QUICS: Quantifying Uncertainty in Integrated Catchment Studies

D1.3 Model structure errors and input uncertainty: Lessons Learned

By: Omar Wani, ESR5

Lead Partner: Eawag

Revision: 31.05.2017

Report Details

Title: Model structure errors and input uncertainty: Lessons Learned

Deliverable Number: D1.3

Author(s): Omar Wani

Dissemination Level: Public

Document History

Version	Date	Status	Submitted by	Checked by	Comment
v 1.0	07/12/2016	Draft First	Jörg Rieckermann	Will Shepherd	
v 1.1	31/05/2017	Final Version	Omar Wani	Jeroen Langeveld	

Acknowledgements

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000.

Executive Summary

This report is intended to serve as an opinion piece – based on the lessons learned in the past few years by the author related to the proper quantification and reduction of uncertainties in hydrologic model predictions.

CONTENTS

1 Introduction

Going by Newtonian mechanics, if we know the state of any hydrologic phenomenon exactly, we should be, in principle, able to predict its future state exactly (Laplace, 1814). However, in practice we make various approximations – among other things, neglecting many processes in the modelling exercise - which result in simulations differing from reality. Therefore, even if we start with the assumptions of classical mechanics, we are beset with uncertainties. There has been extensive research on various sources of uncertainty in environmental modelling and its proper quantification (e.g. Renard et al., 2010; Reichert and Schuwirth, 2012). Epistemic uncertainties, like model structure deficits, and the ignorance about parameter values play a significant role in determining the quality of a model prediction (Del Giudice et al., 2013). It has been documented and researched that oversimplification of models and usage of erroneous inputs reduce the accuracy of environmental predictions (Del Giudice et al., 2015). To put the model predictions to proper use in design and forecast, proper quantification of uncertainty is required.

In this document, we intend to put forth our experience with two uncertainty quantification paradigms:

1) Bayesian Inference: Uncertainty analysis is essentially an attempt to capture the probability of occurrence of events, conditioned on our most updated knowledge of the system. The formal description of such probability can be made using a likelihood function. If used within the Bayesian framework, a likelihood function also allows to learn about model parameters from available data through inference.

2) Post-processors: Other techniques, such as post-processors, are also available to quantify predictive uncertainty, without the need to recalibrate the model. They generally learn about the error processes from the past errors of the model and project that uncertainty into predictions.

We also provide references for those who want to study these paradigms in detail.

2 Likelihood Function Based Uncertainty Estimation

In the recent past, the additive description of bias arising from model structure deficits and uncertainty inputs has been widely studied. Reichert and Schuwirth (2012) describe true system response $\mathrm{Y_t}$ at time t of an environmental system as the sum of a deterministic model *m* and a stochastic process B:

$$\mathrm{Y_t} = m_{\mathrm{t}}(\mathbf{x}, \boldsymbol{\theta}^{\mathbf{m}}) + \mathrm{B_t}(\boldsymbol{\theta}^{\mathbf{B}}). \quad (1)$$

Representing variables by capital letters and write vectors bold. In Eq. (1), the stochastic process B captures the bias of the model due to structural limitations and input errors. The model inputs are denoted with **x** and parameters with $\boldsymbol{\theta}$**.** Knowledge about the physics of a hydrologic system (represented by m) and the error-generating processes can be formulated in terms of a conditional probability distribution, called likelihood function. It is defined as the probability of observing a system response, given a set of parameters and input. Mathematically:

$$\mathrm{p_Y}(\mathrm{y}\,|\mathrm{x}, \theta) \quad (2)$$

Given a set of observations, which constitute the samples from the likelihood function, the parameters of the likelihood function can be inferred using Bayes theorem. It has been suggested to represent the bias **B** as a Gaussian process (Kennedy and O'Hagan, 2001). This makes it possible to write the likelihood function as a multivariate normal distribution with a mean and a covariance matrix.

