

Interactive comment on “Improving uncertainty estimation in urban hydrological modeling by statistically describing bias” by D. Del Giudice et al.

D. Del Giudice et al.

dariode@student.ethz.ch

Received and published: 22 July 2013

Dear Editor and Reviewers,

Thank you very much for your time and for considering our manuscript and providing positive and detailed feedbacks. Thanks to your valuable input we believe that the manuscript will significantly improve. All comments have been considered (see below). Based on the referees' feedback and internal reviews, we made several changes to the article. These are included in the revised manuscript that will be resubmitted as soon as possible.

C3331

On behalf of the authors, sincerely,

Dario Del Giudice

ETHZ - Swiss Federal Institute of Technology Zurich

Eawag - Swiss Federal Institute of Aquatic Science and Technology

Reviewer # 1

- 1) *Figures should be numbered according to their order of appearance in the text (Figures 4 and 5 are mentioned before Figure 3).*

What originally was Fig. 3 will be displaced after Fig. 5.

- 2) *Figure 4, caption: “The observation errors, being very small for this scale (..), the last two uncertainty bands overlap and only the intermediate grey is visible”. This sentence is not much clear, please rephrase.*

The caption of Fig. 4 will be modified to:

“When considering bias, the contribution of uncorrelated observation errors \mathbf{E} to total uncertainty becomes very small ($\lesssim 1$ l/s) and therefore is not visible at this scale. Consequently, the credibility intervals for the system output ($g^{-1}(\tilde{y}_M + \mathbf{B}_M)$) and the observations ($g^{-1}(\tilde{y}_M + \mathbf{B}_M + \mathbf{E})$) are almost identical and overlapping.”

We also noticed that the caption of Fig. 5 could be improved. Therefore we will rewrite it as:

“Probabilistic runoff predictions for part of the calibration (left) and the validation period (right) with the constant bias model and log-sinh transformation. The input time series (hyetograph) is shown on the top. The observed hydrograph is

C3332

represented by dots, with the triangular data points being used only for validation. The 95% credible intervals are interpreted as follows: parametric uncertainty due to y_M (dark gray), parametric plus input and structural uncertainty due to $g^{-1}(\tilde{y}_M + \mathbf{B}_M)$ (intermediate), total uncertainty due to $g^{-1}(\tilde{y}_M + \mathbf{B}_M + \mathbf{E})$ (light gray). Validation data not included in this light gray region are marked in red. The prediction intervals for the system output and the observations are almost indistinguishable and therefore only the intermediate gray band is visible at this scale.”

- 3) *Section 3.1. I would suggest to add a basic description of the calibration and validation rainfall events (e.g. duration, total precipitation and peak intensity).*

A paragraph describing precipitation events will be added at the end of Section 3.1:

“Calibration storms had a peak intensity ranging from 13 to 65 mm/hr, whereas validation events had a maximum rain rate spanning from 8 to 34 mm/hr. The monitored rainstorms had a duration of 0.5-4 hr with a cumulative height varying from 27 mm to 4.1 m.”

- 4) *Section 5.2. It would be interesting to add some discussion on how physical parameters of the catchment (e.g. size, percentage of impervious areas) could influence uncertainty estimation.*

We are not sure to completely understand this point. If the reviewer refers to the effect of model parameters on uncertainty analysis this is our reply. The parameters of a conceptual lumped model, which can have some physical interpretation (e.g. slope, imperviousness), influence the uncertainty estimation in two ways. From one side, the simulator output directly depends on these parameters. On its turn, the difference between model results and calibration data determines the

C3333

error model parameters and, in this way, influences the uncertainty intervals. On the other side, being in a Bayesian context, the width of the posterior distribution of the simulator parameters contributes to a (small) part of total predictive uncertainty.

Please notice, finally, that the physical properties of the watershed are just approximately and loosely translated into the model, whose inferred parameters slightly change according to the error description.

Since this work is more focused on obtaining satisfactory predictions than on solving the inverse problem of parameter inference, we would prefer omitting this discussion in the paper.

If the reviewer refers to the effect of different catchment propriety on uncertainty estimation this is not possible to respond here since it would require the analysis of several hydrosystems.

Reviewer # 2

- 1) *line 21: spell out “iid” on first use*

iid will be defined on first use as independent and identically distributed. The whole paragraph (lines 15-22) will be however eliminated since redundant.

- 2) *p5125, line 7: the cited study does not use Box-Cox transformation, but instead uses separate models to explicitly and separately model the variance and non-normality in the residuals.*

We agree that the citation was ambiguous. The sentence will therefore be modified to:

“(…) a common treatment in hydrology is to transform simulation results and output data.”

