As in the previous study the value for the floodplain roughness was not varied, since the output was found to be insensitive to floodplain roughness. Figure 6.2 shows both the gauged stage at Montford, and the model output at Welsh Bridge in Shrewsbury, together with gauged stage. It can be seen that no parameter value will enable the model to reproduce the output data.

6.5 Calibration

6.5.1 Emulator construction

The emulator chosen for this river reach is due to Romanowicz et al. (2008). This is a nonlinear transfer function, consisting of an ARX model (autoregressive model with exogenous input), applied to the output of a nonlinear transformation of the upstream stage measurement. An ARX model can be described as a relationship between an input time series x_t , $t = 1, \dots, n$ and output time series y_t , $t = 1, \dots, n$ as follows:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + b_0 x_{t-l} + b_1 x_{t-l-1} + \dots + b_q x_{t-l-q} + \epsilon_t$$

where l refers to the lag between input and output, and ϵ_t is an uncorrelated noise series. The model to be used to describe the relationship between input and output is the most parsimonious model justified by the data. In this case, since a nonlinear transformation is applied to the input data before formulating the ARX model, the model order is described

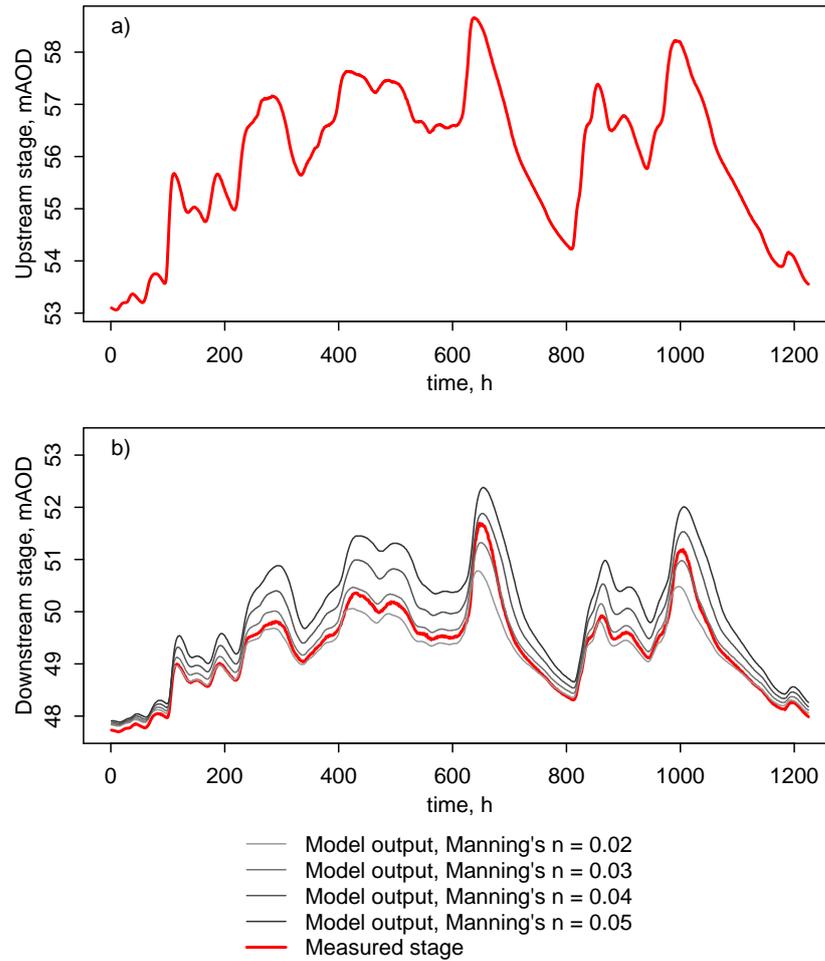


Figure 6.2: a) Gauged upstream stage at Montford, 15th Jan - 7th Mar 2002, b) Output of Hec-Ras model, with input of gauged upstream stage, for different values of Manning's n , compared with gauged downstream stage at Welsh Bridge in Shrewsbury, showing that no value of the parameter will allow the model output to correspond with the data

by $p = 1$, $q = 0$ in the above, giving

$$y_t = ay_{t-1} + bx_{t-1} + \epsilon_t \quad (6.2)$$

Heuristic justification for the choice of such a nonlinear transfer function can be demonstrated with respect to Figure 6.3. If the linear ARX model, Equation 6.2, is applied to the scaled upstream stage (Figure 6.2a), and taking the lag to be 0, the output y_t is approximately linearly related to the exogenous input x_t (Figure 6.3a). If, however, the input is, for example, squared before applying the ARX model, so that the relationship is

$$y_t = ay_{t-1} + bx_t^2 + \epsilon_t \quad (6.3)$$

then the relationship between the input and output (Figure 6.3b) is clearly nonlinear.

A plot of the downstream measured hydrograph against the upstream hydrograph (Fig-

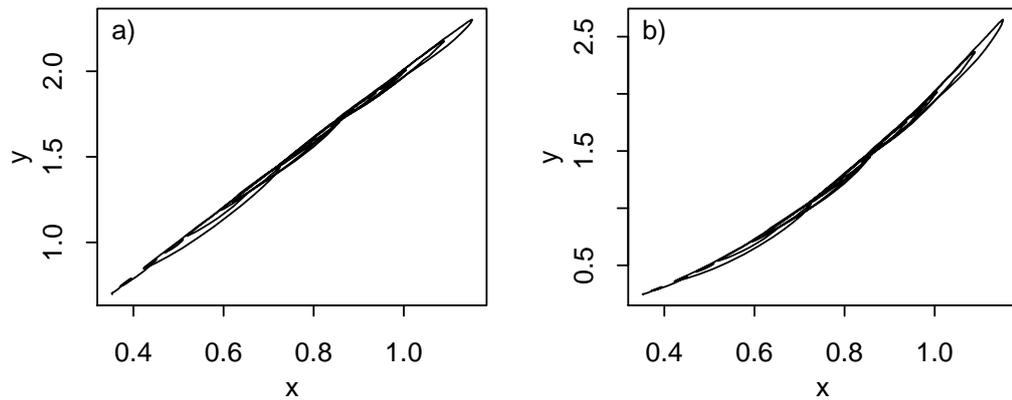


Figure 6.3: Heuristic motivation for emulator strategy: a) input-output plot, when output is determined by Equation (6.2); b) input-output plot, when output is determined by Equation (6.3)

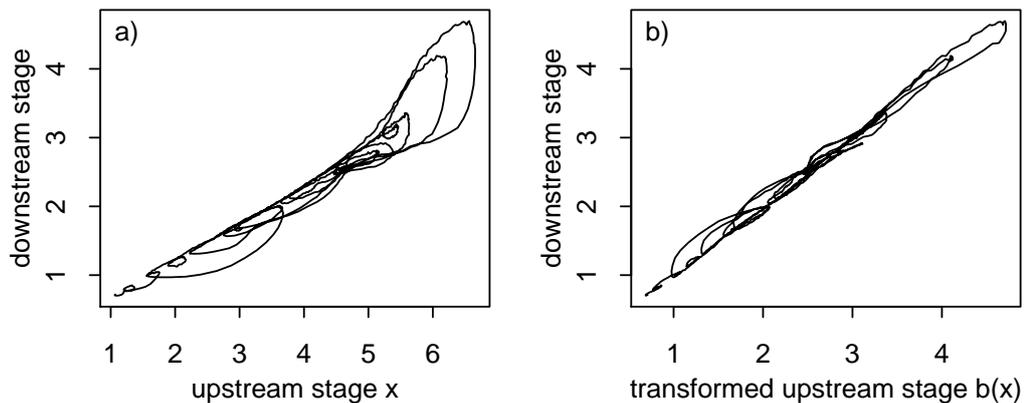


Figure 6.4: Input-output plots: a) downstream v. upstream gauged stage (relative to local data- 47mAOD and 52mAOD respectively); b) downstream stage v. transformed upstream gauged stage

ure 6.4a), demonstrates a nonlinear relationship. The modelling method suggested by Romanowicz et al. is to find a nonlinear function $b(\cdot)$ so that y_t and $b(x_{t-l})$ are related by the first order ARX model $y_t = ay_{t-1} + x_{t-l}b(x_{t-l})$. Figure 6.4b shows the effect of applying an appropriate transformation $b(x_{t-l})$ on the linearity of the relationship with y_t . It should be noted that the hysteretic effect seen in Figure 6.4a is due to the lag between the upstream and downstream signal. Choice of an appropriate lag reduces the amplitude of the “loops” in the relationship, while straightening is effected by the nonlinear transformation.

Romanowicz et al. (2008) used the function SDP of the CAPTAIN time series analysis toolbox (Taylor et al., 2007) to determine the nonlinear transformation. This method involves reordering the input and output data in terms of increasing output stage, and solving for a variable-parameter ARX model with respect to the stage rather than with respect to time. The reordering yields a smooth parameter variation with respect to the input stage. The resulting parameter variation is in effect a non-parametric function

of input data values; in order to use it further, a parametric model has to be fitted to this.

In this study, instead of using the SDP function, the nonlinearity is described as a deterministic spline function of the upstream stage, so that $b(x_t) = b(x_t|x_\tau, \tau = 1, \dots, n_\tau)$. The function values $b(x_\tau)$ at the spline knots are found using the transfer function identification routine RIV from the CAPTAIN toolbox as an inner loop of an optimisation, as illustrated in Figure 6.5. The lag used is that which minimises the fitting error for the optimal model.

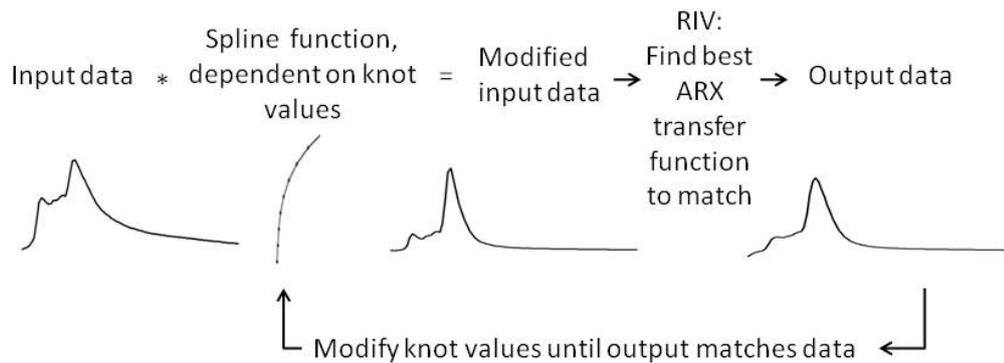


Figure 6.5: Schematic of identification procedure for nonlinear transfer function

Applying this procedure to the hydraulic model input and output for different values of the parameter Manning's n , and also for the upstream and downstream gauged stage, leads to such a representation for each case. These are shown in Figure 6.6. It should be noted that the stage is defined for this transformation relative to a local datum, and the choice of data at the two gauges is significant in this transformation. Data were chosen to ensure that the scaled nonlinear function $\frac{b(x_\tau)}{1-a}$ was nearly constant through the range of x_τ , as this reduced the numerical errors at later stages of the analysis, caused by trend in the scaled nonlinear function swamping the effects of a change in shape. Local data were chosen as large as possible, while ensuring that the relative stages were positive at all times.

Once parameters have been identified for the autoregressive function a and the spline knot values $b(x_\tau)$, time series can be recovered by application of the model $y_t = ay_{t-1} + x_{t-l}b(x_{t-l})$ to the upstream hydrograph, x_1, \dots, x_n . This has been done, both for the calibration period, and for a validation period, 25th October - 15th December, 2002 (Figure 6.7). Agreement between the recovered time series and the output of the hydraulic model is good for both time periods, and all values of Manning's n . Root-mean-squared error is 0.026m, 0.046m, 0.069m and 0.080m respectively for the four cases illustrated for the calibration period, and 0.027m, 0.043m, 0.080m, and 0.105m respectively for the cases for the validation period. This error represents a baseline error in calibration, as the calibration exercise and further analysis is based on the emulator.

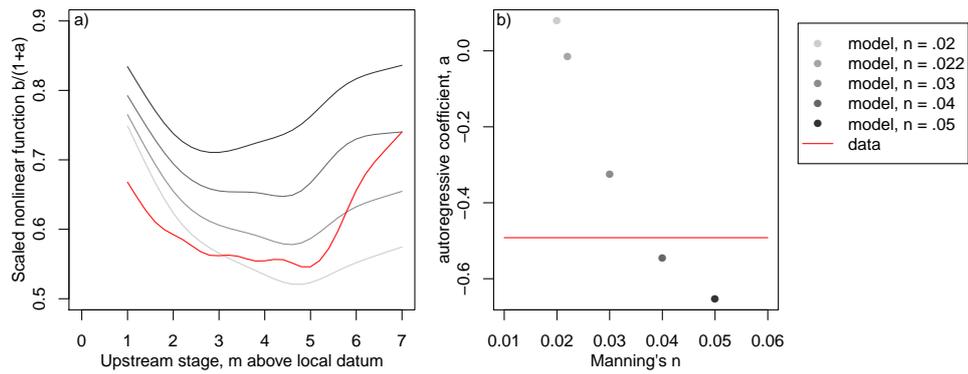


Figure 6.6: Components of nonlinear transfer function for computer output with different values of Manning's n and for data: a) nonlinear function values, $\frac{b(\cdot)}{(1-a)}$, and b) autoregressive coefficient, a .

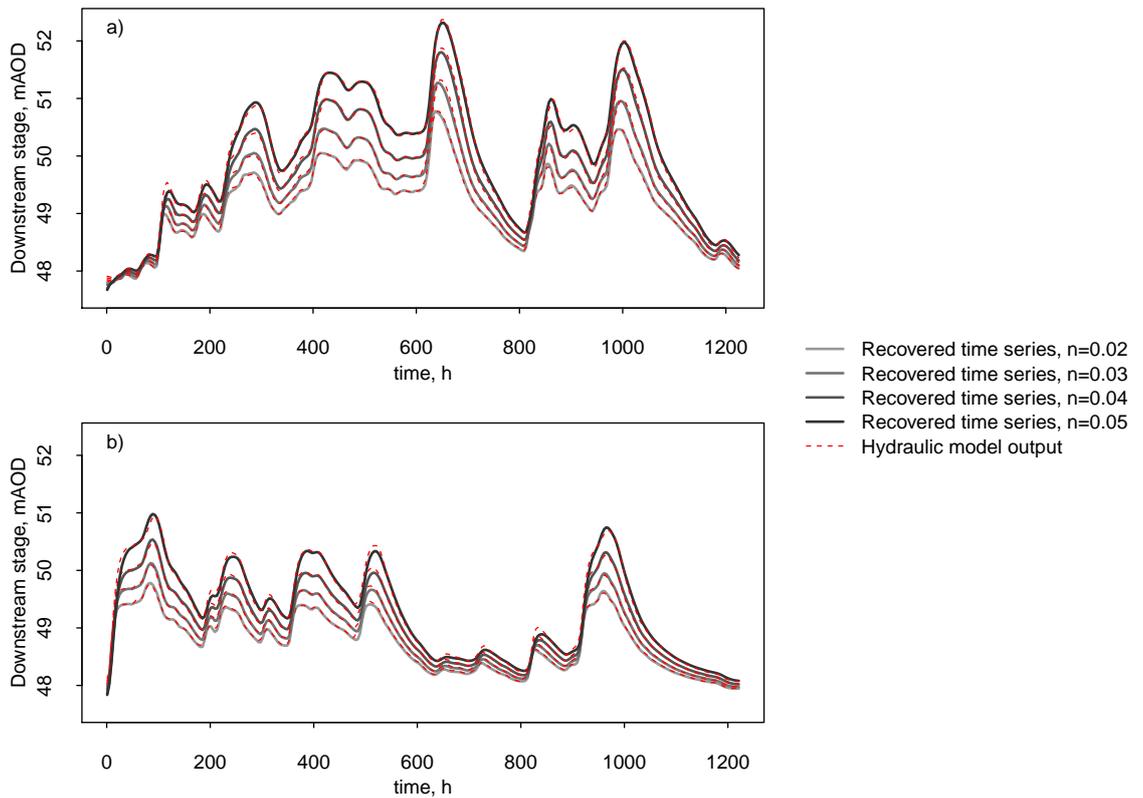


Figure 6.7: Comparison with hydraulic model output of time series recovered from nonlinear functions and autoregressive coefficients applied to the upstream hydrographs, for a) time period (15th January to 7th March 2002) used for estimation and b) validation time period (25th October - 15th December, 2002).

Three formulations have been investigated, to use the above transformation as an emulator. They are described below.

6.5.1.1 Gaussian process emulator, with calibration in the time domain

In what follows in this section, the spline described above and the CAPTAIN package are not used. Instead, a Gaussian process is invoked to serve the purpose of the spline, defining the time-domain nonlinearity in a stochastic manner. In this formulation, the conditioning points of the Gaussian process serve a similar purpose to spline knots.

If x_t and $\eta_t(\theta)$ are time series, representing measured upstream river level, and calculated downstream river level from the hydraulic model with parameter θ , assume that the model runs can be represented by an ARX model with input nonlinearity, as follows:

$$\eta_t(\theta) = a(\theta)\eta_{t-1}(\theta) + x_{t-l}b(x_{t-l}, \theta) \quad (6.4)$$

where $x_{t-l}b(x_{t-l}, \theta)$ is represented by a Gaussian process.

Simultaneous estimation of $\eta_t(\theta)$ and $b(x_{t-l}, \theta)$ for all time values requires the covariance matrix $\Sigma_{i,j} = \sum_t \eta_{t-i}\eta_{t-j}$. This covariance matrix is not easy to formulate in the context of model (6.4); for the simplified model where $\eta_t(\theta) = a(\theta)\eta_{t-1}(\theta) + \epsilon_t$, where ϵ_t is uncorrelated Gaussian noise, the covariance matrix would be $\Sigma_{i,j} = \sigma_a^2 (a(\theta))^{|i-j|}$ (cf. Equation 3.2). Equally, for the model $\eta_t(\theta) = x_{t-l}b(x_{t-l}, \theta)$, the covariance could be reasonably taken as $\Sigma_{i,j} = \sigma_b^2 \exp(-\omega(x_i - x_j)^2)$ (cf. Equation 3.3). However, the formulation of a covariance expression is more complicated for the model (6.4) where both of the terms $a(\theta)\eta_{t-1}(\theta)$ and $x_{t-l}b(x_{t-l}, \theta)$ are present.

Dropping the function arguments and invoking the Gaussian process $b'_t = x_{t-l}b(x_{t-l}, \theta)$ for clarity, the value η_t can be rewritten as follows:

$$\begin{aligned} \eta_t &= a\eta_{t-1} + b'_t \\ &= a^2\eta_{t-2} + ab'_{t-1} + b'_t \\ &= a^r\eta_{t-r} + \sum_{s=0}^{r-1} a^s b'_{t-s} \end{aligned}$$

The covariance term is then

$$\sum_t \eta_t \eta_{t-r} = a^r \sum_t \eta_{t-r}^2 + \sum_t \sum_{s=0}^{r-1} a^s b'_{t-s} \eta_{t-r}$$

It is not obvious how to evaluate the final cross term $\sum a^s b'_{t-s} \eta_{t-r}$, since b'_{t-s} and η_{t-r} are correlated, so it must be eliminated. This is only feasible if b' represents a zero mean Gaussian process; however since in reality this is not the case, the cross terms cannot be eliminated, and the covariance cannot be evaluated. This formulation is thus not practicable. It is possible that the analysis could have been reformulated to estimate the calibration recursively through time. Young (1984) pointed out that the inability to formulate the covariance relationship correctly is one reason why recursive time-series estimation is to be preferred.

The inability to formulate the emulator in the time domain has consequences for the estimation of noise error in calibration. Under these circumstances, calibration has to be undertaken in the transformed domain of the autoregressive coefficient and the spline knot values.

If the autoregressive coefficient a and the spline knot values $b(x_\tau)$ are found by optimisation as described in Section 6.5.1 with reference to Figure 6.5, then the estimation errors are minimised during the optimisation, without taking into account any prior knowledge. Since the same procedure is used for both the hydraulic model output and the measured data, the process of transforming from the time domain to the domain of the autoregressive coefficient and the spline knot values does not in itself allow for measurement error. If now the calibration is performed in the transformed domain, error estimation is undertaken in this domain. However, the structure of the estimated error, transformed back into the time domain, is no longer that of Gaussian noise.