The likelihood function can also be defined by having a stochastic model for the process itself, or having a deterministic model and introducing stochasticity by a multiplicative random variable. In case of explicit consideration of bias, the deterministic model and the additive bias may be equally able to explain the data. To avoid identifiability problems during inference, proper priors can be chosen. As we would like the model to explain the underlying trend of the hydrologic process, the mean of the bias should be zero.

Once a likelihood function is explicitly defined, it can be also extended for different data types. For example, when censored data is available, inference can be carried out by integrating the likelihood function over correct intervals.

The advantages of using such a paradigm are:

1) Proper quantification of parametric uncertainty. The prior knowledge about the system can be captured by the prior probability distribution and data can be used to update the prior.
2) The likelihood function can be formulated to decompose uncertainty due to model structure deficits, parameters and inputs.
3) The paradigm can be extended to different types of data like censored or binary. (Wani et al., 2017b)

4) The dependence between the parameters gets captured in the posterior distribution, thus helping in proper estimation of parametric uncertainty.

The challenges of using such a paradigm are:

1) The likelihood function needs to be defined properly so that the properties of the error process are adequately represented.
2) The computational costs of running a sampling algorithm, which helps provide parameter samples from the posterior, is higher than a simple optimization algorithm, which finds the best fit parameters. (However, the use of emulators is alleviating this concern. (Carbajal et al., 2017))
3) The specification of priors for the parameters of the error process can be a challenge.

For details on the additive description of bias, please be referred to (Del Giudice et al., 2013)