C3334

- 3) p5131, eq.2: I'd suggest using $g(Y)$ for transformed flows, to make the notation more consistent with later sections (same for eq. 4 on next page)

We think that notation using \tilde{Y} is lighter than the one using $g(Y)$. We will modify, however, the notation in later sections (e.g. 2.1.3) to keep it consistent.

- 4) p5136, line 20: "The characteristics of the catchment and the monitoring equipment suggested a setting of $a = 5 \text{ L s}^{-1}$ and $b = 100 \text{ L s}^{-1}$." Can you be more specific how these values were deduced? The same comment applies to selection of Box-Cox transformation parameter. In general it seems that you would want to estimate these parameters directly from the data (i.e. specify a prior and estimate their posteriors); why was this not done here?

We agree that the setting of the transformation parameters could be explain more clearly.

Regarding the non-inference of the calibration parameters, we had two reasons for that, both linked to identifiability problems. First, the parameters of the transformation reflect our assumptions on the error distribution. Calibrating these parameters would be equivalent to fitting the shape of the error distribution, which is seldom performed. Indeed, by increasing the degrees of freedom of our error model, the inference process becomes more and more complex and unstable. Similar considerations can be found in the works of Schoups and Vrugt (2010) and Honti et al. (2013) where the adjustment of the shape of the likelihood function lead to identifiability problems of other error model parameters and unrealistic uncertainty bands. This brings us to our second point. From the experience gained in our previous studies (Frey et al., 2011; Reichert and Mieleitner, 2009; Yang et al., 2007b,a; Sikorska et al., 2012) the estimation of transformation parameters was very challenging and hindering the stability of the MCMC process. We observed the same instabilities when trying to calibrate the transformation

C3335

parameters in a first phase of this study.

Concerning the selection of the Box-Cox parameter λ , we took a value that in the literature is frequently adopted for its good capacity of stabilizing the variance of residuals without excessively compromising simulator matching of the data. λ is scale-independent and therefore well transferable from study to study. Values providing a good compromise between heteroskedasticity reduction and satisfactory model fit usually range from: 0.25 to 0.5. (Willems, 2012; Frey et al., 2011; Reichert and Mieleitner, 2009; Yang et al., 2007b; Honti et al., 2013; Yang et al., 2007a; Sikorska et al., 2012, 2013). For this reason λ was set to an intermediate value of 0.35.

We will emend, for better clarity, the last part of the Box-Cox section as follows: "Profiting from the experience in several hydrological studies (Willems, 2012; Honti et al., 2013; Yang et al., 2007b,a; Wang et al., 2012) and the transferability of this parameter, we choose a $\lambda = 0.35$. Assuming a constant variance in the transformed space, this value yields a moderate increase of variance in non-transformed output. This accounts for an observed increase in residual variance while keeping the weight of high discharge observations sufficiently high for calibration. In other words, this moderate λ assures a good compromise between the performances of the error model and the fit of the simulator. The behavior of the Box-Cox transformation and its derivative for the stormwater runoff in our study are shown in Figs. 1 and S1."

Regarding the setting of the log-sinh parameters we made other considerations, since their value is strongly case- and unit-specific. The attempt to derive sensible values for these parameters without calibrating them is what motivated us to reparametrize the transformation equations.

We will also try to better explain the choice of these values in the central part of the log-sinh section.

"(...) where α (originally a/b) and β (originally $1/b$) are lower and upper reference outputs, respectively. α controls how the error increases for low flows. For

C3336

outputs larger than β , instead, the scaling of the error (derivative of g) is approximately equal to unity. In our study, we chose α to be a runoff in the range of the smallest measured flow and β to be an intermediately high discharge above which uncertainty was assumed not to significantly increase. These considerations are also in agreement with the transformation parameter values determined by Wang et al. (2012). Given the characteristics of our catchment and model we set $\alpha=5$ l/s and $\beta=100$ l/s. The graphs of the transformation function and its derivative with these parameter values are provided in Figs. 1 and S1.”