6.5.1.2 Gaussian process emulator for the spline knot values, with calibration in the transformed domain

An alternative representation of the transformation described in Section 6.5.1.1 above, is to take two Gaussian processes in the domain of the autoregressive coefficients $a(\theta)$ and the scaled nonlinear function values at the spline knots $\frac{b(x_\tau, \theta)}{1 - a(\theta)}$ respectively, defining suitable mean functions as in previous examples. Examination of the variation represented in Figure 6.6 indicates that the scaled function values do not vary independently with parameter θ . Thus, a two-dimensional Gaussian process is invoked for the scaled nonlinear function values:

$$\eta_b(x_\tau, \theta) \sim N(H_b(x_\tau, \theta)\beta_b, V_b)$$

where

$$V_b((x_\tau, \theta), (x'_\tau, \theta')) = \sigma_b^2 \exp(-\omega_{bx}(x_\tau - x'_\tau)^2 - \omega_{b\theta}(\theta - \theta')^2)$$

The autoregressive coefficient depends on θ alone:

$$\eta_a(\theta) \sim N(H_a(\theta)\beta_a, V_a)$$

where

$$V_a(\theta, \theta') = \sigma_a^2 \exp(-\omega_a(\theta - \theta')^2)$$

Emulation is undertaken in the domain of these Gaussian processes, using for data output from an optimisation based on the CAPTAIN package. However, to avoid stability problems in subsequent recovery of the time series, η_a is defined in terms of an arctanh transformation of the data, $a' = \text{arctanh}(a)$; this transformation is also implicit in η_b , which is defined in terms of the transformed data $b'' = \frac{b}{1 - a'}$. As in previous examples, the parameters β_b , ω_{bx} , $\omega_{b\theta}$, σ_b^2 , β_a , ω_a and σ_a^2 are to be estimated by MCMC. An assump-

tion is made that the scaled nonlinear functions are uncorrelated with the autoregressive coefficient, so that there are in effect two independent emulators. This last assumption is an over-simplification of the reality; it has been seen empirically that the values at different knot points have different correlation with a .

Referring to Figure 6.6, regression bases are taken as $(1, x_\tau, x_\tau^2, \theta)$ for the nonlinear functions, and $(1, (\theta - 1.5)^{-2})$ for the autoregressive coefficients. Defining the parameter θ as $100 * \text{Manning's } n$, the emulator equations were solved for the logs of the hyperparameters. Prior and posterior distributions are given in Table 6.1; note that the distributions for the ω values are tighter than for the other variables. This reflects sensitivity of the solution to these values, in particular to the value of ω_{bx} .

Table 6.1: Prior and posterior distributions for emulation

Variable	prior distribution	posterior mean	posterior standard deviation
β_{b1}		0.70	0.01
β_{bx}		-0.13	0.00
β_{bx^2}		0.06	0.00
$\beta_{b\theta}$		0.01	0.04
β_{a1}		-0.42	0.04
$\beta_{a(\theta-1.5)^{-2}}$		0.05	0.02
$\log(\omega_{bx})$	$N(0, 1^2)$	-0.25	0.44
$\log(\omega_{b\theta})$	$N(-1, 1^2)$	-1.83	0.39
$\log(\omega_{a\theta})$	$N(0, 0.7^2)$	-0.12	0.79
$\log(\sigma_b^2)$	$N(-4.6, 2^2)$	-6.31	0.46
$\log(\sigma_a^2)$	$N(-4.6, 2^2)$	-2.76	0.87

As in previous examples, the emulator errors and variance are calculated for particular points in (x, θ) . Figure 6.8 shows the 95% prediction intervals for three values of Manning's n , midway between pairs of values where the hydraulic model has been run, compared with validation output from the hydraulic model. It can be seen, as in calibration output in previous chapters, that the uncertainty increases with distance from the spline knot values. In this case, the variation of the error variance appears more pronounced, since the emulator model does not include observation noise. Note that the widths of the prediction intervals never quite shrink to zero in Figure 6.8, as the Manning's n values are between the spline knots. The root mean variances for the nonlinear functions in Figure 6.8 are all 0.0094, with standard deviations on the autoregressive coefficients of 0.060, 0.101 and 0.109 respectively.

It should be noted that unlike the toy example described in Chapter 3, the variance of the posterior distribution depends heavily on the prior distributions for the ‘‘roughness’’ coefficients ω . These are also not easy to determine using the reference curves in Figure 3.5 (and allowing for scaling) for comparatively small values of ω . The prior distributions here were chosen, since they gave comparatively small posterior variances for the nonlinear function and autoregressive coefficients. Needless to say, a small variance in the transfer

function coefficients can be expected to translate into a small variance in the output time series.

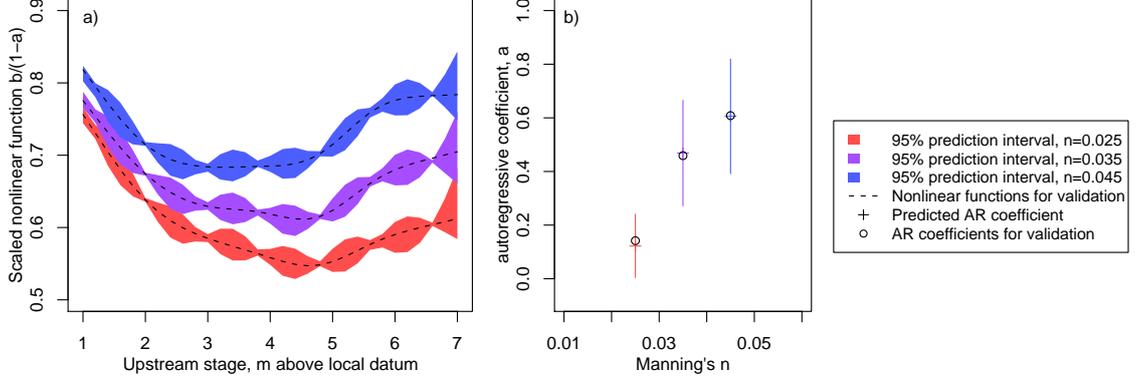


Figure 6.8: Predicted a) nonlinear functions and b) autoregressive coefficients for values of Manning's n between those used to construct the emulator, compared with values obtained directly from the hydraulic model output

Transformation to the time domain has been achieved by simulation, as follows. For each retained member of the Markov chain, estimates are calculated of the conditional means

$$E\left(\eta_b(x_\tau^\dagger, \theta^\dagger) | b''(x_\tau, \theta), \beta_b, \omega_{bx}, \omega_{b\theta}, \sigma_b^2\right) \text{ and } E\left(\eta_a(\theta^\dagger) | a'(\theta), \beta_a, \omega_a, \sigma_a^2\right)$$

and variances

$$\text{Var}\left(\eta_b(x_\tau^\dagger, \theta^\dagger) | b''(x_\tau, \theta), \beta_b, \omega_{bx}, \omega_{b\theta}, \sigma_b^2\right) \text{ and } \text{Var}\left(\eta_a(\theta^\dagger) | a'(\theta), \beta_a, \omega_a, \sigma_a^2\right)$$

of the nonlinear function and (arctanh) autoregressive coefficient values, where we recall that $a'(\theta) = \text{arctanh}(a(\theta))$ and $b''(x_\tau, \theta) = \frac{b(x_\tau, \theta)}{1 - a'(\theta)}$. Values of $\eta_b(x_\tau^\dagger, \theta^\dagger)$ and $\eta_a(\theta^\dagger)$ can be drawn from these distributions, and a number of such draws used to evaluate a conditional time series, using the time series model, $\eta_t(\theta) = \tanh(\eta_a(\theta))\eta_{t-1}(\theta) + x_{t-l}(1 - \tanh(\eta_a(\theta)))\eta_b(x_{t-l}, |x_\tau, \theta)$. A full conditional distribution is achieved by repeating the process for each member of the Markov chain. In practice, only one draw is needed from each conditional distribution, as further draws do not add to the accuracy of the unconditional distribution. Statistics of the output time series are then calculated from the ensemble of the output distributions at each time value. Use of the arctanh transformation for a ensures that the simulated conditional time series remain stable, even when the mean and variance of η_a are not small.

On transformation to the time domain, use of the scaling in the calculation of $b'' = \frac{b(\cdot)}{(1 - a')}$ ensures that the errors are comparatively small, by minimising the impact of the uncertainty in a . However, Figure 6.9 shows that the emulation errors are not negligible, in particular for the curve with the higher Manning's n value; the mean of the 95% interquartile range of the three curves are 0.107m, 0.154m and 0.267m respectively, although

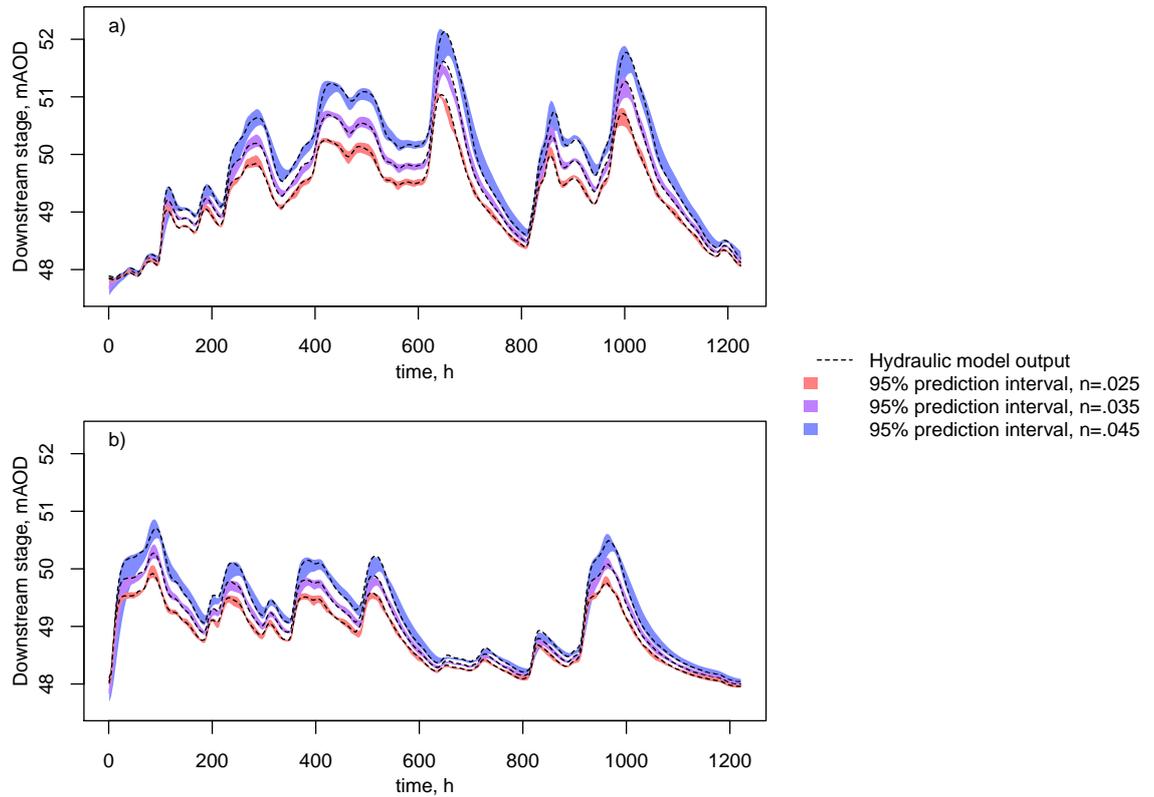


Figure 6.9: Predicted time series for values of Manning’s n away from those used to construct the emulator, compared with hydraulic model output for a) original time period (15th January to 7th March 2002) and b) validation time period (25th October - 15th December, 2002).

the root mean squared error of the median time series with regard to the output of the hydraulic model are 0.031m, 0.052m and 0.063m respectively. For the validation period, the equivalent values for the 95% inter-quantile ranges are 0.086m, 0.122m and 0.213m, while the root mean squared errors of the median time series are 0.025m, 0.052m and 0.070m. The prediction range increases, not with mean predicted stage, but with the slope of the predicted stage. Figure 6.10 shows the relationship between the prediction range and the slope of the predicted stage, with a 2-hour lag, for Manning’s $n = 0.045$, for both time periods considered, indicating a strong linear dependence. This is consistent with error caused by uncertainty in the autoregressive coefficient, which will affect the timing of the predicted peak.

The errors in extrapolation have not been estimated. This is because the hydraulic model failed to compute outside the range which had been used for emulator construction.

There is some potential confusion in this formulation between the representation of the nonlinear functions as splines $b(x_t|x_\tau, \tau = 1, \dots, n_\tau)$, and the variation of the knot values $b(x_\tau)$ as Gaussian process functions of (x, θ) . This would have been avoided if it had been possible to represent the nonlinear functions as Gaussian processes, as was proposed in Section 6.5.1.1.

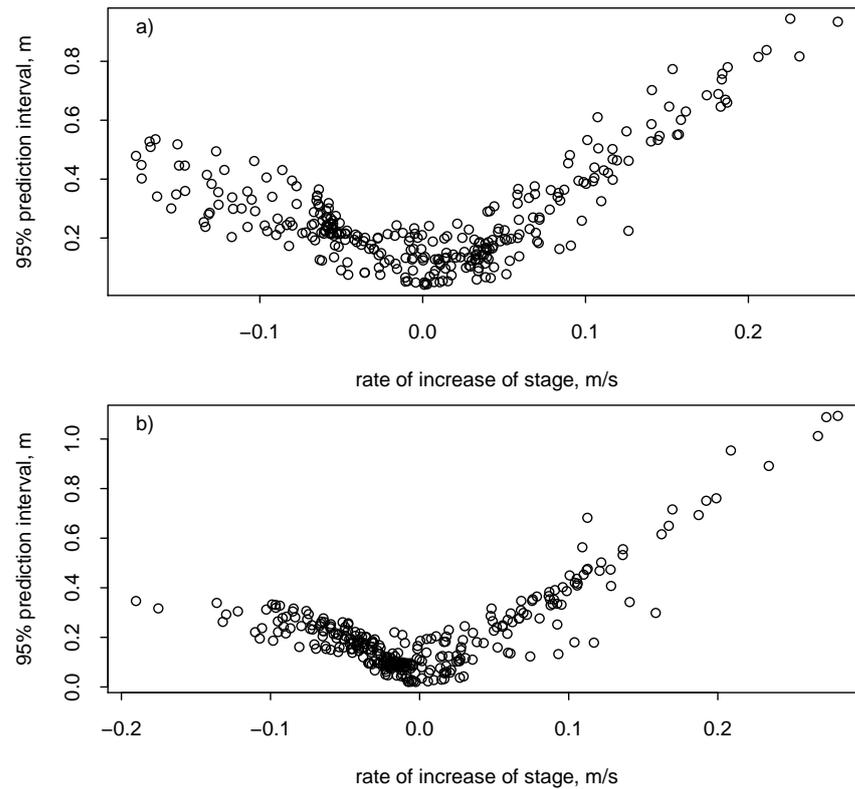


Figure 6.10: 95% prediction interval for Manning’s $n = 0.045$, compared with the slope of the predicted stage for a) original time period (15th January to 7th March 2002) and b) validation time period (25th October - 15th December, 2002).

6.5.1.3 Spline emulator, with calibration in the transformed domain

Avoiding the difficulty in specification of prior distributions for the “roughness” coefficients, and confusion in switching between spline and Gaussian process representation of the nonlinear functions, the transformation can simply be represented by a pair of spline functions, the arctanh of the autoregressive coefficient depending simply on θ and the scaled nonlinear function depending on both x and θ , again using for data output from optimisations based on the CAPTAIN package. This is a deterministic representation, and makes no allowance for the errors in determination of the original values, or of the interpolation errors for different θ values. For comparison with Figures 6.8 and 6.9, Figures 6.11 and 6.12 show the equivalent projection of the spline functions for parameter values away from those used to construct the splines. While there are no estimates involved in these projections, a very slight disagreement between the spline projections for high input stage in Figure 6.11 is reflected in a slight underestimate in the peaks of the output time series in Figure 6.12a.

6.5.2 Calibration

Calibration has been undertaken separately using the last two of the emulator formulations described above. As with these emulator formulations, calibration is undertaken in the

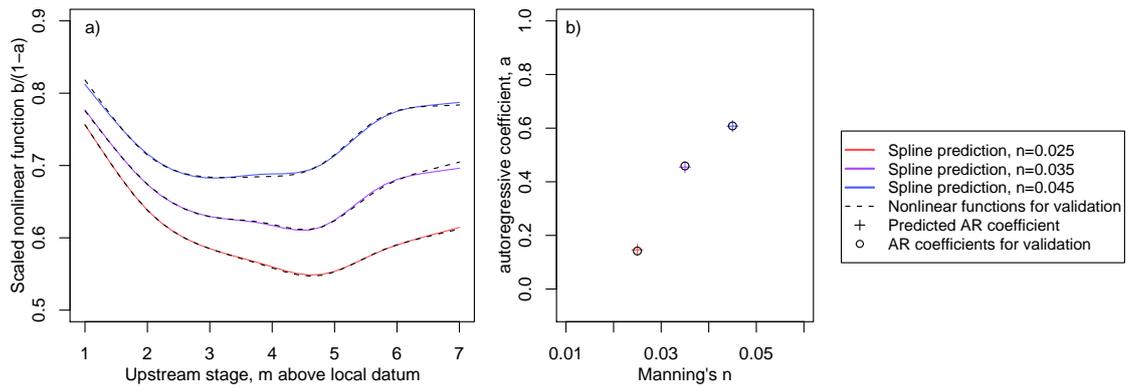


Figure 6.11: Projected a) nonlinear functions and b) autoregressive coefficients, for values of Manning's n between those used to construct the spline, compared with values obtained directly from the hydraulic model output

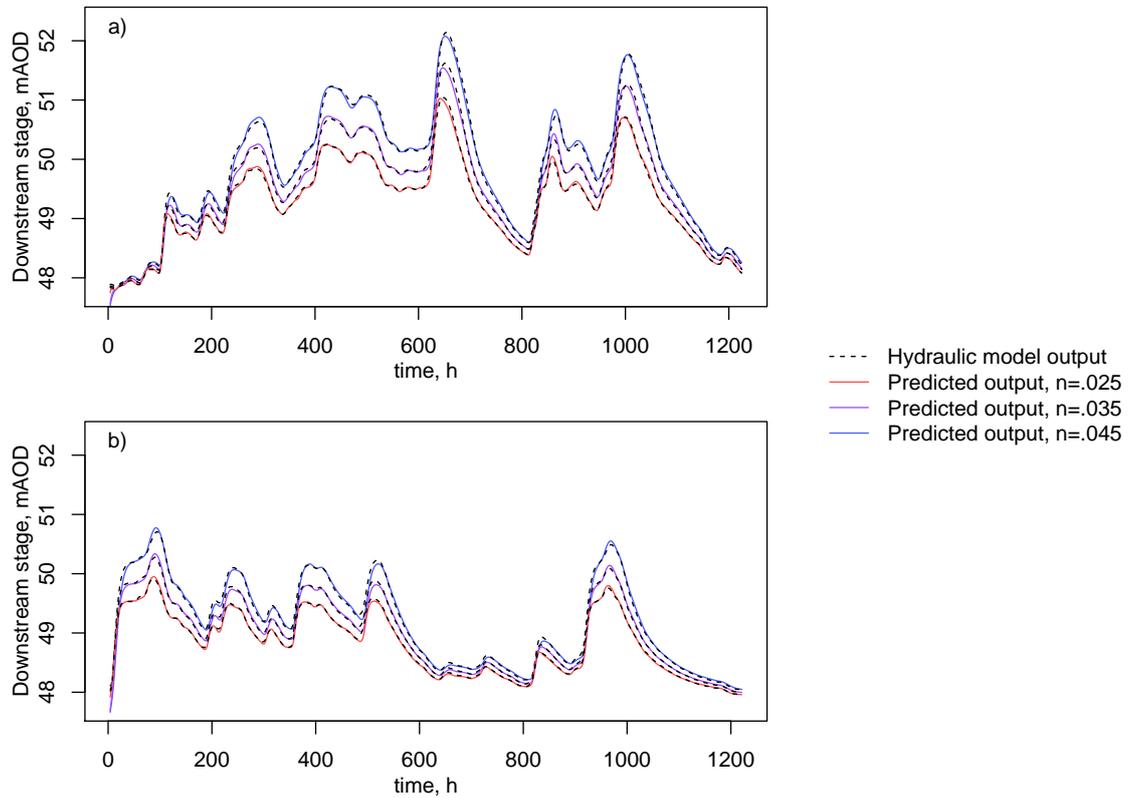


Figure 6.12: Projected time series for values of Manning's n away from those used to construct the spline, compared with hydraulic model output for a) original time period (15th January to 7th March 2002) and b) validation time period (25th October - 15th December, 2002).

transformed domain of the scaled nonlinear function and autoregressive coefficient, shown in Figure 6.6.