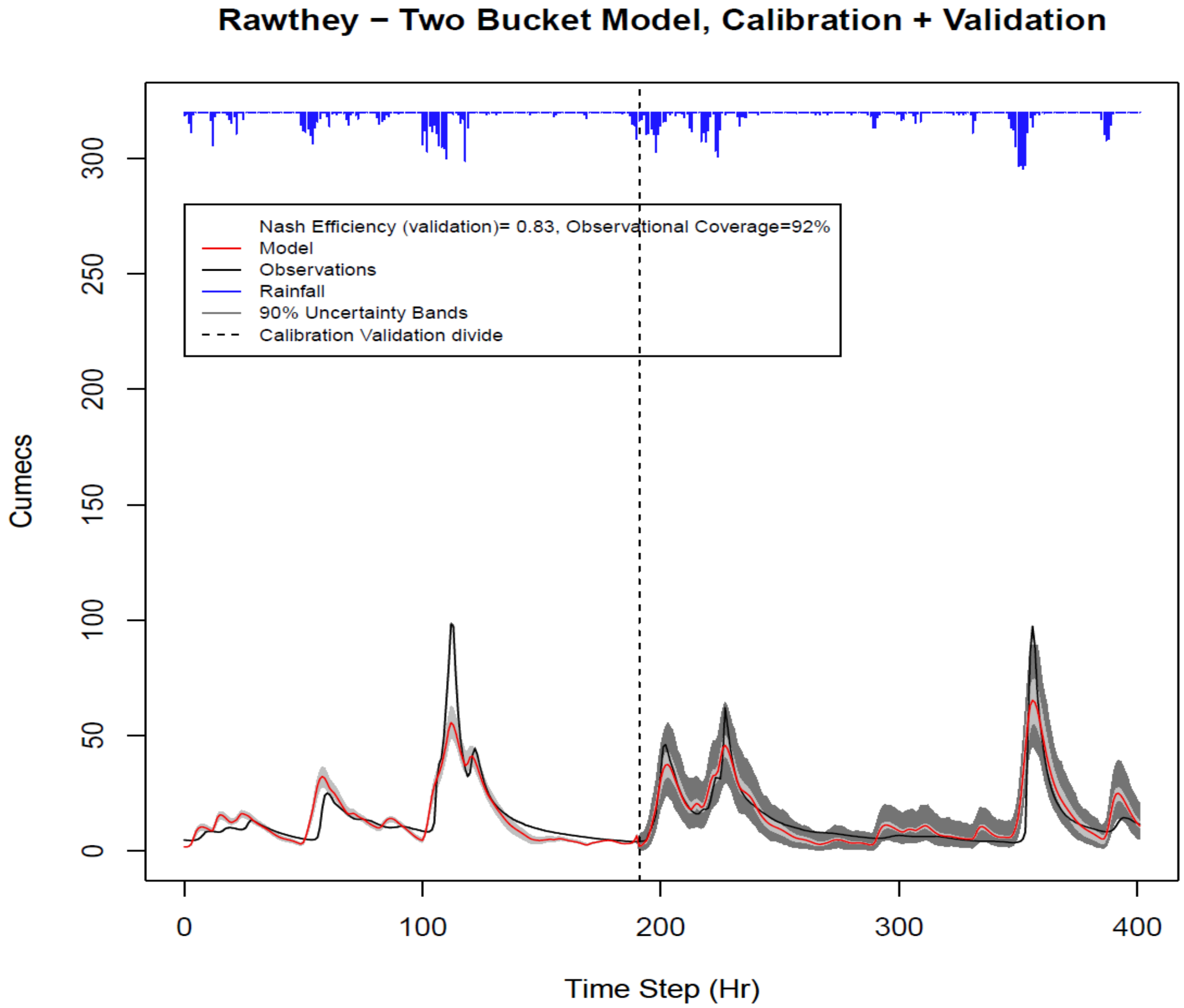

Fig. 1. Inference of model parameters in the calibration phase of a two-bucket conceptual model and the uncertainty intervals in the prediction phase using the likelihood function description in Eq. 1 and 2.