6.5.2.1 Calibration with Gaussian process emulator

As in the steady state flooding example, the emulator and model inadequacy are jointly estimated. However, in formulating the calibration no model inadequacy has been allowed for in the autoregressive coefficient, since a one-dimensional variable does not provide sufficient information to estimate such a quantity. Joint calibration of two quantities, one of which is estimated without model inadequacy, implies that there is a direct trade-off between the estimated parameter value and the posterior variance of that quantity. It is noted that the emulator analysis already resulted in a large uncertainty on the estimate of the autoregressive coefficient.

The model used is as follows:

$$d = \begin{pmatrix} y_b \\ y_a \\ z_b \\ z_a \end{pmatrix} \sim N(H\beta, V) \quad (6.5)$$

where

$$H = \begin{pmatrix} H_{1b} & 0 & 0 \\ 0 & H_{1a} & 0 \\ H_{1b} & 0 & H_{2b} \\ 0 & H_{1a} & 0 \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \quad \text{and}$$

$$V((x, \theta), (x', \theta')) = \begin{pmatrix} \Sigma_{1b} & 0 & \Sigma_{1b} & 0 \\ 0 & \Sigma_{1a} & 0 & \Sigma_{1a} \\ \Sigma_{1b} & 0 & \Sigma_{1b} + \Sigma_{2b} + \sigma_{eb}^2 I & 0 \\ 0 & \Sigma_{1a} & 0 & \sigma_{ea}^2 \end{pmatrix}$$

where definitions are analogous to those in Chapter 5. Note that there are three regression basis submatrices, H_{1a} , H_{1b} and H_{2b} , two for the emulators for the autoregressive coefficient and the nonlinear function, and one for the model inadequacy for the nonlinear function. Similarly, the covariance matrices Σ_{1b} and Σ_{2b} refer to the emulator and model inadequacy for the nonlinear function. Σ_{1a} refers to the variance of the autoregressive model, while σ_{eb}^2 and σ_{ea}^2 are scalar quantities arising from ARX model representation errors in the nonlinear function and autoregressive coefficient respectively. The interpretation of σ_{eb}^2 and σ_{ea}^2 as physical quantities are difficult, as they refer to a transformed variables, and are difficult to understand in the transformed space.

Once again, the input data are transformed to $a' = \text{arctanh}(a)$ and $b'' = \frac{b}{1-a'}$. The regression bases for the emulator are $h_{1b} = (1, x, x^2, \theta)$ for the nonlinear function, and

$h_{1a} = (1, (\theta - 1.5)^{-2})$ for the autoregressive coefficient, while for the model inadequacy $h_{2b} = (1, x)$ has been used for the nonlinear function. The covariance functions are taken to be smooth as in Equation (3.3), with different hyperparameters for each of the three covariance functions. These hyperparameters are to be estimated, together with the parameter θ and the variance of two sources of error, σ_{cb}^2 for inaccuracies in finding the nonlinear functions, and σ_{ca}^2 for the autoregressive coefficient. Prior and posterior distributions for the parameter θ ($=100 * \text{Manning's } n$) and the log hyperparameters are shown in Table 6.2.

Table 6.2: Prior and posterior distributions for calibration (first method)

Variable	prior distribution	posterior mean	posterior standard deviation
β_{1b1}		0.60	0.09
β_{1bx}		-0.02	0.01
β_{1bx^2}		0.02	0.05
$\beta_{1b\theta}$		0.01	0.01
β_{1a1}		-0.42	0.21
$\beta_{1a(\theta-1.5)^{-2}}$		0.05	0.04
β_{2b1}		-0.13	0.08
β_{2bx}		0.03	0.02
θ	$N(3.5, 0.5^2)$	3.23	0.39
$\log(\omega_{1bx})$	$N(0, 1^2)$	-0.32	0.42
$\log(\omega_{1b\theta})$	$N(-1, 1^2)$	-2.02	0.39
$\log(\omega_{1a\theta})$	$N(0, 0.7^2)$	-0.12	0.69
$\log(\omega_{2bx})$	$N(0, 1.4^2)$	-1.06	1.14
$\log(\sigma_{1b}^2)$	$N(-4.6, 2^2)$	-5.70	0.41
$\log(\sigma_{1a}^2)$	$N(-4.6, 2^2)$	-2.79	0.83
$\log(\sigma_{2b}^2)$	$N(-4.6, 2^2)$	-6.09	1.55
$\log(\sigma_{cb}^2)$	$N(-4.6, 2^2)$	-8.85	1.05
$\log(\sigma_{ca}^2)$	$N(-4.6, 2^2)$	-5.12	1.82

Calibrated prediction of the input data yields the posterior distributions shown in Figure 6.13 for the parameter, Manning's n , for the nonlinear functions and the autoregressive coefficient. The root mean variance of the estimate for the nonlinear function is 0.014, with a mean error with respect to the transformed output data of 0.0012, while the standard deviation of the estimate for the autoregressive function is 0.094 with a mean error with respect to the autoregressive coefficient associated with the data of 0.013. Transformation to the time domain is done in much the same way as the transformation of the emulator estimate, and has again been done for both the calibration and validation time series (Figure 6.14). For the calibration period, the mean of the 95% inter-quantile range is 0.2m, although the root mean squared error of the median time series with regard to the output of the hydraulic model is 0.004m. For the validation period, the equivalent value for the 95% inter-quantile range is 0.15m, while the root mean squared error of the median time series is 0.017m.

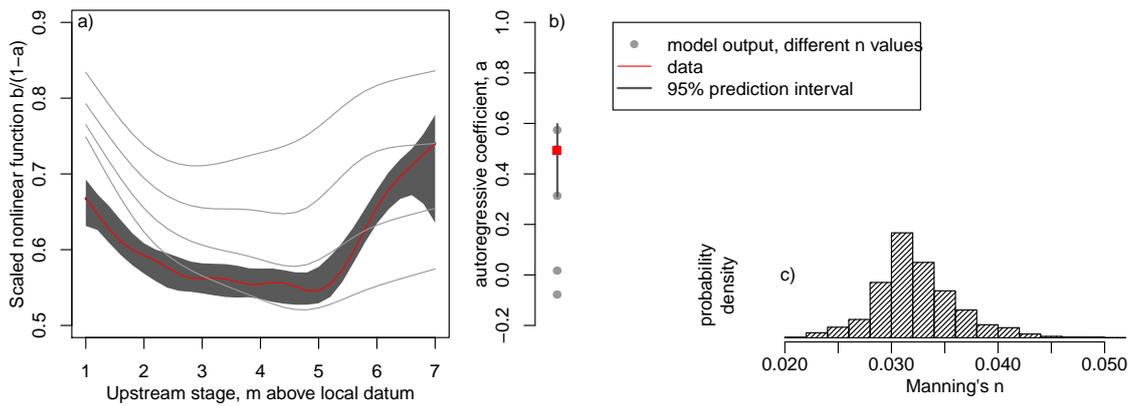


Figure 6.13: Calibrated prediction, using first calibration method, for a) nonlinear function, b) autoregressive coefficient a and c) predictive distribution for Manning's n

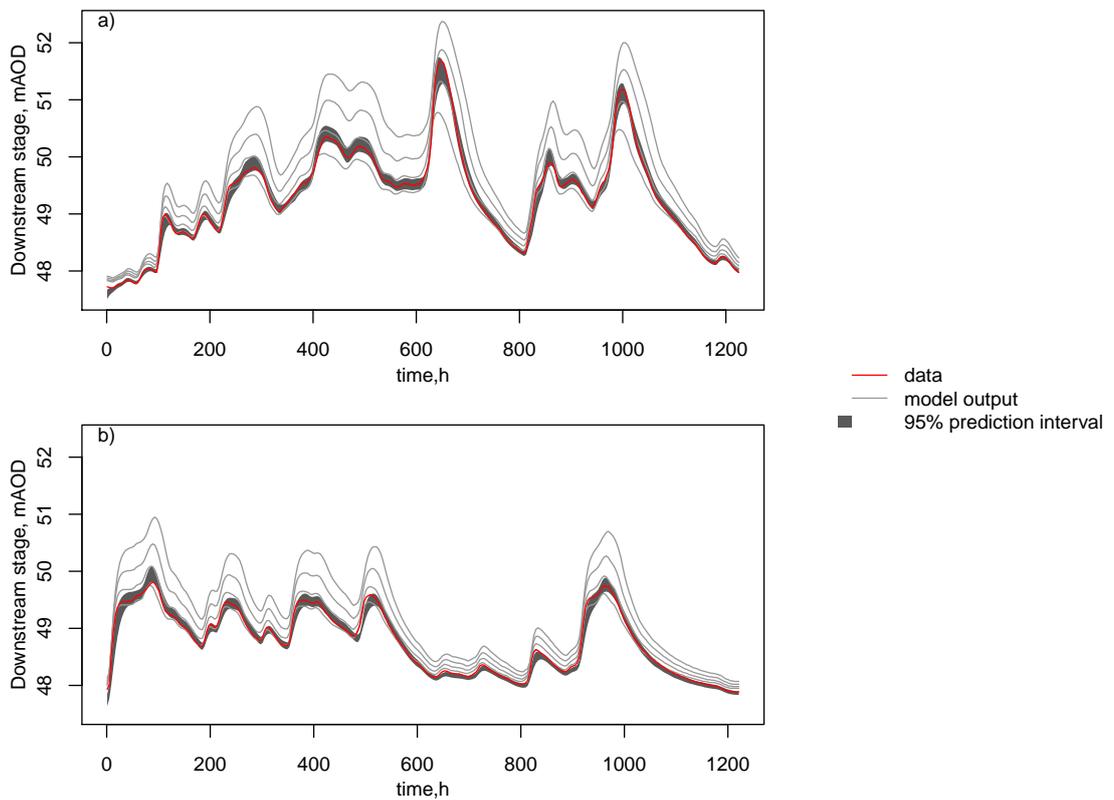


Figure 6.14: Calibrated prediction for stage at Welsh Bridge in Shrewsbury, using first calibration method, compared with model predictions and observed data: a) time period used in calibration, b) validation time period

6.5.2.2 Spline emulator

Use of a deterministic spline to encode the model output makes the problem much easier, as there is no uncertainty associated with the emulator. Once more, the calibration problem may be formulated in the domain of the autoregressive coefficients and nonlinear functions, this time without emulator uncertainty, referencing the spline functions directly. As with

the formulation in section 6.5.2.1 above, calibration is performed simultaneously for the autoregressive coefficient and nonlinear function, again, with a model inadequacy for the nonlinear function, but not for the autoregressive coefficient.

Thus, the model used is:

$$d = \begin{pmatrix} z_b \\ z_a \end{pmatrix} - \begin{pmatrix} b''(x_\tau, \theta) \\ a'(\theta) \end{pmatrix} \sim N(H\beta, V) \quad (6.6)$$

where

$$H = \begin{pmatrix} H_{\delta b} \\ 0 \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_b \\ 0 \end{pmatrix} \text{ and}$$

$$V((x, \theta), (x^*, \theta^*)) = \begin{pmatrix} \Sigma_{\delta b} + \sigma_{\epsilon b}^2 I & 0 \\ 0 & \sigma_{\epsilon a}^2 \end{pmatrix}$$

and $b''(x_\tau, \theta)$ and $a'(\theta)$ are the spline functions embodying the transfer functions, respectively the scaled nonlinear functions and the transformed autoregressive coefficients. Other definitions are analogous to those in Chapter 5.

As before, the regression basis for the model inadequacy $h_{\delta b}(x) = (1, x)$ has been used for the nonlinear function. The covariance functions are taken to be smooth, as in Equation (3.3). These hyperparameters are to be estimated, together with the parameter θ , and the variance of two sources of error, $\sigma_{\epsilon b}^2$ for inaccuracies in finding the nonlinear functions, and $\sigma_{\epsilon a}^2$ for the autoregressive coefficient. Prior and posterior distributions for the parameter θ and the log hyperparameters are given in Table 6.3.

Table 6.3: Prior and posterior distributions for calibration (second method)

Variable	prior distribution	posterior mean	posterior standard deviation
$\beta_{\delta b1}$		-0.14	0.02
$\beta_{\delta bx}$		0.09	0.02
θ	$N(3.5, 0.5^2)$	3.59	0.35
$\log(\omega_{\delta bx})$	$N(0, 1.4^2)$	-0.81	1.22
$\log(\sigma_{\delta b}^2)$	$N(-4.6, 2^2)$	-5.23	1.69
$\log(\sigma_{\epsilon b}^2)$	$N(-4.6, 2^2)$	-6.30	1.38
$\log(\sigma_{\epsilon a}^2)$	$N(-4.6, 2^2)$	-8.53	0.93

Calibrated prediction of the input data yields the posterior distributions shown in Figure 6.15 for the parameter, Manning's n , for the nonlinear functions and the autoregressive coefficient. The root mean variance of the estimate for the nonlinear function is 0.014, with a root mean squared error with respect to the transformed output data of 0.0006, while the standard deviation of the estimate for the autoregressive function is 0.109 with a mean error with respect to the autoregressive coefficient associated with the data of 0.028.

Transformation to the time domain is done in much the same way as the transformation of the emulator estimate, and has again been done for both the calibration and validation time series (Figure 6.16). For the calibration period, the mean of the 95% inter-quantile range is 0.2m, although the root mean squared error of the median time series with regard to the output of the hydraulic model is 0.006m. For the validation period, the equivalent value for the 95% inter-quantile range is 0.15m, while the root mean squared error of the median time series is 0.018m.

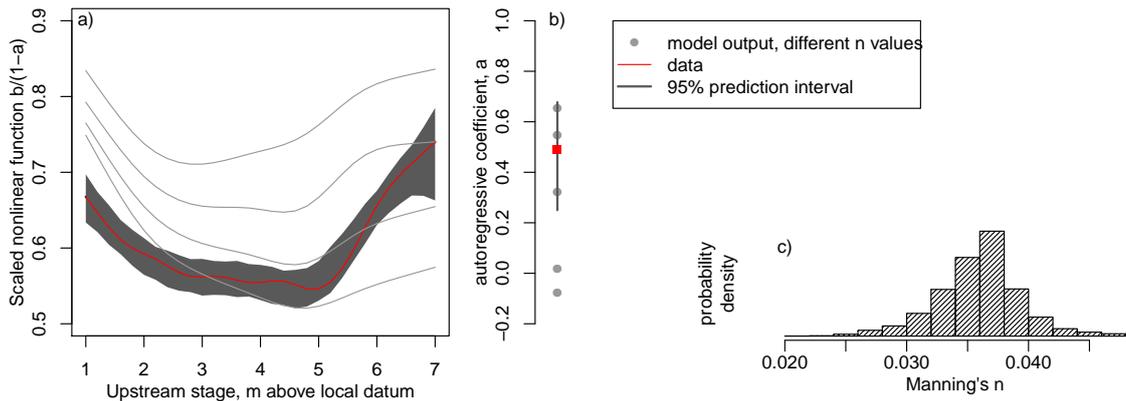


Figure 6.15: Calibrated prediction, using second calibration method, for a) nonlinear function, b) autoregressive coefficient a and c) predictive distribution for Manning's n

A comparison between the predictive distributions for Manning's n in Tables 6.3 and 6.3 indicates at first sight that the distributions are different. In addition, the distribution in Figure 6.13c is positively skewed, while that in Figure 6.15c is negatively skewed. However, further analysis shows that the two posterior distributions are not significantly different. The width of the posterior intervals is in addition influenced by the vague prior distribution on the parameter. With more information about the river reach, it would be possible to specify a tighter prior distribution. There thus appears to be little to choose between the output of the two calibration methods detailed above. However, the second calibration method has been chosen for further use, for simplicity.

6.6 Discussion

In applying the foregoing analysis to further examples, it is necessary to consider the resilience of the uncertainty estimates to a number of implementational and methodological choices. These include the choice of time interval for the calibration data and the choice of spline knot position, the stability of the calibration to input noise, and to the determination of lag in the ARX model.

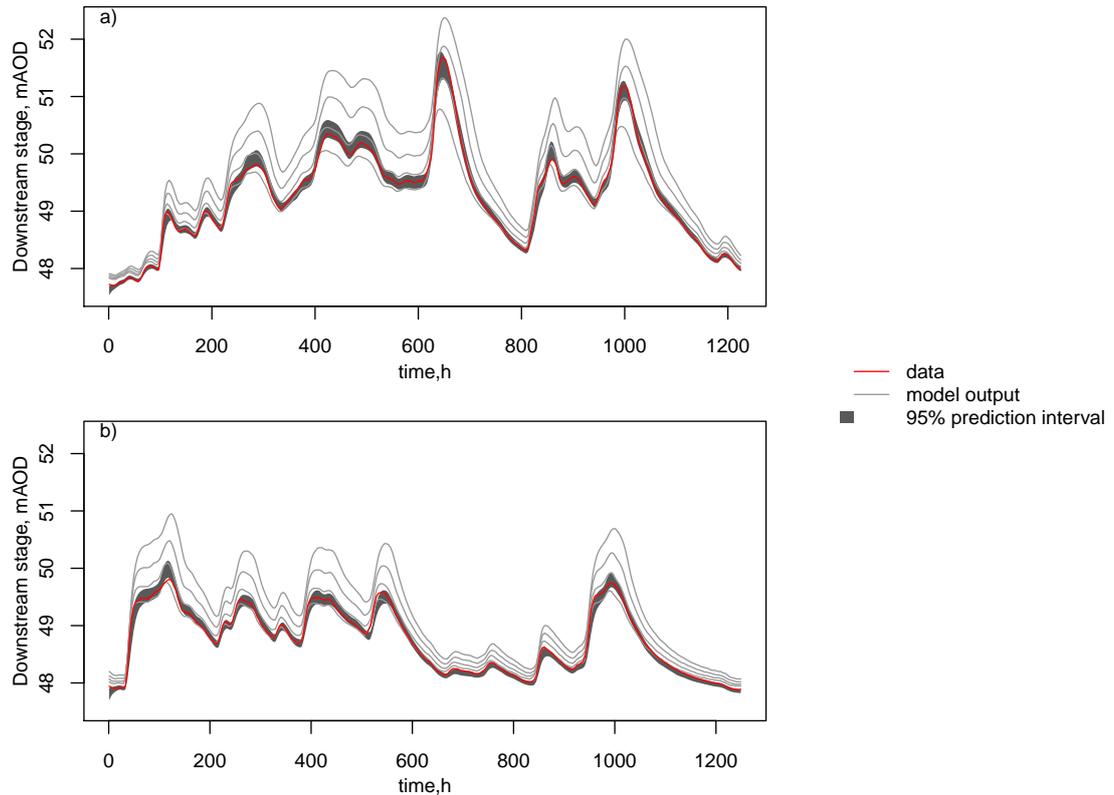


Figure 6.16: Calibrated prediction for stage at Welsh Bridge in Shrewsbury, using second calibration method, compared with model predictions and observed data: a) time period used in calibration, b) validation time period

6.6.1 Stability of the emulator formulation to spline knot position and to lag value

In order to check the stability of the optimisation calculation, it was repeated for 5, 7, 9, and 11 equally spaced spline knots, and at different lag values. It was found that the scaled spline values were stable to the number of knots, although 5 knots appeared to be somewhat coarse for the shape of the curve. The autoregressive coefficient was also stable, provided that the calculation was done with the correct lag.

However, if a lag was used other than the optimal lag identified for the transfer function, it was found that the autoregressive coefficient could change as the number of knot locations were increased. In addition, with a suboptimal lag, the autoregressive coefficient was not stable to uncorrelated noise, added to the input hydrograph, before running the hydraulic model. In this case, if the test was repeated using a number of different noise series, there appeared to be two groups of a values; however, the spline values, scaled by $(1 - a)$, were always stable.

This instability of the calculation when undertaken with the wrong lag, highlights a problem with the assumption of uniform lag. Since the emulator is constructed by running an optimisation to find the nonlinear spline function and autoregressive coefficient for each

Manning's n value, repeating the calculation at different lags to find the best one, the optimal lag is found to increase with increasing Manning's n . The emulator, however, has not been constructed to allow for different lags at different Manning's n values. In the construction of the emulator, therefore, a compromise lag is chosen, giving a suboptimal solution for some Manning's n values, but which leads to consistent autoregressive coefficients.

Application of this emulator to a longer reach may give rise to difficulties in the choice of a compromise lag. While the difference between lags for different Manning's n values is not great for the reach between Montford and Welsh Bridge, it can be expected to increase for a longer reach. This is illustrated in Table 6.4, which shows the optimal lags found from linear ARX(1,0) models applied to Hec-Ras input and output for this reach, and for the reach between Montford and Buildwas, a further 35km downstream.

Table 6.4: Lag found in best ARX(1,0) model applied to Hec-Ras output

Manning's n	lag at Welsh Bridge (hours)	lag at Buildwas (hours)
.02	0	7
.03	2	10
.04	4	12
.05	5	15

6.6.2 Comparison with Bayesian calibration, without model inadequacy

In order to understand the impact of a model inadequacy representation, a simple Bayesian calibration has been performed in the time domain, using the spline emulator instead of embedding the hydraulic model in the calibration. The data model used was:

$$\begin{aligned}
 e_t &= z_t - (a(\theta)z_{t-1} + x_{t-l}b(x_{t-l}, \theta)) \\
 &\sim N(0, \sigma^2) f(e|\theta, \sigma^2, a, b) \\
 \mathcal{L}(e_t|\theta, \sigma^2) &\propto \frac{1}{\sigma^n} \exp\left(-\frac{\sum e_t^2}{2\sigma^2}\right)
 \end{aligned}$$

Priors for θ and σ^2 were the same as those in the other calibrations in this chapter. The posterior distribution for θ is very narrow (Figure 6.17), with mean 0.0274, and standard deviation 0.0003. The comparative narrowness of the distribution reflects the lack of a model inadequacy function, since in the other formulations examined in this thesis, it could be said that the model inadequacy function has weakened

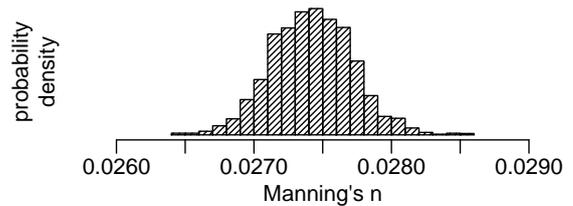


Figure 6.17: Posterior distribution of parameter Manning's n , found with simple calibration method

the model parameter identifiability. Bearing in mind this narrow distribution, it is not surprising that the calibrated prediction is sufficiently narrow, that the 95% prediction interval cannot be seen at the scale of Figure 6.18; the mean standard deviation is 0.12m for the calibration period and 0.08m for the validation period. In spite of this narrow prediction interval, it is plain that the model does not fit the data, as was already evident from Figure 6.2. This is thus a case of an over-fitted model; the effect of the calibration with model inadequacy is effectively to prevent over-fitting.

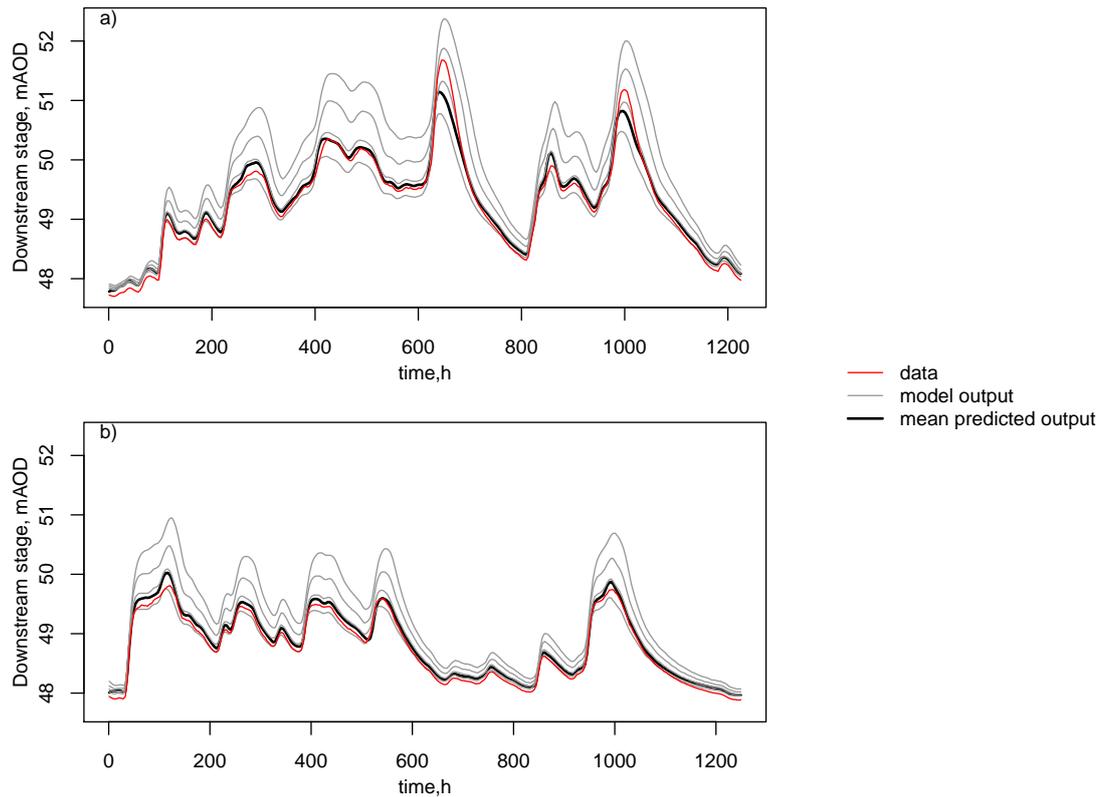


Figure 6.18: Calibrated prediction for stage at Welsh Bridge in Shrewsbury, using simple calibration method, compared with model predictions and observed data: a) time period used in calibration, b) validation time period

6.7 Channel modification

The object of using a physically based flow model is to be able to analyse the effect of future physical changes in the channel, including options for flood risk management. Thus it is necessary to consider whether the calibration method is applicable to situations where there has been a modification to the river channel. It is possible to use the hydraulic model to simulate flow in the modified channel, but the new channel model cannot be separately calibrated, because the modified channel has not been observed. Instead, it is necessary to use the existing calibration. The difference in the application comes in the calibrated prediction. In Chapter 5, the expected value for the calibrated prediction $\zeta(x^\dagger)$ at input location x^\dagger , in the case where an emulator is not used, conditional on data z , parameters

θ , model inadequacy hyperparameters ψ_δ and observation noise variance σ_ϵ^2 , is given by Equation (3.10), repeated below:

$$E\left(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2\right) = M(x^\dagger, \theta) + h_\delta(x^\dagger, \theta)^T \hat{\beta}_\delta + \tau(x^\dagger, \theta)^T \Sigma^{-1} \left(z - M(x, \theta) - H_\delta(x, \theta) \hat{\beta}_\delta \right)$$

where

$\tau(x^\dagger, \theta) = V_\delta(x^\dagger, \theta, (x, \theta))$, with (x, θ) referring to the locations where the data z have been collected,

$\Sigma = V_\delta(x, \theta, (x', \theta')) + \sigma_\epsilon^2 I$ is the covariance matrix, and

$V_\delta((x, \theta), (x', \theta'))$ is the covariance matrix representing the model inadequacy alone.

Now suppose that the modified model is denoted \mathcal{M} . Then the expected value for the calibrated prediction is given by

$$E\left(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2\right) = \mathcal{M}(x^\dagger, \theta) + h_\delta(x^\dagger, \theta)^T \hat{\beta}_\delta + \tau(x^\dagger, \theta)^T \Sigma^{-1} \left(z - M(x, \theta) - H_\delta(x, \theta) \hat{\beta}_\delta \right)$$

The variance of the conditional calibrated prediction remains unchanged, being given by

$$\text{Var}\left(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2\right) = \Sigma - \tau(x^\dagger, \theta)^T \Sigma^{-1} \tau(x^\dagger, \theta) + \Lambda^T \mathcal{W} \Lambda \quad (6.7)$$

where the elements of the second correction term are

\mathcal{W} , the covariance matrix of the regression coefficients β_δ , and

Λ , defined by $(h_\delta(x^\dagger, \theta) - \tau(x^\dagger, \theta)^T \Sigma^{-1} H_\delta(x, \theta))$.

In order to find the calibrated prediction in practice, the conditional distribution for $\zeta(x^\dagger)$ must be found for each member of the Markov chain; the mean of the unconditional calibration is given by

$$E\left(E\left(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2\right)\right)$$

while the variance is

$$\text{Var}\left(E\left(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2\right)\right) + E\left(\text{Var}\left(\zeta(x^\dagger)|z, \theta, \psi_\delta, \sigma_\epsilon^2\right)\right) \quad (6.8)$$

Predicted time series are recovered as before, by simulation of the output, drawing from the conditional posterior distributions of the autoregressive coefficient and spline knot values, and filtering the input waveform for each member of the Markov chain.

As an illustration of the application of the calibration to a modified channel, consider a relief channel in Shrewsbury itself, crossing the neck of the meander, illustrated by a dotted line in Figure 6.1. Note that this is not a realistic option for flood relief; the Environment Agency recommended increasing upstream storage in the Welsh mountains to protect the entire catchment (Environment Agency, 2009b). However, the use of this relief channel provides an illustration of the methodology. The relief channel has been modelled in Hec-Ras, using two different cross sections: a triangular cross-section, of depth 5m and ground-level width 6m, and a rectangular cross section of depth 5m and ground-level width 10m. Figure 6.19 shows the calibrated prediction for (a) the original channel, (b) the

triangular cross-section relief channel, and (c) the rectangular cross-section relief channel. The equivalent predictions for the time series are shown in Figure 6.20.

Note that although the expression (6.7) for the conditional variance remains the same, the total unconditional variance (6.8) is the sum of the mean of the (unaltered) variance and the variance of the (altered) mean; thus the variance of the unconditional estimates in Figure 6.20 are not the same.

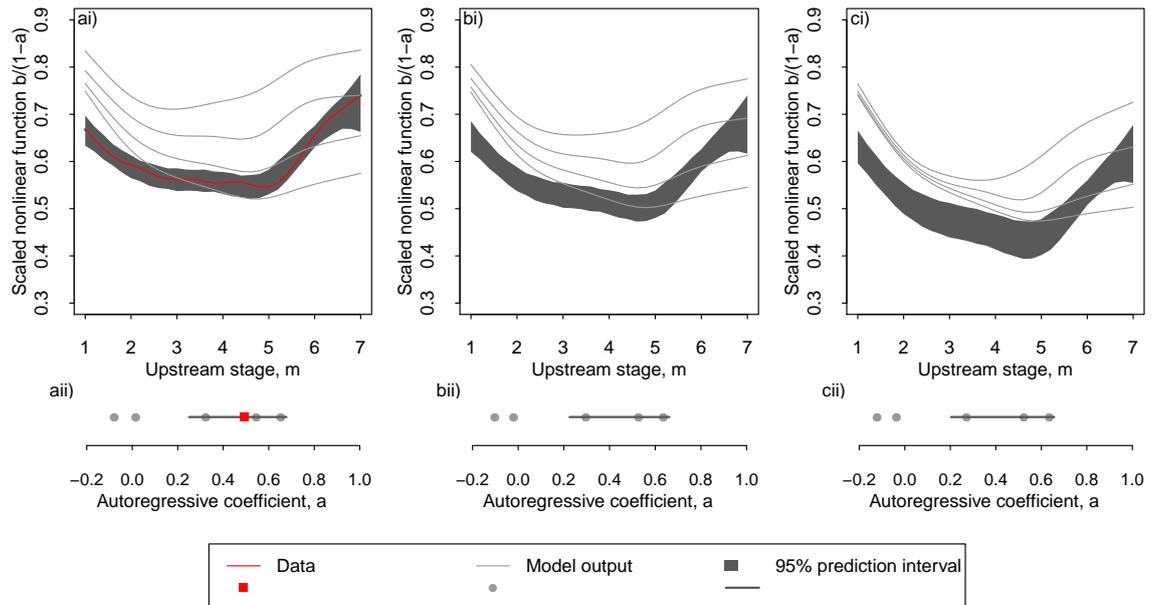


Figure 6.19: Calibrated prediction for a) original river channel, b) channel modified with triangular relief channel and c) channel modified with rectangular relief channel: i) nonlinear function and ii) autoregressive coefficient

It is questionable whether the prediction intervals for these modifications are plausible. In Figure 6.19c the transformed model output for low stage varies much less than those for the unmodified channel, calling into question the large translation at low stage for the calibrated prediction in the context of more major modification. Similarly, in the time domain (Figure 6.20), where the predicted outputs are compared with the outputs for the modified models, the predicted output in case (c) is for most of the time period below the lowest model output. However, without other information, from another model for example, since there are clearly no observations of the modified channel, it is impossible to say how large a perturbation to the original model is justifiable.

Other authors have considered the translation of the outcome of a calibration to a different model. Bayarri et al. (2007), in their calibration of vehicle crashworthiness, translated the results of the calibration from one vehicle to another, and demonstrated success in their calibration by comparison with subsequent measurements. However, it is clear that the practicality of extrapolation of a model is problem-dependent. Goldstein and Rougier (2009) suggested the principle of “reification”, where the outcome of a model could be compared in a thought experiment to a more complex model, which may or may not physically exist. The two models are compared by using an emulator for the original model,

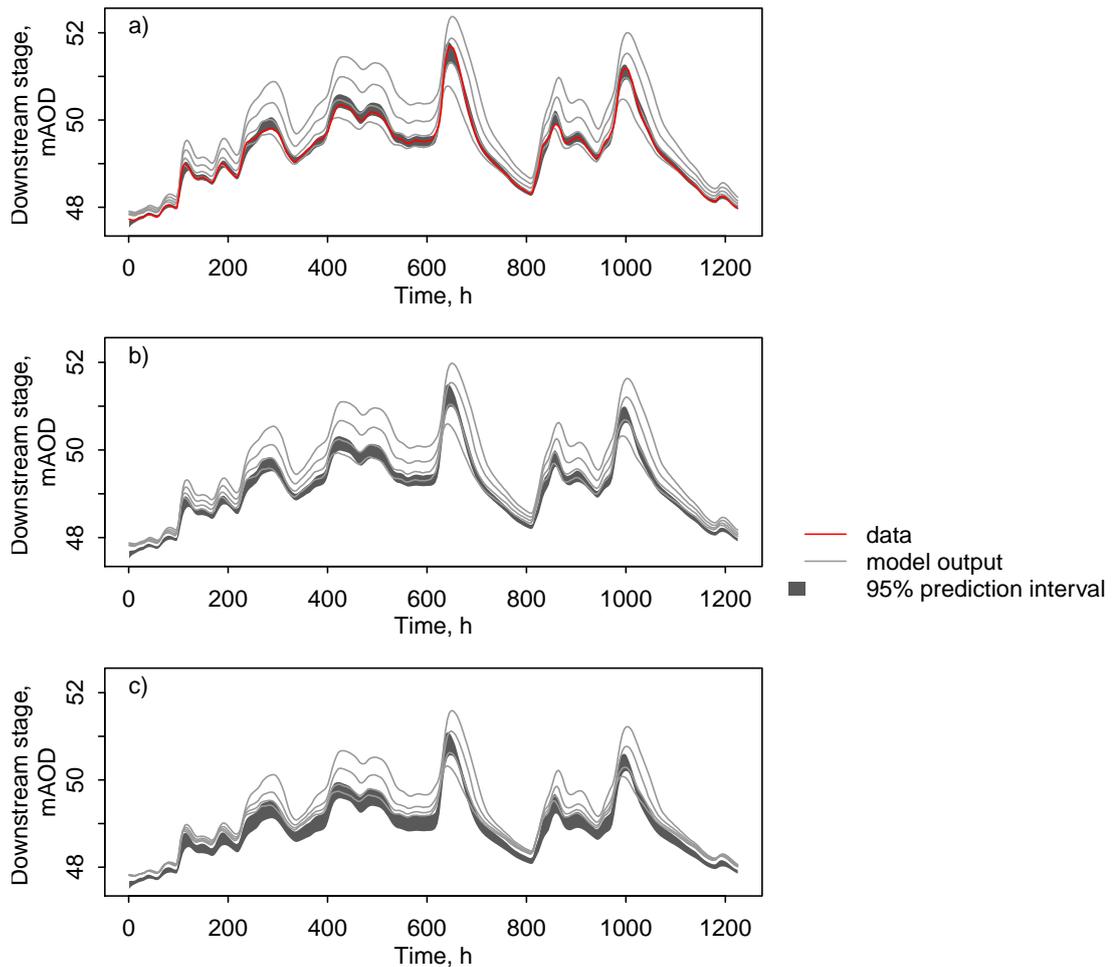


Figure 6.20: Calibrated prediction for a) original river channel, b) channel modified with triangular relief channel and c) channel modified with rectangular relief channel

which can be extended to account for the additional complexity of the second model, using Bayes Linear analysis to relate the models to the data. House (2009) applied the reification principle to a rainfall-runoff model, comparing the output of a reduced complexity rainfall-runoff model to the full model at a particular time. 67% of the outputs from complete model runs were contained within the 95% prediction interval from the reified model, indicating some success for the reification method. However, the choices made in the extension of the emulator are based on expert judgement about the expected alterations to the model output. Under some circumstances, it may be difficult to anticipate the difference in output arising from an extended model.

6.8 Summary

A methodology has been presented in this chapter for the calibration of dynamic flood models using an emulator suggested by Romanowicz et al. (2008), in the form of a transfer function. This has been demonstrated with respect to a hydraulic model for a reach of the river Severn above Shrewsbury, with gauged river stage data from a historical flood. A

number of algorithmic choices have been examined, and the robustness of the calibration methodology has been explored in terms of a number of its variables.

The methodology used involves performing the calibration in the domain of the transfer function. This choice suffers from the problem that observation noise is estimated in the domain of the transfer function parameters, and will no longer have the structure of uncorrelated Gaussian noise on retransformation to the time domain. In addition, the uncertainty accrued in translating from the time domain to the domain of the transfer function is ignored in the analysis.

Bayesian calibration of the hydraulic model, using the emulator but not the model inadequacy function, results in a calibrated prediction which underestimates the peak events, demonstrating the efficacy of the model inadequacy representation in counteracting the model bias towards lower peaks.

The emulator has been shown to form an effective transfer function for this river reach, by applying it to a different flood event. However, application of the calibration to a model representing modifications to the river reach is more problematical, as it is not clear under what circumstances the model calibration and model inadequacy can be transferred to unobserved situations.

The transfer function emulator and calibration approach have been chosen for ease of incorporation of calibrated prediction into a flood risk calculation; this will be demonstrated in the next chapter.

Chapter 7

Use of calibrated prediction in calculating probability of inundation

7.1 Introduction

The previous three chapters have demonstrated, for different types of models, how the Bayesian calibration methodology can be used to provide a calibrated prediction for flood model output, taking into account model inadequacy. However, the motivation of this study is to provide a means of incorporating the uncertainties in flood model calibration into a risk analysis study. The work in this chapter outlines, for the dynamic model of the Severn presented in the previous chapter, how the calibrated predictions from that model may be used to provide a probability of inundation.

Reformulating the statement of expected damage in expression (1.1) in the introduction to this thesis in terms of water height, the expected damage $E(c)$ is

$$E(c) = \int_h c(h)\phi(h)f(h)dh$$

where the height of water at some point in a river or floodplain is denoted h , and the damage function is denoted $c(h)$, $\phi(h)$ is the probability of a failure of the flood defence system at water height h , and $f(h)$ is the probability density of h . The distribution $f(h)$ is to be determined from available information, which includes historical field observations and computer models.