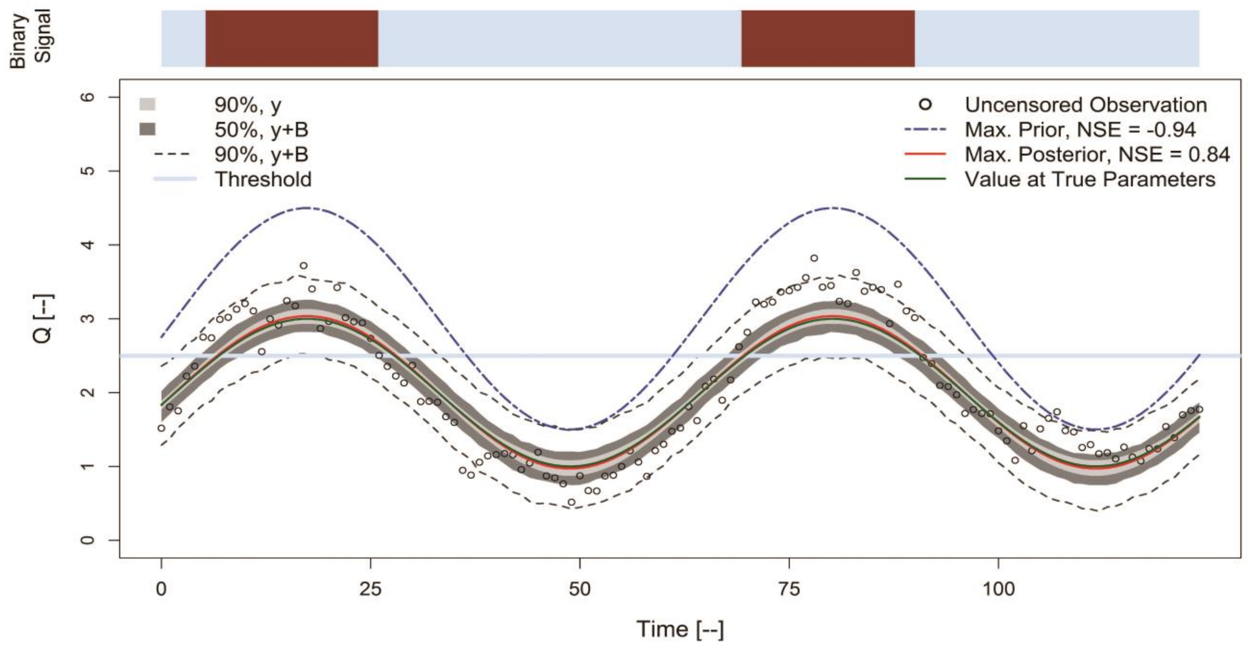

Fig. 2. The improvement in the model performance in the validation phase after inference using five hundred binary data points. The model is a two parameter linear model, with a sinusoidal input. (source: (Wani et al., 2017b))

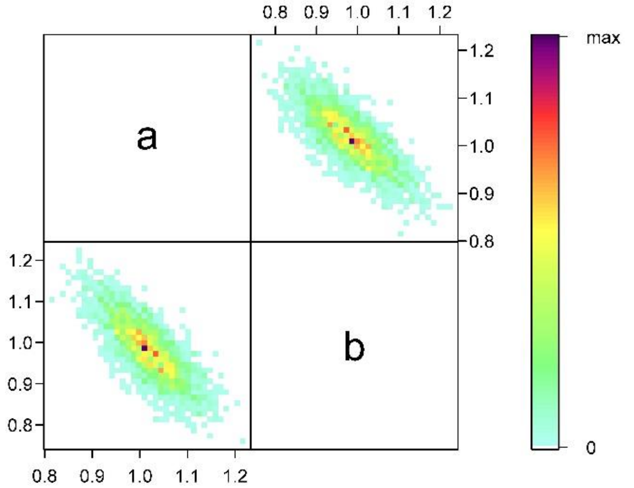

Fig. 3. Bivariate posterior distribution of model parameters a and b of the linear model (Fig. 2). The negative correlation that is not assumed by the prior gets captured in the posterior. (source: (Wani et al., 2017b))

3 Post-Processor Based Uncertainty Estimation

Post-processing uncertainty estimation techniques generally work with a defined set of model parameter values. Once the model structure and the parameter vector is defined, the properties of the errors made by the model for the past time series are used to predict future errors. These techniques are heavily data driven and learn from the statistical properties of model-observation mismatch of the past simulations. One of the simple uncertainty estimation techniques is based on instance-based learning (Wani et al., 2017a). It is a non-parametric method and thus does not make an explicit assumption about the nature of the error distribution. It employs a k-nearest neighbour search for similar historical hydrometeorological conditions and determines the uncertainty intervals from historical errors. This technique has the advantage of being conceptually simple and computationally inexpensive. However, the post processing techniques do not allow for the disaggregation of uncertainty. Also, the uncertainty is not reduced, as no inference is performed. Many of these techniques are generally unable to capture the observational uncertainty.

For more details on the different post-processor uncertainty estimators, please be referred to (Dogulu et al., 2015; Wani et al., 2017a).