As before, taking the approach that failure of flood defence is inevitable when a threshold water height is exceeded, and that damage is a fixed cost once the flood defence has failed,

then the expected damage is proportional to the probability of inundation:

$$\begin{aligned} P(h > h_0) &= \int_{h_0}^{\infty} I(h > h_0) f(h) dh \\ &= \int_{h_0}^{\infty} f(h) dh \end{aligned}$$

In this case, where the model used to predict water height is conditional on the upstream water height the integral can be further developed:

$$P(h > h_0) = P(h > h_0 | h^\dagger) P(h^\dagger)$$

where h^\dagger is the upstream water height. Thus,

$$P(h > h_0) = \int_{h^\dagger}^{\infty} I(h > h_0) f(h | h^\dagger) f(h^\dagger) dh^\dagger \quad (7.1)$$

Evaluation of the integrand is then reduced to two problems; evaluation of the two distributions $f(h | h^\dagger)$ and $f(h^\dagger)$. The second of these two is known as the flood frequency curve, and an approach to this is described in section 7.2. The first of these distributions is to be evaluated with reference to synthetic upstream waveforms, for which downstream calibrated prediction can be evaluated. The treatment of the synthetic upstream waveforms is dealt with in section 7.3. Subsequently, the probability of inundation can be evaluated for both the existing channel, and a proposed alteration to the channel, as described in section 6.7. A sensitivity analysis is then performed, to identify which parts of the model are most influential, and a comparison is made with the simple calibration of section 6.6.2.

7.2 Flood frequency curve

The Flood Estimation Handbook (Robson and Reed, 1999) gives a set of statistical procedures for the estimation of flood frequency in a catchment. Robson and Reed recommend estimating separately the median annual maximum daily mean flow, and fitting a “growth curve” or extreme value distribution to the scaled data. Although this separation is unnecessary when dealing with data from a single catchment, the reason for it becomes more obvious when it is recognised that many catchments do not have a long record of annual maximum flows. Robson and Reed recommend that when the flood frequency of interest is smaller than the reciprocal of half of the length of reliable record, data are combined from different catchments to extend the record. The recommended threshold frequency in this procedure is a rule of thumb, resulting from the increasing prediction interval with decreasing frequency. While care must be taken to pool data from catchments with similar characteristics, errors are reduced when the data are first scaled by the median annual maximum flow.

Using the methodology set out in the Flood Estimation Handbook gives a frequency curve for upstream daily mean flow; this then has to be converted to upstream daily peak

flow using a methodology such as that of Fill and Steiner (2003), and for this study a further conversion would be required for peak height, with the aid of a rating curve. Both procedures add extra layers of uncertainty. However, since the fitting of a flood frequency curve is not the main thrust of this study, a simpler methodology is used here, fitting an extreme value distribution to the annual maxima of instantaneous upstream river peak heights.

A Generalised Extreme Value distribution (Coles, 2001, p 48) gives the probability that the level z will be exceeded in any given year, and is described by:

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (7.2)$$

where the parameters μ represents the location, σ the spread, and ξ the shape of the curve. This distribution has been fitted to the available records of instantaneous stage at Montford, which stretch from 1952 to 2008, comprising 55 complete years of data. Parameter estimates are given in Table 7.1. The annual probability of exceedance p is usually described by its reciprocal, the return period, so the flood height dependence on return period is found by substituting $G(z_p) = 1 - p$ into Equation (7.2), giving

$$z_p = \mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p)\}^{-\xi}]$$

z_p is called the return level associated with the return period $\frac{1}{p}$, and is exceeded by the annual maximum in any particular year with probability p . A return period plot is shown in Figure 7.1.

Table 7.1: Parameter estimates for Generalised Extreme Value function fitted to historical annual maxima of flood height at Montford

Variable	Mean	Standard error
μ , location	5.5332	0.08761
σ , scale	0.5741	0.06404
ξ , shape	-0.2557	0.11471

7.3 Synthetic upstream data

The integral in expression (7.1) needs to be evaluated over all possible input conditions. The section above has dealt with the distribution of peak height, but it is important to investigate to what extent the shape of the hydrograph around the peak influences the probability of inundation. If variables χ describing peak shape are influential on the probability of inundation, the integral (7.1) should be rewritten incorporating the variables

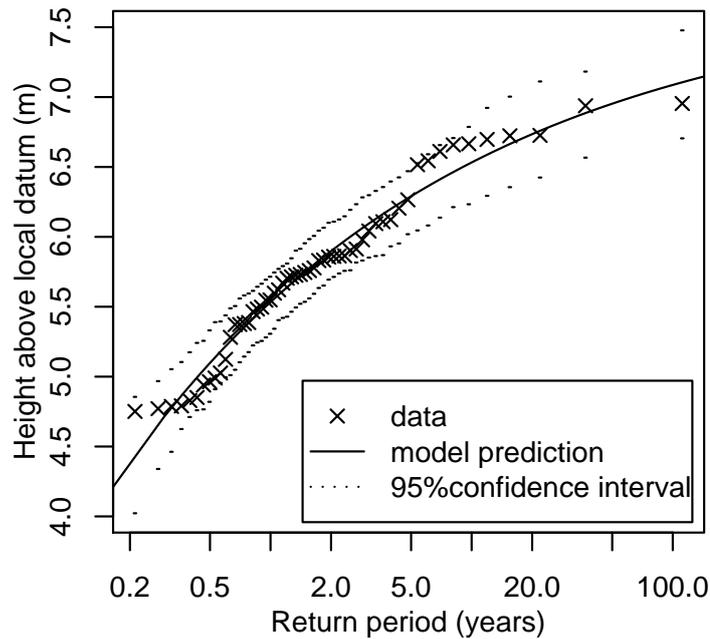


Figure 7.1: Variation of flood height with return period at Montford, as predicted by fitted Generalised Extreme Value model

χ , thus:

$$P(h > h_0) = \int_{h^\dagger} \int_{\chi} I(h > h_0) f(h|h^\dagger, \chi) f(h^\dagger, \chi) d\chi dh^\dagger \quad (7.3)$$

A sensitivity analysis has been undertaken with regard to characteristics of peak shape. This involves identifying the extent of upstream time series required before and after the peak to capture all dependencies, and a parameterisation of relevant shape indicators.

7.3.1 Selection of appropriate window length for flood peak

Figure 7.2 shows the annual peaks of river height at Montford, from 1952 to 2007, inclusive, scaled for comparison, and centred in a 20-day window. It is clear that while some of the peaks are isolated, such as 1971 and 1972, most of the others have nearby subsidiary peaks. In order to examine the proximity at which a subsidiary peak does not affect the output, synthetic curves have been constructed to represent the peak shape, and truncated at varying distances before and after the peak, substituting different constant water heights for the truncated sections. These synthetic waveforms were then introduced to the calibrated model for the transfer function representing the river section, and the sensitivity was examined of the quantiles of the predicted output peak height to the truncation window of the input peak.

The synthetic input peaks were constructed with reference to the annual maximum peaks of 1960 and 1995, both of which appear to be fairly isolated (Figure 7.3). A power curve is used for the rising limb, while the falling limb is represented by the difference between

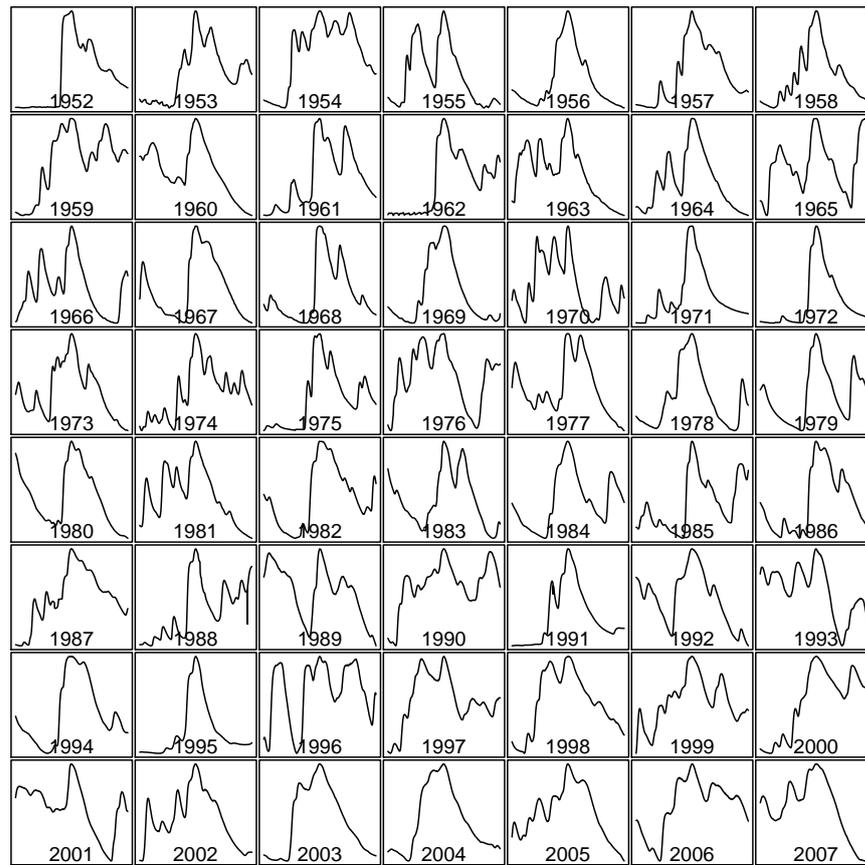


Figure 7.2: Flood height at Montford, at annual maximum peaks (scaled by maximum height), for a window of 10 days before and after the peak

two exponentials, thus:

$$x_{rising} = 1 - a(t_p - t)^b$$

$$x_{falling} = c \exp(-d(t - t_0)) - e \exp(-f(t - t_0))$$

where

t_p is the location of the peak, and

a, b, c, d, e, f and t_0 are constants to be chosen.

Synthetic curves are used to avoid the interference of irregularities found in real data. These synthetic curves are scaled to have peak height 7m above local datum (a peak height with annual probability of 0.016), have then been truncated at (72, 48, 36, 24, 12) hours before the peak, and (60, 36, 24, 12, 8, 4) hours after the peak, with the remaining parts of the curve replaced by values of 2,4, or 6. One example of this is shown in Figure 7.4.

Given a truncated input waveform, a distribution of predicted output waveforms can be found from the calibrated model as follows. For each member of the Markov chain, a draw is made from the posterior conditional distribution of the autoregressive coefficient and nonlinear function spline values. These are then used to form a transfer function which is applied to the truncated synthetic waveforms, chosen using Monte Carlo sampling. Taking

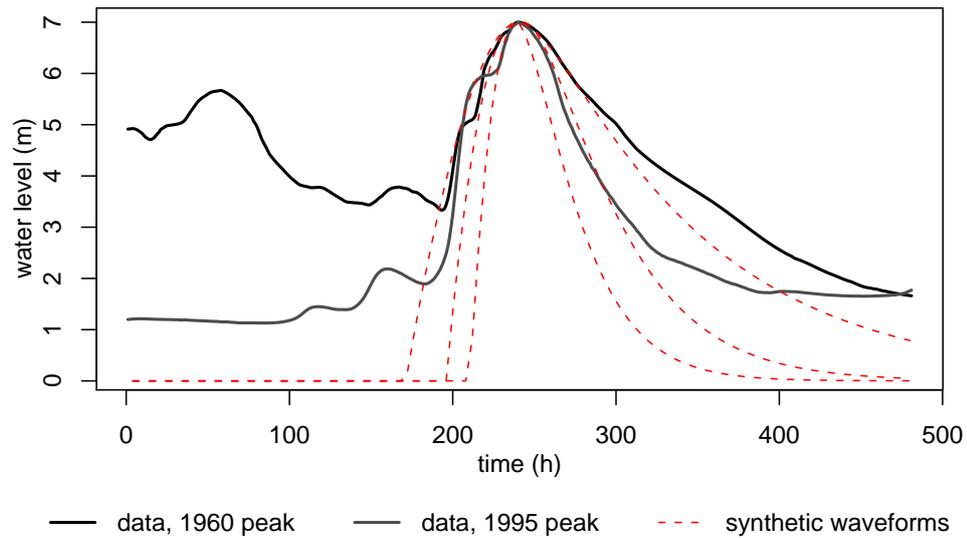


Figure 7.3: Scaled data for annual maximum peaks in 1960 and 1995, compared with synthetic waveforms used for sensitivity analysis

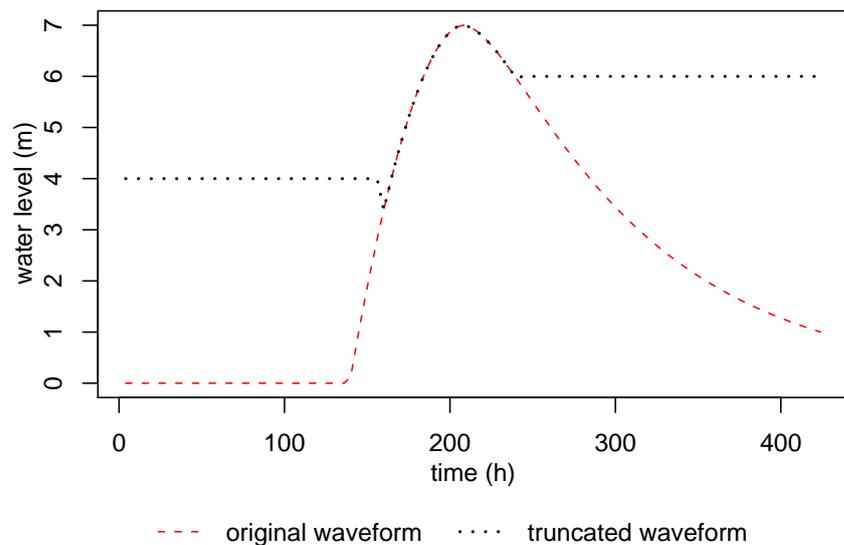


Figure 7.4: Illustration of the truncation of the flood peak waveform used in the sensitivity analysis

together the elements of the Markov chain to represent the posterior distribution, and Monte Carlo samples from the distributions of the predicted transfer function coefficients, a distribution of output peaks results for each truncated waveform. 5%, 50% and 95% quantiles of this distribution were taken, and compared with similar distributions for the untruncated input waveform. The differences at each of these quantiles are shown in Table 7.2. From this it was determined that the input waveform could be truncated 36 hours before and 24 hours after the peak.

Table 7.2: Maximum difference in quantile peak height (m), for truncated compared with untruncated upstream waveform

Quantile	Length before peak	Length after peak					
		60 hours	36 hours	24 hours	12 hours	8 hours	4 hours
0.025	72 hours	0.000	0.000	0.000	0.000	0.001	0.064
	48 hours	0.009	0.009	0.009	0.009	0.009	0.060
	36 hours	0.022	0.022	0.022	0.022	0.022	0.058
	24 hours	0.070	0.070	0.070	0.070	0.070	0.070
	12 hours	0.187	0.187	0.187	0.187	0.187	0.215
0.5	72 hours	0.000	0.000	0.000	0.000	0.000	0.013
	48 hours	0.001	0.001	0.001	0.001	0.001	0.013
	36 hours	0.004	0.004	0.004	0.004	0.004	0.012
	24 hours	0.017	0.017	0.017	0.017	0.017	0.017
	12 hours	0.166	0.166	0.166	0.166	0.166	0.166
0.975	72 hours	0.000	0.000	0.000	0.000	0.000	0.000
	48 hours	0.000	0.000	0.000	0.000	0.000	0.000
	36 hours	0.001	0.001	0.001	0.001	0.001	0.001
	24 hours	0.006	0.006	0.006	0.006	0.006	0.006
	12 hours	0.059	0.059	0.059	0.059	0.059	0.059

7.3.2 Parameterisation of peak shape

The rising and falling limbs of the annual maximum peaks, truncated to 36 hours before and 24 hours after the peak, have been modelled separately. Recognising that the rising slope can be steep, and can occur in a short space of time, at any time in the 36 hours before the peak, the rising limb has been modelled with a logistic curve (Equation 7.4), while the falling limb is modelled by a power curve (Equation 7.5). Expressions for the water heights, as a function of time are thus:

$$x = x_- + \frac{(x_p - x_-)}{1 + \exp\left(-\gamma \left(\frac{t - t_0}{t_p - t_-}\right)\right)} \quad (7.4)$$

$$x = x_p - (x_p - x_+) \left(\frac{t - t_p}{t_+ - t_p}\right)^\delta \quad (7.5)$$

where

t represents time and x represents upstream peak height, with suffices $(\cdot)_p$ referring to peak time, and $(\cdot)_-$ and $(\cdot)_+$ to 36 hours before and 24 hours after the peak respectively, and

t_0, γ and δ are parameters to be determined.

The effectiveness of these functions to represent the rising and falling limbs of the historical annual peaks is shown in Figure 7.5. While it can be seen that the falling limb is well represented, the rising limb is approximated in most, but not all cases, if preceding

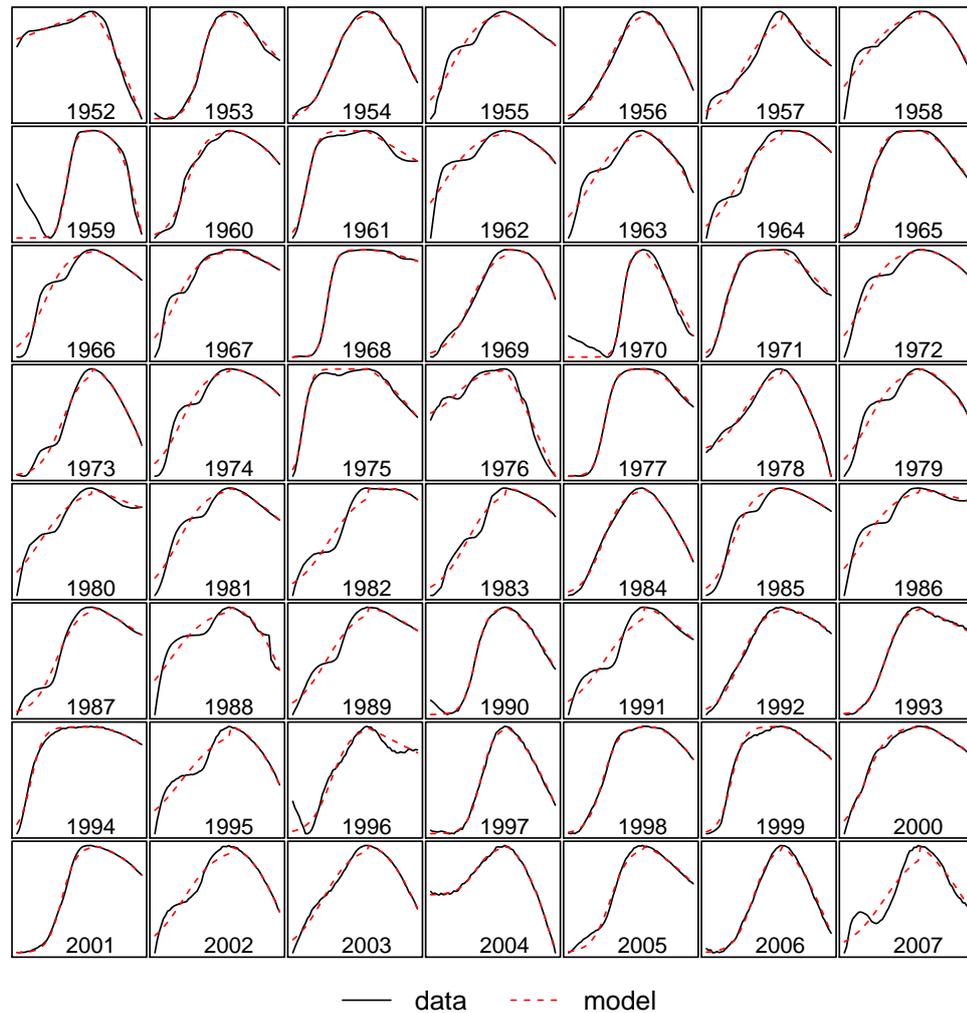


Figure 7.5: Scaled flood height at Montford, at annual maximum peaks, for a window of 36 hours before and 24 hours after the peak. Hydrograph model superimposed.

secondary peaks are ignored.

The sensitivity of the downstream peak height has been assessed to the peak shape parameters x_- and x_+ (given peak height x_p), and t_0 , γ and δ . It was found that downstream peak height was sensitive to all of these parameters; however, it was most sensitive to the parameters x_- and x_+ representing height before and after the peak.

7.3.3 Joint distributions for peak characteristic parameters

The evaluation of integral (7.1) requires a joint distribution to be found of all the variables describing distribution of peak height and upstream shape, namely, x_p , $\Delta_- = 1 - \frac{x_-}{x_p}$, $\Delta_+ = 1 - \frac{x_+}{x_p}$, t_0 , γ and δ . In order to evaluate the integral, it is necessary to characterise both the marginal distributions of the independent variates, and their correlations.

Since the number of annual maximum peaks is limited, marginal distributions and cor-

relations of the parameters describing peak shape have been estimated from the monthly maximum peaks. Discarding those months where the maximum peak is overshadowed by the adjacent month's peak, and those where the models were not able to well estimate the peak shape, there remained a sample of 540 peaks.

The marginal distributions were characterised empirically, by interpolation between the percentiles of the data. However, it was considered that while some of the variables were bounded at the upper end, others were not; the nature of the tails is listed in Table 7.3. For each of those distributions whose tails were unbounded, a Generalised Pareto distribution was fitted to the highest 5% of the data. This distribution

$$G(u) = 1 - \zeta \left\{ 1 + \frac{\xi(u - u_o)}{\sigma} \right\}^{-1/\xi}, \quad u > u_o$$

where the model is defined above threshold u_o by parameters ζ , σ and ξ . The threshold is taken here to be the 95th centile of the observed data.

Table 7.3: Nature of marginal distributions for variables describing flood peak shapes

Variable	Lower tail	Upper tail
y_p	bounded	unbounded
y_-	bounded	bounded
y_+	bounded	bounded
t_0	bounded	bounded
γ	bounded	unbounded
δ	bounded	unbounded

Simulating from the joint distribution for the flood peak heights and the other variables requires knowledge not only of their marginal distributions, but also of their correlation. Correlated covariates are to be generated by postmultiplying Normal random variates with the Choleski decomposition of the covariance matrix of appropriately transformed data. Accordingly, the parameter distributions were transformed to Normal, to investigate their correlation. Figure 7.6 is a scatterplot matrix of the transformed variables, showing the correlations between each pair. The figure has been greyscale coded for the height of the flood peak, with darker points referring to the characteristics of higher flood peaks.

It can be seen that the correlation is not constant for all peak heights; indeed it changes sign for correlation of the scaled heights before and after the peak with the peak height. There are three ways in which this could be addressed; the first would be to transform the data once more, to a set of variables with more consistent correlations, bearing in mind that the parameters are required to simulate realistic flood peaks. The second approach would be to identify copula functions to model the changing covariance directly. The third approach, which has been used here, is to partition the data into subsets with stable correlations. After investigation, correlations have been performed for peak height centile intervals of $\{(0,0.2), (0,0.3), (0.1,0.4), (0.2,0.5), (0.3,0.6), (0.4,0.7), (0.5,0.8), (0.6,0.9), (0.7,1), (0.8,1)\}$. These overlapping intervals provided as large as possible a data set from which to model the correlations, while significance tests showed that for the most

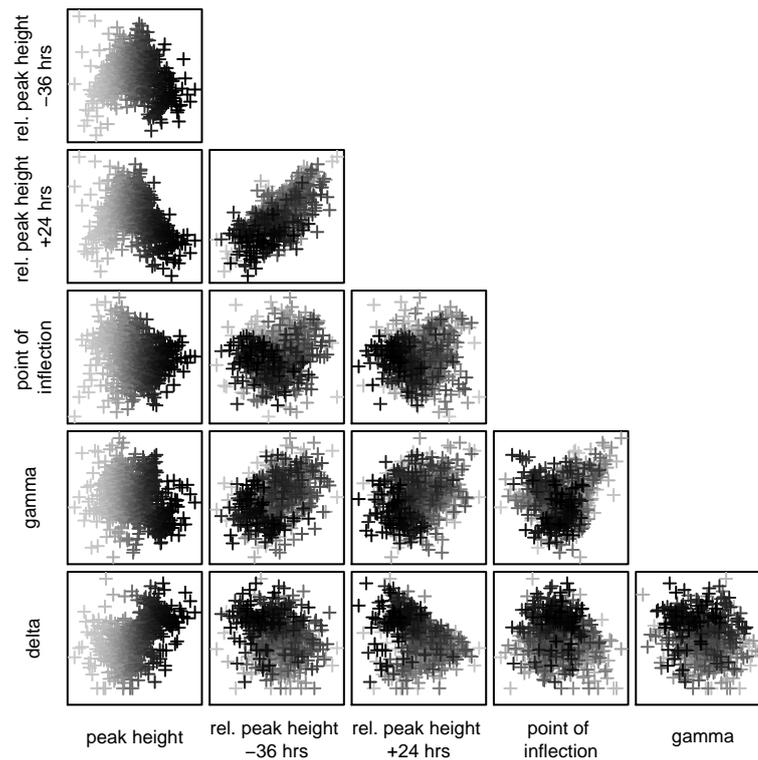


Figure 7.6: Correlation of transformed variables describing the flood peak shape at Montford. Darker shades correspond to higher peaks.

part there was not a significant difference between the correlation coefficients of adjacent overlapping intervals, ensuring that the change was gradual. The correlations were used for all variables, with the peak height in the intervals $\{(0,0.1), (0.1,0.2), (0.2,0.3), (0.3,0.4), (0.4,0.5), (0.5,0.6), (0.6,0.7), (0.7,0.98), (0.8,0.9), (0.9,1)\}$ respectively (Figure 7.7).

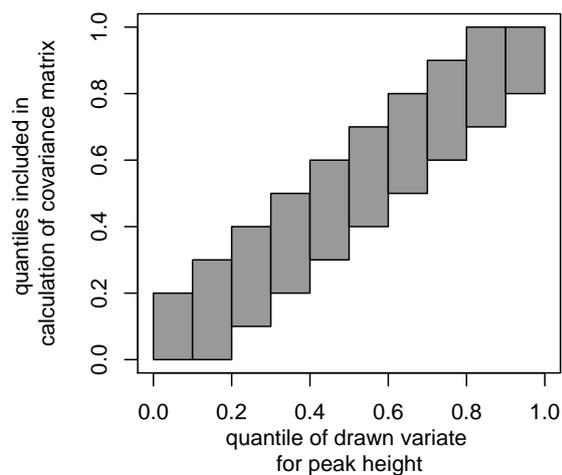


Figure 7.7: Choice of covariance matrix for correlated variables.

Use of the different correlation matrices alone, for different parts of the distribution is not adequate to ensure the correct correlation between the variates, and marginal distributions have to be used which correspond to the correlation matrices. On generating a random

sample, transforming, and transforming the correlated sample back to physical units using the appropriate marginal distribution, an appropriately correlated sample can be drawn. An example is shown in Figure 7.3.3, indicating that the method is well able to represent the covariance structure.

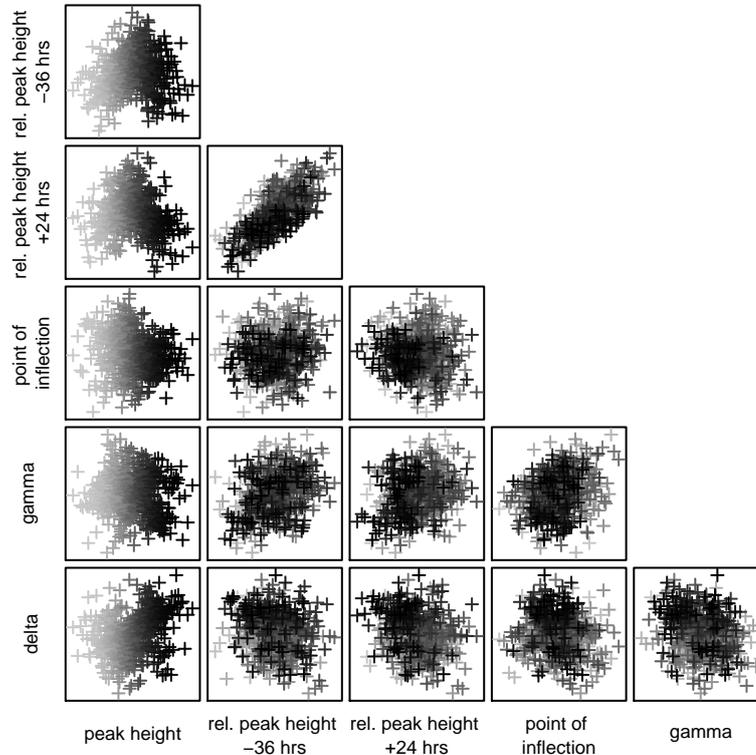


Figure 7.8: Correlation of simulated transformed variables describing the flood peak shape at Montford. Darker shades correspond to higher peaks.

7.4 Probability of inundation

Given these model parameters, the integral in Equation (7.3) is repeated here:

$$P(h > h_0) = \int_{h^\dagger} \int_{\chi} I(h > h_0) f(h|h^\dagger, \chi) f(h^\dagger, \chi) d\chi dh^\dagger$$

However, the estimate of the downstream water height, h , is dependent not only on the upstream peak height, h^\dagger and peak shape parameters χ , but also on the calibration, and in particular on the parameters and hyperparameters ψ of the statistical model used for the calibration; thus

$$P(h > h_0) = \int_{h^\dagger} \int_{\chi} \int_{\psi} I(h > h_0) f(h|h^\dagger, \chi, \psi) f(h^\dagger, \chi) f(\psi) d\psi d\chi dh^\dagger \quad (7.6)$$

This integral is evaluated as follows. For each iteration, six random numbers are drawn from the Normal distribution. The first of these represents flood peak height, and its level determines the covariance matrix for the peak shape parameters, as described in

the previous section. The vector of random variates is transformed to a correlated set by postmultiplying the vector by the Choleski decomposition of the covariance matrix. The correlated variates are then transformed from Normal to their empirical marginal distributions. The distribution for peak height is the Generalised Extreme Value distribution whose fitted parameters are listed in Table 7.1, while the distributions of the other variables are the marginal distributions found from the historical monthly flood peaks. With these variables, an input waveform is generated. A downstream waveform is the result of filtering this input waveform using a transfer function derived from the calibration process, in the same way that a calibrated prediction was found, using the procedure described in Section 6.5.1.2, and repeated below.

Recall that during the calibration process, a Markov chain is generated embodying the distributions of the parameters and hyperparameters of the statistical data model. For each member of this Markov chain, estimates are calculated of the conditional means

$$E\left(\eta_b(x_\tau^\dagger, \theta^\dagger) | b''(x_\tau, \theta), \beta_b, \omega_{bx}, \omega_{b\theta}, \sigma_b^2\right) \text{ and } E\left(\eta_a(\theta^\dagger) | a'(\theta), \beta_a, \omega_a, \sigma_a^2\right)$$

and variances

$$\text{Var}\left(\eta_b(x_\tau^\dagger, \theta^\dagger) | b''(x_\tau, \theta), \beta_b, \omega_{bx}, \omega_{b\theta}, \sigma_b^2\right) \text{ and } \text{Var}\left(\eta_a(\theta^\dagger) | a'(\theta), \beta_a, \omega_a, \sigma_a^2\right)$$

of the nonlinear function and (arctanh) autoregressive coefficient values describing the calibrated transfer function. Values of $\eta_b(x_\tau^\dagger, \theta^\dagger)$ and $\eta_a(\theta^\dagger)$ are drawn from these distributions, which are used to filter the input waveform, using the time series model, $\eta_t(\theta) = \tanh(\eta_a(\theta))\eta_{t-1}(\theta) + x_{t-l}(1 - \tanh(\eta_a(\theta)))\eta_b(x_{t-l}, |x_\tau, \theta)$. A positive contribution is made to the integral from a draw of the variates representing the shape of the input waveform and a member drawn from the Markov chain, if the output waveform, $\eta_t(\theta)$ at any point exceeds the threshold level h_0 .

The integral has been evaluated with regard to the indicator water levels set by the Environment Agency. A flood warning is issued to at risk properties when the recorded water level at Welsh Bridge is 3.15m above datum; this is upgraded to a severe flood warning when the water level reaches 4.5m above datum (Environment Agency, 2007). The integral was evaluated both for the existing channel, and using the smaller of the two modifications described in Section 6.7, for both flood warning levels.

Figure 7.9 illustrates the convergence of the probability of inundation to a final value, with an increasing number of draws from the distribution of input waveform shape parameters. Means and standard deviations are given in Table 7.4 of the annual probability of issuing the two levels of flood warning, both before and after modification. At both levels, the modification would result in a reduction of the annual probability.

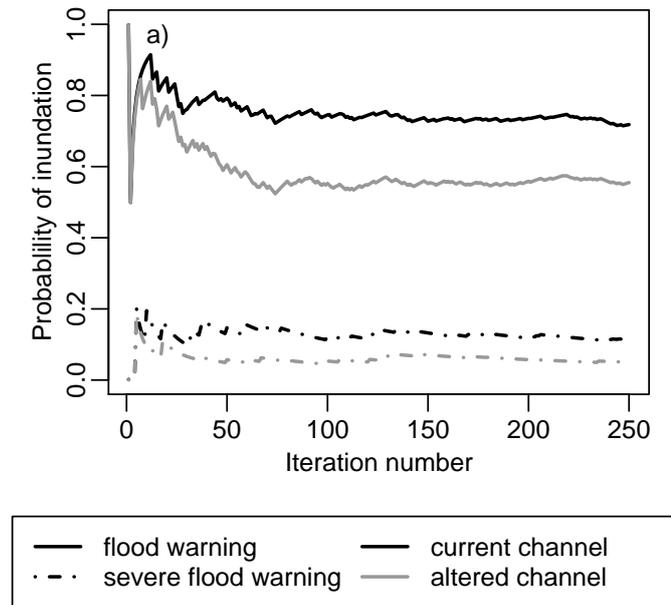


Figure 7.9: Probability of issuing a flood warning at Welsh Bridge in Shrewsbury, for unmodified and modified channels.

Table 7.4: Annual probability of issuing flood warnings, before and after channel modification

	Mean
Flood warning, unmodified channel	0.72
Flood warning, modified channel	0.56
Severe flood warning, unmodified channel	0.11
Severe flood warning, modified channel	0.05

7.5 Sensitivity analysis

In order to find which are the most important factors in the probability of inundation calculation, a sensitivity analysis was conducted. The sensitivity measure used is a variance ratio method for determining the most important factors in the evaluation of a function. If a function y is dependent on inputs $x = x_1, \dots, x_n$, then a first order measure of the sensitivity of y to input x_i is (McKay,1995)

$$S_i = \frac{V(E(y|x_i = \tilde{x}_i))}{V(y)} \quad (7.7)$$

$V(E(y|x_i = \tilde{x}_i))$ is known as the variance of conditional expectation.

Evaluation of the sensitivity of the probability of inundation integral (7.6) to the input variables is not a practicable option, as these are integrated out of the expression. The sensitivity measure is thus required of the predicted downstream peak flood height $f(h|h^\dagger, \chi, \psi)$ to upstream peak height and shape, Manning's n , the regression parameters β for the model inadequacy function, which were introduced explicitly into the calibration

for the purpose of this sensitivity analysis, variances and the other hyperparameters for the model. These are the same variates as the probability of inundation calculation, with the addition of the model inadequacy regression parameters.

It should be noted that the sensitivity analysis used here has been formulated under the implicit assumption that the function y is a deterministic one, while the conditional predicted peak downstream flood height is a stochastic function, being derived from posterior Gaussian processes. Oakley and O'Hagan (2004) used a Gaussian process emulator to increase the speed of sensitivity analysis calculations for expensive deterministic computer simulators. This analysis could be adapted to the predicted Gaussian processes arising from a calibration exercise, such as those undertaken in Chapters 4 and 5. The conditional predicted peak downstream flood height derived in Chapter 6 arises from filtering the input data with the posterior transfer function, so its distribution is a finite sum of Gaussian processes, and is thus a Gaussian process. The analysis of Oakley and O'Hagan could thus be used to examine the sensitivity of the predicted downstream peak flood height to upstream peak height and shape and Manning's n . However, in this instance a sensitivity analysis is required which includes not only the input conditions and the model parameter but also the additional parameters introduced in the calibration. The method of Oakley and O'Hagan does not allow for this. An alternative approach was taken by Degasperi and Gilmore (2008), who reported the sensitivity analysis of a stochastic biochemical model, by examining the changes in histogram distance between different realisations. Marrel et al. (2010) undertook a sensitivity analysis of stochastic models by jointly modelling the mean and variance, comparing the performance of Gaussian processes with Generalised Linear Models and Generalised Additive Models.

To evaluate the first order sensitivity measure, correlated random variates are drawn on a replicated modified Latin hypercube scheme. The modification to the usual Latin hypercube scheme is as follows. Recall that the choice of a sample of size N from a K dimensional input domain $[0, 1]^K$. For each dimension $k \in \{1, \dots, K\}$, the value of the i^{th} sample ($i \in \{1, \dots, N\}$) is taken to be at

$$\frac{\pi_k(i) - U(0,1)}{N}$$

where $U(\cdot, \cdot)$ refers to the uniform distribution, and $\pi_k(i)$ refers to the i^{th} member of a permutation of the N intervals, the suffix emphasizing that a different permutation is taken for each dimension k . In this case, the sample values are given by:

$$\frac{\pi_k(i) - 0.5}{N}$$

A replicated Latin hypercube scheme involves repeating this scheme R times, so that there are NR samples of the variates.

Iman and Conover (1982) suggested a method for generating correlated variables under a Latin hypercube scheme. The method hinges on the rank correlation, and the fact that a random draw of the variates can be transformed to have a correlation C by postmultiplying

by \mathcal{C} , the Choleski decomposition of C . Iman and Conover suggest that a Latin hypercube L is drawn for the set of variates, using the appropriate marginal distributions. For each column, the rank of each member is found. Let R be the matrix of Normal quantiles corresponding to the ranks. If R were uncorrelated, then the matrix RC would have the appropriate correlation between the variates. However, R will not be uncorrelated, so its empirical rank correlation D must be found, and \mathcal{D} , the Choleski decomposition; $\mathcal{D}^{-1}RC$ will have the required rank correlation. The final step is to reorder the individual columns of the original matrix L to have the same rank order as this new matrix. It is then a Latin hypercube with approximately the appropriate rank correlation. Note that the correlation scheme described in Section 7.3.3 was effected using variates transformed to Normal distributions. A minor modification is made here, transforming the variates to a Uniform distribution to define the correlation, thus ensuring the distribution tails are better sampled.

Evaluation of these first order sensitivity indices for a replicated correlated Latin hypercube of $m = 100$ levels and $r = 5000$ replications (Table 7.5, first column), shows that the output of the conditional posterior is indeed sensitive to the upstream peak height, and to a far lesser extent, to the peak shape parameters. However, the first order indices indicate no sensitivity, either to the model parameter, or to the parameters characterising the model inadequacy. A similar evaluation (Table 7.5, second column), using the simple calibration of Section 6.6.2 without model inadequacy, does not show sensitivity to the model parameter, either. It appears that the contribution of the upstream peak height swamps all other relationships, as one might expect.

Table 7.5: First order sensitivity analysis

Variable	Full model	Emulator only
y_p	0.96	1.01
Δ_-	0.11	0.1
Δ_+	0.15	0.12
t_0	0.09	0.1
γ	0.16	0.15
δ	0.33	0.32
θ	0	0
β_1	0	-
β_2	0	-
ω_b	0	-
σ_a^2	0	-
σ_b^2	0	-
σ_ϵ^2	0	0

In an effort to overcome the swamping effect of peak height on the sensitivity analysis, first order sensitivity was calculated for specific upstream peak height, for heights at different quantiles of the distribution. However, in this case, the sensitivity to all variables was extremely low. It should be noted that the sensitivity analysis is designed for functions with deterministic output, while the predicted downstream height in this case is a stochastic

quantity, resulting in lower sensitivities being found.

A more realistic test is to investigate the sensitivity of the downstream peak height to the upstream flood peak shape parameters and to the parameters of the nonlinear transfer function; that is the autoregressive coefficient and the spline values at different water heights, representing the calibrated model. The output of this sensitivity analysis is shown in Table 7.6, where it can be seen that the downstream peak height sensitivity changes with upstream peak height, and is greatest at that part of the spline corresponding to the upstream peak height.

7.6 Summary

This chapter has shown how the output from the calibrated prediction may be used to formulate a probability of inundation for a risk calculation. In addition to the posterior distribution of the parameters and hyperparameters from the calibration exercise, the probability of inundation depends on the distribution of upstream peak heights and peak shapes. By combining all of these uncertain factors it is possible to generate a probability of flooding (in this case the probability of exceeding a flood warning trigger level) that takes into account all of the sources of uncertainty in the flooding prediction.