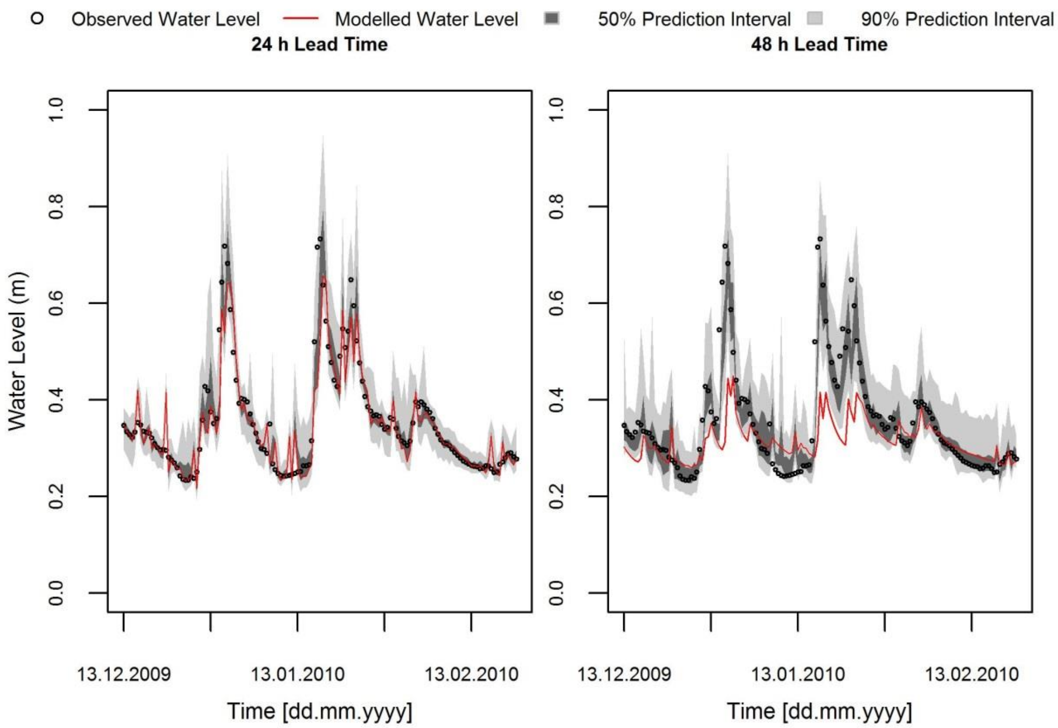

Fig 4. Prediction intervals for a hydrologic catchment case study generated using kNN resampling. The hydrographs are shown for the two different lead times. (source: (Wani et al., 2017a))

4 References

Carbajal, J.P., Leitao, J.P., Albert, C., Rieckermann, J., 2017. Appraisal of data-driven and mechanistic emulators of nonlinear simulators: The case of hydrodynamic urban drainage models. Environ. Model. Softw. 92, 17–27. doi:10.1016/j.envsoft.2017.02.006

Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., Rieckermann, J., 2013. Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. Hydrol. Earth Syst. Sci. 17, 4209–4225. doi:10.5194/hess-17-4209-2013

Del Giudice, D., Reichert, P., Bareš, V., Albert, C., Rieckermann, J., 2015. Model bias and complexity - Understanding the effects of structural deficits and input errors on runoff predictions. Environ. Model. Softw. 64, 205–214. doi:10.1016/j.envsoft.2014.11.006

Dogulu, N., López López, P., Solomatine, D.P., Weerts, A.H., Shrestha, D.L., 2015. Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. Hydrol. Earth Syst. Sci. 19, 3181–3201. doi:10.5194/hess-19-3181-2015

Laplace, Pierre Simon, 1814, translated into English from the original French, 6th ed. by Truscott,F.W. and Emory,F.L. (1951). A Philosophical Essay on Probabilities. Dover Publications, New York.

Reichert, P., Schuwirth, N., 2012. Linking statistical bias description to multiobjective model calibration. Water Resour. Res. 48, 1–20. doi:10.1029/2011WR011391

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resour. Res. 46, 1–22. doi:10.1029/2009WR008328

Wani, O., Beckers, J.V.L., Weerts, A.H., Solomatine, D.P., 2017a. Residual uncertainty estimation using instance-based learning with applications to hydrologic forecasting. Hydrol. Earth Syst. Sci. Discuss. 1–30. doi:10.5194/hess-2017-75

Wani, O., Scheidegger, A., Carbajal, J.P., Rieckermann, J., Blumensaat, F., 2017b. Parameter estimation of hydrologic models using a likelihood function for censored and binary observations. Water Res. 121, 290–301. doi:10.1016/j.watres.2017.05.038