The sensitivity has been explored, of the prediction of downstream peak height, conditional on the upstream peak height and shape, and on the statistical model. It has been demonstrated that the downstream peak height is sensitive to different parts of the statistical model at different upstream peak heights, a sensitivity not available for a more simple calibration.

Table 7.6: First order sensitivity analysis, conditional on upstream peak height

Peak height quantile	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
Peak height (m)	5.24	5.42	5.58	5.73	5.89	6.05	6.25	6.52	6.73
Δ_-	0	0	0	0	0	0	0	0	0
Δ_+	0	0	0	0	0	0	0	0	0
t_0	0	0	0	0	0	0	0	0	0
γ	0	0.01	0.01	0.01	0	0	0	0	0
δ	0	0	0	0	0	0	0	0	0
a	0.09	0.12	0.12	0.14	0.13	0.11	0.08	0.06	0.05
$b(1m)$	0	0	0	0	0	0	0	0	0.01
$b(1.2m)$	0.01	0.01	0.01	0.01	0	0	0	0.01	0.01
$b(1.4m)$	0.02	0.02	0.02	0.01	0	0	0.01	0.01	0.01
$b(1.6m)$	0.01	0.01	0.01	0.01	0	0	0.01	0.01	0
$b(1.8m)$	0	0	0	0	0	0	0	0	0
$b(2m)$	0.01	0.01	0.01	0.01	0	0	0	0.01	0.01
$b(2.2m)$	0.03	0.03	0.03	0.02	0.01	0	0.01	0.02	0.02
$b(2.4m)$	0.04	0.04	0.04	0.03	0.01	0	0.01	0.03	0.03
$b(2.6m)$	0.04	0.03	0.03	0.02	0.01	0	0.01	0.03	0.04
$b(2.8m)$	0.03	0.02	0.02	0.02	0.01	0	0	0.02	0.04
$b(3m)$	0.03	0.02	0.02	0.02	0.01	0	0	0.02	0.04
$b(3.2m)$	0.03	0.02	0.02	0.02	0.01	0	0	0.02	0.04
$b(3.4m)$	0.03	0.03	0.03	0.02	0.01	0	0	0.02	0.04
$b(3.6m)$	0.04	0.04	0.03	0.03	0.02	0	0.01	0.03	0.04
$b(3.8m)$	0.06	0.05	0.05	0.04	0.02	0	0.01	0.04	0.04
$b(4m)$	0.08	0.08	0.08	0.06	0.02	0	0.03	0.05	0.04
$b(4.2m)$	0.07	0.08	0.07	0.05	0.02	0	0.04	0.05	0.02
$b(4.4m)$	0.03	0.04	0.03	0.02	0	0.01	0.03	0.02	0
$b(4.6m)$	0.01	0	0	0	0.01	0.01	0.01	0	0.01
$b(4.8m)$	0.17	0.12	0.1	0.09	0.05	0.01	0.01	0.06	0.08
$b(5m)$	0.36	0.33	0.27	0.21	0.1	0.01	0.05	0.15	0.14
$b(5.2m)$	0.82	0.42	0.35	0.28	0.12	0.01	0.08	0.18	0.15
$b(5.4m)$	0.44	0.84	0.46	0.29	0.15	0.02	0.09	0.19	0.16
$b(5.6m)$	0.29	0.33	0.77	0.48	0.14	0.03	0.07	0.16	0.13
$b(5.8m)$	0.13	0.12	0.1	0.54	0.57	0.06	0.03	0.08	0.07
$b(6m)$	0	0	0	0.02	0.18	0.7	0.12	0.02	0
$b(6.2m)$	0.12	0.12	0.11	0.06	0.05	0.04	0.67	0.12	0.06
$b(6.4m)$	0.19	0.19	0.17	0.13	0.04	0	0.12	0.55	0.14
$b(6.6m)$	0.16	0.15	0.14	0.11	0.05	0	0.03	0.48	0.61
$b(6.8m)$	0.09	0.08	0.07	0.06	0.03	0	0.01	0.06	0.55
$b(7m)$	0.05	0.04	0.03	0.03	0.02	0	0	0.06	0.11

Chapter 8

Conclusions and recommendations for further work

8.1 Introduction

The study described in this thesis has demonstrated a methodology for calibration of flood models which includes known information about measurement accuracy, allows for model bias, is statistically coherent and can be used with models and data of different types. While the literature abounds with calibration methods, this one is particularly suitable for use in risk analysis, which is one of the most important purposes of flood modelling. Risk analysis has particular requirements in terms of model calibration; the calibration method must produce a probability distribution, and this must have statistical credibility. It is also important that the method should allow the input of upstream conditions other than those used for calibration.

The formulation of the current method allows not only an emulator to increase the speed of the calculation, but also a model inadequacy function to allow for bias. In the spatial domain and in the transformed time domain, both the emulator for the flood model output, and the model inadequacy are described as Gaussian processes. These are correlated Gaussian distributions, defined at any point in space, time and parameter space, but conditioned on the known model output at locations in this domain where model output has been obtained, and on measurements, where it is assumed that there is some understanding of model error. The method has been used with some success in other fields of application.

The study has been conducted through calibration of three hydraulic models of increasing complexity. Calibration of all of these involved estimation of a single parameter, the channel roughness, Manning's n .

8.1.1 Analytical model

The first application of the methodology concerned calibration of steady state laboratory experiments, using an analytical hydraulic model which was clearly inadequate to describe the physical processes. In this simple case, where there was no need for an emulator, it was possible to concentrate on the formulation of the model inadequacy representation, whose flexibility arises from the formulation as the sum of a regression relationship and a correlated Gaussian distribution. The analytical model exhibited a step behaviour which was not evident in experimental measurements, and it was shown that this behaviour needed to be represented both in the regression relationship and in the correlation structure. As more experimental series were included, a greater sophistication was required for the regression basis for the model inadequacy function to account for increasing dimensionality, while the complexity of the correlation structure increased more slowly. Eventually, however, lack of data made it impossible to specify the correlation structure completely.

The ability to describe flow as a function of stage naturally leads to the suggestion that this method could be used for the determination of rating curves, an area where there has been much activity, and where the results are of everyday use in practice.

8.1.2 Steady-state flood model

The demonstration of calibration of a two dimensional steady-state flood model using information from a satellite image represents a straightforward application of the Bayesian calibration methodology, incorporating a simple emulator for the level of the water surface. The resulting map of inundation probability is comparable with previous results in the literature. This example demonstrated the issue of identifiability of the observation error and model inadequacy, illustrating the need to have prior knowledge of the observation error variance, an issue which was noted by Wynn (2001). However, the proposed method is superior to previous work on this topic which has been based upon the GLUE methodology, and requires subjective judgment with regard to an arbitrary threshold which does not have any physical significance.

The calibration of the steady state flood model could have been made more general by including not only the roughness parameter but also the upstream flow in the calibration. As formulated, flow was taken as given, being the quantity measured at the time of the satellite overpass. However, a more thorough investigation would have couched the upstream flow as a parameter to be determined, and the analysis could thus have furnished an estimate of the accuracy of the measurement. While to do so would have exacerbated the identifiability problem, it is commonplace in flood modelling practice to update the flow estimate as part of the calibration process, so the Bayesian procedure provides a statistical framework within which to formalise this practical necessity.

One issue which was raised in the calibration of the two dimensional steady-state flood model is as follows. Computation of the Gaussian processes describing the posterior dis-

tribution of the water surface involves inversion of the covariance matrix, conditioned on the locations of the output from the flood model, and measured data. In practice, this matrix is unstable if computer output or data are used for locations which are too close together, so it may well not be possible to use all of the available data in calibration.

A second issue which appeared in this example, is that the covariance structure investigated was not able to accommodate model variability over a number of spatial scales. Kennedy and O'Hagan (2001a) did attempt to investigate other structures, such as a Matèrn correlation function, noting that the Gaussian correlation function did show a tendency to smooth the model output, but found no significant difference between the structures they investigated.

8.1.3 Dynamic flood model

Using the same methodology for calibration of a dynamic flood model presents a considerably larger challenge than the previous examples, since the Gaussian process representation is essentially spatially inspired, being derived from geostatistical techniques. Two significant differences arise when dealing with the output of time-varying rather than spatially-varying models; one is that rate of change of the output may be more rapid by comparison with the time scale of interest, and the second is that the dependency in the output is on historical information alone, whereas spatial models involve dependency in any spatial direction. These complications are compounded when an emulator or a model inadequacy function are to be identified. Thus, the solutions proposed in the literature involve treating the time domain as another spatial dimension for models with simple variation, or creating emulators and model inadequacy function for a single step of the computer model. An alternative is to use time series methods for the time domain, and to link the time series with Gaussian process techniques. However, care is required in this last approach, since the end goal is not forecasting, but risk analysis, where calibrated prediction is required for the full distribution of possible input time series.

The method chosen here is to parameterise the transfer function from input to output time series, and to define the model inadequacy in terms of this parameterisation. It did not prove feasible to apply the calibration directly in the time domain, so the parameterisation of the transfer function was done off-line, and calibration was performed directly in the domain of the parameterisation. The resulting calibration problem reduces to the same class of spatial problems as the previous examples. By separating the transfer function parameterisation from the calibration, and in particular by using optimisation to perform the parameterisation, the errors were minimised during the process, thus overriding any prior knowledge which may exist about observation error variance, and potentially introducing bias in the model parameterisation.

The calibration of this model was found to be influenced by the prior distribution of a parameter in the Gaussian process covariance matrix representing the smoothness of the parameterisation of the transfer function. While this is easy to estimate when the transfer

function parameterisation is fairly irregular, it is less so for a smooth transfer function. This represents a potential weakness of the method.

Two features are required for a calibration method to be suitable for risk analysis. The first is that the calibrated model should be capable of being applied to input time series other than the time series used in calibration. This was ensured by defining the model inadequacy in terms of a transfer function, and was demonstrated for a second, validation time period. The second requirement is that the calibration can be extended to a perturbed model to be able to compare different flood management options, or other future changes. The demonstration of this capability was less convincing, as it would require a more complex model to validate the perturbation, in order to determine what size of perturbation it is feasible to use. A coherent comparison would employ a formal combination of the output from the different computer models. Kennedy and O'Hagan (2000) suggested a method of combining the output from computer programs at different levels of complexity, while Rougier et al. (2009) demonstrated the combination of ensembles of model output from different climate models, using Gaussian process emulators.

8.1.4 Risk calculation

Once it has been demonstrated that calibrated prediction can be produced for other inputs than that used in calibration, it becomes feasible to undertake a risk analysis. In the case of this thesis, the probability of inundation for an existing channel was compared with a modified channel.

Some effort was required to characterise the possible input states, in other words the historical upstream flow patterns, and to determine appropriate statistical distributions for the parameters that described these flow patterns. A sensitivity analysis showed that the output peak flood height is dependent on the entire upstream flood peak. For a given flood peak, dependency on other uncertain quantities that determine flood depth is more complex. Where no model inadequacy has been included, predicted output peak height depends directly on the model calibration parameter, in this case Manning's n . However, where there is a model inadequacy function, dependence of the predicted output peak height is sensitive to several input factors, including the parameters defining the model inadequacy function, indicating that its contribution to the calibrated prediction, and therefore to the risk analysis, is significant.

8.1.5 General comments

Other than optimisation, the automatic calibration method currently in most widespread use in hydrology is the Generalised Likelihood Uncertainty Estimation method of Beven and Binley (1992). The relaxation of the necessity of a formal likelihood in this method, and the possibility of discarding model runs deemed non-behavioural, means that the method is easy to use, but also that its results are somewhat arbitrary. In addition, the

failure of the method to distinguish between different error sources potentially leads to biased parameterisations.

The difficulty of the calibration problem is indicated by the fact that no calibration methodology has been adopted that addresses the well-known deficiencies of GLUE. While the Bayesian methodology proposed and demonstrated in this thesis appears to offer a clear improvement over GLUE for the calibration of steady state models using satellite image information, comparison of the steady state calibration using the current methodology with one undertaken using GLUE does not show a large difference in the flood probability map. This similarity could be used as an argument that the difficulties in applying the current method are not justified by an obvious difference in the results. However, the example of calibration of a dynamic model of the river Severn shows a clear advantage for a method which can incorporate bias correction.

It cannot be denied that there are difficulties in using the current method. The statistical model equations take some effort to understand, and are not straightforward to program. The use of Markov chain Monte Carlo is undeniably not straightforward, and cannot be entirely automated, although the flood extent model does not pose a difficult problem for the calculation method. Covariance matrix stability is a significant problem for adjacent input or output locations, and necessitates the discarding of model output or, more seriously, of data. The choice of prior distributions did not pose a difficult problem for the most part, although it did require some thought.

The greatest need for model calibration in hydrology is for calibration of dynamic models, using gauged data. The work in this thesis represents a partial solution of this problem. No other calibration study of a physically based hydrological model has demonstrated bias correction, and at the same time produced output suitable for risk analysis.

Integral to the current dynamic calibration methodology is the definition of an emulator. While an emulator is not an absolute necessity for model calibration, it is not really practicable to calibrate a complex physically-based spatio-temporal model without one. However, the emulator used here, while both parsimonious and effective, has posed difficulties in the formulation of the error model in the time domain, undermining the statistical credibility of the calibration method.

8.2 Recommendations for further work

The study in this thesis has demonstrated a feasible method for calibration of hydraulic models in the presence of model inadequacy, and has indicated how this calibration method can be incorporated in a risk analysis calculation. However, in order to be of significant practical use in the improvement of risk analysis, a number of issues need to be addressed:

Calibration has been demonstrated separately using spatial data, and using temporally varying gauged data. Pappenberger et al. (2005) noted an improvement in parameter

determination where both spatial and temporal data are used for calibration.

Model inadequacy represents a major potential source of bias in the output of a calibration. However, there are other sources of bias which have not been addressed in this study. In particular, the calibration and output of a model is affected by errors in the determination of input forcing. This error source is most significant where the model input is rainfall, which varies spatially to an extent that it cannot be precisely determined by a typical raingauge network, or flow, which may be determined by rating curve thus introducing bias, or where there is significant unmeasured lateral inflow. Naturally, there may well be an identifiability issue here, because of the number of uncertain quantities relative to the information contained in the observations (Renard et al., 2010). Problems of identifiability represent a fundamental limitation in the absence of more, and more accurate, observations.

The proposed methodology should be extended to different types of models; specifically rainfall-runoff models, where the input is rainfall, and the model represents the hydrological behaviour of the entire catchment, thus allowing a broader range of flood defence measures to be considered. Besides the issue of input errors, another potential problem with the calibration of rainfall-runoff models is the number of parameters used in these models. It is likely that the distributions of some of these parameters may be correlated, causing further difficulties in solution.

A full risk analysis requires consideration of long term consequences of processes of change; this necessitates analysis of the impacts of climate change. To address these processes, analysis will entail a sequence of models; climate change models, a rainfall model, a rainfall-runoff model, a hydraulic model, a model of flood defence failure, and a model of propagation of a flood wave through the floodplain. Each of these modelling stages will introduce uncertainties and errors into the final risk calculation. The errors need to be propagated through the modelling cascade.

It was demonstrated that in the extension of the calibration to a modified channel, it is unclear at what point a modification is too large for the calibration to remain credible. Under these circumstances, it would be helpful to use a second model, possibly a more complex one, as arbitrator. Kennedy and O'Hagan (2000) considered the use of a simpler model to serve as an emulator for a more complex model. An alternative might be to calibrate two models in parallel, using the more complex model to inform the emulation and calibration of the simpler model.

Finally, the work in this thesis has concentrated on the contribution of fluvial flood models to flood risk analysis. Similar techniques can be used for calibration of other models, both for risk analysis, and for other applications.

References

- Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E. and Rasmussen, J., 1986. An introduction to the European hydrological system - Systeme Hydrologique Europeen, "SHE", 2: Structure of a physically-based, distributed modelling system. *Journal of Hydrology*, 87, 61-77.
- Ajami, N.K., Duan, Q. and Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, 43, W01403.
- Anderson, T.W., 1958. *An introduction to multivariate statistical analysis*. Wiley, New York.
- Andrieu, C. and Thoms, J., 2008. A tutorial on adaptive MCMC. *Statistics and Computing*, 18, 343-373.
- Apel, H., Thielen, A.H., Merz, B. and Blöschl, G., 2004. Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Sciences*, 4, 295-308.
- Apel, H., Thielen, A.H., Merz, B. and Blöschl, G., 2006. A probabilistic modeling system for assessing flood risks. *Natural Hazards*, 38, 79-100
- Apel, H., Merz, B. and Thielen, A.H., 2008. Quantification of uncertainties in flood risk assessments. *International Journal of River Basin Management*, 6(2), 149-162.
- Aronica G., Bates, P.D. and Horritt, M.S., 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes*, 16, 2001-2016.
- Bastos, L.S. and O'Hagan A., 2009. Diagnostics for Gaussian process emulators. *Technometrics*, 51(4), 425-438.
- Batchelor, G.K., 1967. *An introduction to fluid dynamics*. Cambridge.
- Bates, P.D. and de Roo, A.P.J., 2000. A simple raster-based model for flood inundation simulation. *Journal of Hydrology* 236, 54-77.
- Bates, P.D., Horritt, M.S., Aronica, G. and Beven, K., 2004. Bayesian updating of flood inundation likelihoods conditioned on flood extent data. *Hydrological Processes*, 18, 3347-3370.
- Bathurst, J.C., 1986. Physically-based distributed modelling of an upland catchment using the Systeme Hydraulique Europeen. *Journal of Hydrology*, 87, 79-102.
- Bayarri, M.J., Berger, J.O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R.J., Paulo, R., Sacks, J., Walsh, D., 2007. Computer model validation with functional output. *Annals of Statistics*, 35(5), 1874-1906.

- Beaumont, M.A., Zhang, W. and Balding, D.J., 2002. Approximate Bayesian Computation in Population Genetics. *Genetics*, 162, 2025-2035.
- Best, N.G., Cowles, M.K. and Vines, K., 1995. *CODA: Convergence diagnostics and output analysis software for Gibbs sampling output, version 0.30*, Technical report, University of Cambridge, MRC Biostatistics unit.
- Beven, K. and Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, 6, 279-298.
- Beven, K.J., Young, P.C., Leedal, D.T. and Romanowicz, R., 2009. Computationally efficient flood water level prediction (with uncertainty). In: P. Samuels, S. Huntingdon, W. Allsop and J. Harrop eds., *Flood risk management : research and practice*. CRC Press, London.
- Bhattacharya, S., 2007. A simulation approach to Bayesian emulation of complex dynamic computer models. *Bayesian Analysis*, 2(4) 783-816.
- Boorman, D.B., Hollis, J.M. and Lilly, A., 1995. *Hydrology of Soil Types: a hydrologically based classification of the soils of the UK* Institute of Hydrology Report 126.
- Box, G.E.P. and Cox, D.R., 1964. An analysis of transformations. *Journal of the Royal Statistical Society, series B*, 26(2), 211-252.
- Boyle, D.P., Gupta, H.V. and Sorooshian S., 2000. Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resources Research*, 36(12), 3663-3674.
- Campolongo F., Cariboni, J. and Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models *Environmental modelling and Software*, 22, 1509-1518.
- Chow, V.T., Maidment, D.R. and Mays, L.W., 1988. *Applied Hydrology*. McGraw-Hill.
- Ciarapica, L. and Todini, E., 2002. TOPKAPI: a model for the representation of the rainfall-runoff process at different scales. *Hydrological Processes*, 16, 207-229.
- Coles, S., 2001. *An introduction to statistical modelling of extreme values*. Springer.
- Conti, S. Gosling, J.P., Oakley, J. and O'Hagan, A., 2009. Gaussian process emulation of dynamic computer codes. *Biometrika*, 96, 663-676.
- Coomber, J.R., 2006. Natural and large catastrophes - changing risk characteristics and challenges for the insurance industry. *The Geneva Papers*, 31, 88-95.
- Cowles, M.K. and Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91, 883-904.
- Craig, P.S., Goldstein, M., Rougier, J.C. and Seheult, A.H., 2001. Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96(454), 717-729.
- Dawson, R., Hall, J., Sayers, P., Bates, P. and Rosu, C., 2005. Sampling-based flood risk analysis for fluvial dike systems. *Stochastic Environmental Research and Risk Assessment*, 19: 388-402.
- Degasperi, A. and Gilmore, S., 2008. Sensitivity Analysis of Stochastic Models of Bistable Biochemical Reactions. In: M. Bernardo, P. Degano and G. Zavattaro eds., *Formal Methods in Computational Systems Biology*. Springer.

- Deutsch, C.V. and Journel, A.G., 1998. *GSLIB: Geostatistical software library and user's guide*. O.U.P.
- Di Baldassarre, G., Castellarin, A., Montanari, A. and Brath, A., 2009. Probability-weighted hazard maps for comparing different flood risk management strategies: a case study. *Natural Hazards*, 50, 479-496.
- Drécourt, J.-P., Madsen, H. and Rosbjerg, D., 2006. Bias aware Kalman filters: comparison and improvements. *Advances in Water Resources*, 29, 707-718.
- Duan, Q., Ajami, N.K., Gao, X. and Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging *Advances in Water Resources*, 30, 1371-1386.
- Duan, Q., Sorooshian, S. and Gupta, V.K., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28, 1015-1031.
- Environment Agency, 2007. *Flooding in Shrewsbury: joint flood action plan*. Environment Agency.
- Environment Agency, 2009a. *Flood risk and coastal management in England*. Environment Agency.
- Environment Agency, 2009b. *River Severn catchment flood management plan - final report*. Environment Agency.
- Environment Agency, 2010. *Station summary: Severn at Montford (54005)*. <http://www.environment-agency.gov.uk/hiflows/station.aspx?54005> accessed 9 Aug, 2010
- Ervine, D.A., Willets, B.B., Sellin R.H.J. and Lorena, M., 1993. Factors affecting conveyance in Meandering Compound Flows. *Journal of Hydraulic Engineering*, 119, 1383-1399.
- European Union, 2007. *Directive 2007/60/EC of the European Parliament and of the Council, of 23 October 2007 on the assessment and management of floods*. Available from: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007L0060:EN:NOT>
- Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4), 343-367.
- Evensen, G. and van Leeuwen, P.J., 2000. An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128, 1852-1867.
- Ewen, J., Parkin, G. and O'Connell, P.E., 2000 SHETRAN: Distributed River Basin Flow and Transport Modelling System. *ASCE Journal of Hydrologic Engineering* 5, 250-258.
- Ewen, J., O'Donnell, G.M., Burton, A. and O'Connell P.E., 2006. Errors and uncertainty in physically-based rainfall-runoff modelling of catchment change effects. *Journal of Hydrology*, 330(3-4), 641-650.
- Feyen, L., Vrugt, J.A., Ó Nualláin, B., van der Knijff, J. and De Roo, A., 2007. Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. *Journal of Hydrology*, 332(3-4), 276-289.
- Fiering, M.B., 1967. *Streamflow synthesis*. Harvard Univ. Press, Cambridge, Mass.
- Fill, H.D. and Steiner, A.A., 2003. Estimating instantaneous peak flow from mean daily flow data. *ASCE Journal of Hydraulic Engineering*, 109, 549-563.

- Freeze, R.A. and Harlan, R.L., 1969. Blueprint for a physically-based, digitally-simulated hydrologic response model. *Journal of Hydrology*, 9, 237-258.
- Gaggiotti O.E., 2010. Preface to the special issue: advances in the analysis of spatial genetic data. *Molecular Ecology Resources*, 10, 757-759.
- Gelman, A. and Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7, 457-511.
- Georgakakos, K.P., D.-J. Seo, H. Gupta, J. Schaake and M.B. Butts, 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, 298, 222-241.
- Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds. *Bayesian Statistics 4*, O.U.P.
- Goldstein, M. and Rougier, J.C., 2004. Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM Journal on Scientific Computing*, 26(2), 467-487.
- Goldstein, M. and Rougier, J.C., 2006. Bayes linear calibrated prediction for complex systems. *Journal of the American Statistical Association*, 101(475), 1132-1143.
- Goldstein, M. and Rougier, J.C., 2009. Reified Bayesian Modelling and Inference for Physical Systems (with discussion and rejoinder). *Journal of Statistical Planning and Inference*, 139(3), 1221-1239.
- Goldstein, M. and Wooff, D., 2007. *Bayes Linear Statistics, Theory and Methods*. Wiley. ISBN 978-0-470-01562-9
- Haario, H., Laine, M., Mira, A. and Saksman, E., 2006. DRAM: Efficient adaptive MCMC. *Statistics and Computing*, 16, 339:354.
- Haario, H., Saksman, E. and Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, 14, 375-395.
- Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223-242.
- Hall, J.W., Dawson, R. J., Sayers, P. B., Rosu, C., Chatterton, J. B. and Deakin, R., 2003. A methodology for national-scale flood risk assessment. *Proceedings of the ICE - Maritime Engineering*, 156(3), 235-247.
- Hall, J.W., Manning, L.J. and Hankin, R.K.S., 2011. Bayesian calibration of a flood inundation model using spatial data. *Water Resources Research*, in press.
- Hall, J.W., Tarantola, S., Bates, P.D. and Horritt, M.S., 2005. Distributed sensitivity analysis of flood inundation calibration. *Journal of Hydraulic Engineering*, 131(2), 117-126.
- Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2, 84-90.
- Hankin, R.K.S., 2005. Introducing BACCO, an R bundle for Bayesian analysis of computer code output. *Journal of Statistical Software*, 14(16).
- Harrison, P.J. and Stevens, C.F., 1976. Bayesian Forecasting (with discussion). *Journal of the Royal Statistical Society, series B*, 38, 205-247.

- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, 97-109.
- Heidelberger, P. and Welch, P.D., 1981. A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM*, 24, 233-245.
- Higdon, D., Gattiker, J., Williams, B. and Rightley, M., 2008a. Computer Model Calibration Using High-Dimensional Output. *Journal of the American Statistical Association*, 103(482), 570-583.
- Higdon, D., Kennedy, M.C., Cavendish, J., Cafoe, J. and Ryne, R.D., 2004. Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2), 448-466.
- Higdon, D., Nakhleh, C., Gattiker, J. and Williams, B., 2008b. A Bayesian calibration approach to the thermal problem. *Computer Methods in Applied Mechechanics and Engineering*, 197, 2431-2441.
- Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. *Statistical Science*, 14(4), 382-417.
- Horritt, M.S., Mason, D.C. and Luckman, A.J., 2001. Flood boundary delineation from Synthetic Aperture Radar imagery using a statistical active contour model. *International Journal of Remote Sensing*, 22(13), 2489-2507.
- House, L., 2009. *An application of reification to a rainfall-runoff computer model* Unpublished internal report 2.1.4, Managing Uncertainty in Computer Models research project.
- Huard, D. and Maillot, A., 2006. A Bayesian perspective on input uncertainty in model calibration: application to hydrological model 'abc'. *Water Resources Research*, 42, W07416.
- Iman, R.L. and Conover, W.J., 1982. A distribution-free approach to inducing rank correlation among input variables. *Communications in statistics - simulation and computation*, B11, 311-334.
- International Disaster Database, 2010. www.emdat.net, accessed 6-05-2010.
- Jacques, J., Lavergne, C. and Devictor, N., (2006). Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering and System Safety*, 91, 1126-1134.
- Jeffreys, H., 1961. *Theory of probability* Oxford: Clarendon Press
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME Journal of Basic Engineering, (Series D)*, 82, 35-45.
- Kavetski, D., Kuczera, G. and Franks, S.W., 2002. Confronting input uncertainty in environmental modelling. In Q. Duan, H.V. Gupta, S. Sorooshian, A.N. Rousseau and R. Turcotte (eds) *Calibration of watershed models*. AGU, Washington.
- Kavetski, D., Kuczera, G. and Franks, S.W., 2006. Bayesian analysis of input uncertainty in hydrological modelling 2: Application. *Water Resources Research*, 42, W03408.
- Kennedy, M.C. and O'Hagan, A., 2000. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87, 1-13.
- Kennedy, M.C. and O'Hagan, A., 2001a. Bayesian calibration of computer models. *Journal of the Royal Statistical Society, series B*, 63(3), 425-464.

- Kennedy, M.C. and O'Hagan, A., 2001b. Supplementary details on Bayesian calibration of computer models. Available at www.shef.ac.uk/~st1ao/ps/calsup.ps
- Knight, D.W. and Sellin, R.H.J., 1987. The SERC Flood Channel Facility. *Journal of the Institution of Water and Environmental Management*, 1(2), 198-204.
- Knight, D.W., Shiono, K., 1996. River Channel and Floodplain Hydraulics. In M.G. Anderson, D.E. Walling and P.D. Bates (eds), *Floodplain Processes*. Wiley.
- Kollat J.B., Reed, P.M. and Rizzo, D.M., 2008. Assessing model bias and uncertainty in three-dimensional groundwater transport forecasts for a physical aquifer experiment. *Geophysical Research Letters*, 35, L17402.
- Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, 35(9), 2739-2750.
- Krzysztofowicz, R., 2002. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, 268, 16-40.
- Krzysztofowicz, R. and Herr, H.D., 2001. Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. *Journal of Hydrology*, 249, 46-68.
- Krzysztofowicz, R. and Kelly, K.S., 2000. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research*, 36(11), 3265-3277.
- Kuczera, G., Kavetski, D., Franks, S.W. and Thyer, M., 2006. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterising model error using storm-dependent parameters. *Journal of Hydrology*, 331, 161-77.
- Lin, Z. and Beck, M.B., 2007a. On the identification of model structure in hydrological and environmental systems. *Water Resources Research*, 43, W02402.
- Lin, Z. and Beck, M.B., 2007b. Understanding dual complex environmental systems: a dual approach. *Environmetrics*, 18, 11-26.
- Linsley, R.K.Jr., Kohler, M.A. and Paulus, J.L.H., 1988. *Hydrology for engineers*. SI edition, McGraw-Hill
- Little, J., Goldstein, M. and Jonathan P., 2004. Efficient Bayesian sampling inspection for industrial processes based on transformed spatio-temporal data. *Statistical Modelling*, 4, 299-313.
- Liu, F. and West, M., 2009. A dynamic modelling strategy for Bayesian computer model emulation. *Bayesian Analysis*, 4(2), 393-412.
- Manning, L.J. and Hall, J.W., 2010. A Bayesian approach to analysing structural uncertainties in flood inundation models. Proceedings of the 1st European IAHR Congress, Edinburgh 2010, 4th-6th May 2010.
- Mantovan, P. and Todini, E., 2006. Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *Journal of Hydrology*, 330(1-2) 368-381.
- Marrel, A., Iooss, B., da Viegua, S. and Ribatet, M., 2011. Global sensitivity analysis of stochastic computer models with joint metamodels. *Statistics and Computing*, in review. <http://fr.arxiv.org/abs/0911.1189>
- Marshall, L., Nott, D. and Sharma, A., 2004. A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research*, 40(2), W02501.

- McKay, M.D., 1995. *Evaluating prediction uncertainty*. Internal report NUREG/GR-6311 Los Alamos National Laboratory.
- McKay, M.D., Conover, W.J. and Beckman, R.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21, 239-245.
- McNeall, D.J., 2008. *Dimension reduction in the Bayesian analysis of a numerical climate model*. Doctoral Thesis, University of Southampton.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller A.H. and Teller, E., 1953. Equations of State calculations by fast computing machine. *Journal of Chemical Physics*, 21, 1087-1091.
- Moradkhani, H., Sorooshian S., Gupta, H.V., Houser, P., 2005a. Dual State-Parameter Estimation of Hydrological Models using Ensemble Kalman Filter. *Advances in Water Resources*, 28(2),135-147.
- Moradkhani, H., Hsu, K., Gupta, H. V. and Sorooshian, S., 2005b. Uncertainty Assessment of Hydrologic Model States and Parameters: Sequential Data Assimilation Using Particle Filter. *Water Resources Research*, 41, W05012.
- Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2), 161-174.
- Montanari, A., 2005. Large sample behaviors of GLUE in assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, 41, W08406.
- Moore, R.J. and Clark, R.T., 1981. A distribution function approach to rainfall runoff modeling *Water Resources Research*, 17(5), 1367-1382.
- Myers W.R.C. and Brennan E.K., 1990. Flow resistance in compound channels, *Journal of Hydraulic Research*, 28(2),141-155.
- Nash, J.E., 1959. Systematic determination of unit hydrograph parameters. *Journal of Geophysical Research*, 64, 111-115.
- Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part 1 - A discussion of principles. *Journal of Hydrology*, 10, 282-290.
- Neuman, S.P., 2003. Maximum likelihood Bayesian averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment*, 17,(5), 291-305.
- Oakley, J., 2002. Eliciting Gaussian process priors for complex computer codes. *The Statistician*, 51, 81-97.
- Oakley, J. and O'Hagan, A., 2002. Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4), 769-784.
- Oakley, J. and O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society, series B*, 66(3), 751-769
- O'Hagan, A. and Forster, J., 2004. *Kendall's advanced theory of statistics: Vol 2B Bayesian inference (2nd edition)*. Arnold.
- Office of Public Sector Information, 2009. *Flood Risk Regulations 2009*. http://www.opsi.gov.uk/si/si2009/uksi_20093042_en_1 Accessed 13-06-2010

- Office of Public Sector Information, 2010. *Flood and Water Management Act, 2010, Chapter 29*. http://www.opsi.gov.uk/acts/acts2010/pdf/ukpga_20100029_en.pdf Accessed 13-06-2010
- Owen, A.B., 1994. Controlling Correlations in Latin Hypercube Samples. *Journal of the American Statistical Society* 89, 1517-1522.
- Pappenberger, F., Beven, K., Horritt, M. and Blazkova, S., 2005. Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations *Journal of Hydrology*, 302, 46-69.
- Parkin, G., O'Donnell, G.M., Ewen, J., Bathurst, J.C., O'Connell, P.E. and Lavabre, J., 1996. Validation of catchment models for predicting land-use and climate change impacts. 2: Case study for a Mediterranean catchment. *Journal of Hydrology*, 175, 595-613.
- Plummer, M., Best, N., Cowles, M. and Vines, K., 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R-news*, 6(1) 7-11.
http://CRAN.R-project.org/doc/Rnews/Rnews_2006-1.pdf (accessed 27-06-2010)
- Price, R.K., Fernando, W.A.Y.S and Solomantine, D.P., 2006. Inverse modelling for flood propagation. In: P. Goubesville, J. Cunge, V. Guinot and S. Y. Liong, eds, *Proceedings of the 7th International Conference in Hydroinformatics*, Research Publishing Services, Chennai, India.
- R Development Core Team, 2009. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Raftery, A.E. and Lewis, S.M., 1996. Implementing MCMC. In W.R. Gilks, S. Richardson and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* Chapman and Hall.
- Reichert P. and Mieleitner, J., 2009. Analyzing input and structural uncertainty of non-linear dynamic models with stochastic, time-dependent parameters. *Water Resources Research*, 45, W10402.
- Renard, B., Kavetski, D. and Kuczera, G., 2009. Comment on "An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction". *Water Resources Research*, 45, W03603.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modelling: the challenge of identifying input and structural errors. *Water Resources Research*, 46, W05521.
- Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, 7(1), 110-120.
- Robson, A.J. and Reed, D.W., 1999. *Statistical procedures for flood frequency estimation*. Volume 3 of the Flood Estimation Handbook. Centre for Ecology and Hydrology
- Rojas, R., Feyen, L. and Dassargues A., 2008. Conceptual model uncertainty in groundwater modeling: combining generalized likelihood uncertainty estimation and Bayesian model averaging. *Water Resources Research*, 44, W12418.

- Romanowicz, R., Beven, K.J. and Tawn J., 1996. Bayesian calibration of flood inundation models. In M.G. Anderson, D.E. Walling and P.D. Bates (eds), *Floodplain Processes*. Wiley.
- Romanowicz, R.J., Young, P.C., Beven, K.J. and Pappenberger, F., 2008. A data based mechanistic approach to nonlinear flood routing and adaptive flood level forecasting. *Advances in Water Resources*, 31, 1048-1056.
- Rougier, J.C., 2008. Efficient emulators for multivariate deterministic functions. *Journal of Computational and Graphical Statistics*, 17(4), 827-843.
- Rougier J.C., Sexton, D.M.H., Murphy, J.M. and Stainforth, D., 2009. Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. *Journal of Climate*, 22, 3540-3557.
- Sacks J., Welch W.J. Mitchell T.J. and Wynn, H.P., 1989. Design and analysis of computer experiments. *Statistical Science*, 4(4), 409-435.
- Sallaberry, C.J., Helton, J.C. and Hora, S.C., 2008. Extension of Latin hypercube samples with correlated variables. *Reliability Engineering and System Safety*, 93, 1047-1059.
- Sellin R.H.J., 1964. A laboratory investigation into the interaction between the flow in the channel of a river, and that over its floodplain. *La Houille Blanche*, 7, 793-801.
- Sivapalan, M., Blöschl, G., Zhang, L. and Vertessy, R., 2003. Downward approach to hydrological prediction *Hydrological Processes*, 17(11), 2101-2111.
- Smith, M.B., Seo, D.-J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F. and Cong, S., 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. *Journal of Hydrology*, 298(1-4), 4-26.
- Sobol', I.M., 1998. On quasi-Monte Carlo algorithms *Mathematics and computers in simulation*, 47, 103-112.
- Sorooshian, S. and Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water Resources Research*, 16, 430-442.
- Stedinger, J.R., Vogel, R.M., Lee, S.U. and Batchelder, R., 2008. Appraisal of Generalized Likelihood Uncertainty Estimation (GLUE) methodology. *Water Resources Research*, 44, W00B06.
- Stigter, J.D. and Beck, M.B., 2004. On the development and application of a continuous-discrete recursive prediction error algorithm. *Mathematical Biosciences*, 191, 141-158.
- Taylor, C.J., Pedregal, D.J., Young, P.C. and Tych, W., 2007. Environmental time series analysis and forecasting with the Captain toolbox. *Environmental Modelling and Software*, 22, 797-814.
- ter Braak, C.J.F., 2006. A Markov chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16, 239-249.
- Thyer, M.A., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W. and Srikanthan, S., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modelling: a case study using Bayesian total error analysis. *Water Resources Research*, 45, W00B14.

- Todini, E., 1996. The ARNO rainfall-runoff model. *Journal of Hydrology*, 175(1-4), 339-382.
- Todini, E., 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. *Journal of River Basin Management*, 6(2), 123-137.
- U.S. Army Corps of Engineers, 2002. *Hec Ras River Analysis System*.
<http://www.hec.usace.army.mil/software/hecras/index.html>
- Vrugt, J.A., Gupta, H.V., Bouten, W. and Sorooshian, S., 2003. A shuffled complex evolution metropolis algorithm for optimisation and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8), 1201.
- Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M. and Robinson, B.A., 2008. Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, 44, W00B09.
- Wagner, T., McIntyre, N., Lees, M.J., Wheater, H.S. and Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes*, 17(2), 455-476.
- Werner, M.G.F., Hunter, N.M. and Bates, P.D., 2005. Identifiability of distributed floodplain roughness values in flood extent estimation. *Journal of Hydrology*, 314, 139-157.
- West, M. and Harrison, J. 1989. *Bayesian forecasting and dynamic models*. Springer, New York.
- Wilkinson, R.D., 2010. Bayesian calibration of expensive computer experiments. In L.T. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B.K. Mallick, L. Tenorio, B. Van Bloemen Waanders and K. Wilcox eds. *Large scale inverse problems and quantification of uncertainty*, Wiley.
- Wood, E.F., Lettenmaier, D.P. and Zartarian, V.G., 1992. A land-surface hydrology parameterization with subgrid variability for General Circulation Models. *Journal of Geophysical Research*, 97, 2717-2728.
- Woodhead, S.P.B., 2007. *Bayesian calibration of flood inundation simulators using an observation of flood extent*. Doctoral thesis, University of Bristol.
- Wynn, H.J., 2001. Contribution to the discussion on the paper by Kennedy and O'Hagan. *Journal of the Royal Statistical Society, series B*, 63(3), 425-464.
- Young, P.C., 1984. *Recursive estimation and time series analysis: an introduction*. Springer-Verlag.
- Young, P.C., 2003. Top-down and data-based mechanistic modelling of rainfall-flow dynamics at catchment scale. *Hydrological Processes*, 17(11), 2195-2217.
- Zhao, R.J., 1992. The Xinanjiang model applied in China, *Journal of Hydrology*, 135, 371-381.