- 5) *Relating to the previous comment, a potential disadvantage of the transformation is that it applies to the sum of the two terms (bias and output error). Is that correct? An approach that allows one to separately treat heteroscedasticity in these two terms seems preferable. For example, output error parameters could then be estimated a priori, independent of any heteroscedasticity in the bias term (which has a different source).*

It is correct that the transformation acts on Y_o and y_M and subsequently the bias B_M and the uncorrelated output observation errors E are in the same transformed space. We agree that it would be conceptually more appealing to have different degrees of heteroskedasticity for B_M and E . Indeed, this is what we considered with the two input-dependent transformed error models. In these cases, the bias variance in the backtransformed space varies with input and output, whereas the scatter of the random observation errors just depends on the output. As we discussed in the paper, however, this increased complexity of the error model does not necessarily improves its performances.

Regarding estimating the parameter of E , it is not clear to us how and why this term should be omitted from the updating process in which all other parameters are inferred. Please notice that E cannot be estimated alone since only the sum $\tilde{y}_M + B_M + E$ is observed. Furthermore E is not exactly equal to the error of the

C3337

measurement device but it is a stochastic process representing the independent part of the output observation errors. For that reason its a priori estimation, even assuming the model to be inadequacy-free, is not a trivial task. It is possible, however, by analyzing output data, to estimate the white measurement noise and set a strong prior on E 's parameter.

- 6) *section 2.2.5: beyond measures such as reliability and sharpness, the entire predictive distribution can be checked by constructing predictive quantile-quantile plots (comparing observations to probabilistic model predictions)*

We agree that PQQ plots could be an additional way of analyzing the predictive capabilities of the error descriptions. We think, however, that care should be applied when interpreting these plots due to the presence of epistemic uncertainty expressed by parameter distributions. Actually, due to the presence of parametric uncertainty, we cannot expect the output observations to exactly correspond to realizations of the distribution of $\tilde{y}_M + B_M + E$. In fact, as explained on p. 5145, line 10, our lack of knowledge about the “true” parameters will make the predictive distributions of the modeled observations wider than the actual distribution of output data, unless one of the uncertainty sources was underestimated. Due to these open issues, we plan to investigate these diagnostics in future studies.

- 7) *“The frequentist component of the residuals (the estimate of the observation errors) is virtually normally distributed, has an almost constant variance and no autocorrelation.” This is not that easy to deduce from figure 6. The diagnostic plots are better for that; these are now in supplement (figs S1 and S2) and should be included in the main body of the paper. Based on fig S1 I don't necessarily agree that the residuals are “virtually normally distributed”. This statement should be qualified.*

C3338

This is a valuable observation and we recognize that the sentence about the distribution of \mathbb{E} was imprecise. Actually, no error model was capable of fulfilling the normality assumption. However, as it is observable in the updated quantile-quantile plot with p-values from the Shapiro-Wilks test for normality, the selected error model was among the most likely to produce normally distributed observation error residuals. Rigorously speaking, however, we cannot test which distribution is more normal. All that considering, we will modify our statement to:

“The frequentist component of the residuals (the estimate of the observation errors) has an almost no autocorrelation and relatively low heteroskedasticity.”

As we discussed in the article, in the context of bias description, the residuals are so heavily dominated by model discrepancy that their frequentist part plays a secondary role in the uncertainty estimation. In order not to distract the reader with less relevant information, we would prefer to leave the diagnostic plots in the supporting information.

- 8) *figure S3 (and following): specify in caption what solid and dashed lines represent; I assume prior and posterior, but it is not stated explicitly*

Caption amended as: “Prior (dashed lines) and posterior marginal distributions (gray areas) for (...)”.

Other amendments following informal reviews

1. All plots will be modified after setting slightly more realistic priors of the error model parameters (as displayed in Table S1), having run longer and more stable chains and having improved the graphical characteristics of most of the plots. Please refer to the updated supplementary material and figures.
2. Abstract, from line 25 on rephrased as: “further research will focus on quantifying
C3339

and reducing the causes of bias by improving the model structure and propagating input uncertainty”

3. Several sentences will be reformulated in the Introduction to make it more concise.
4. p. 5132, line 3: “additive bias term” instead of “bias term”
5. Section 2.1.3: equations for the transformation derivatives will be added
6. p. 5137, line 9: sentence added: “these last two phases imply the derivation of credible intervals via uncertainty propagation.”
7. p. 5137, line 24: sentence modified to “On the other hand, the maximum value of κ is in the same order of magnitude of the maximum discharge divided by the corresponding maximum precipitation of a previously monitored storm event.”
8. p. 5138, line 3: sentence added: “The parameters σ_E , $\sigma_{B_{ct}}$, and κ have values and units depending on the transformation”
9. p. 5138, line 15: full Bayes’ theorem written and sentence added: “Since the integral in the denominator is usually not analytically solvable, numerical techniques have to be applied. In this context, Markov Chain Monte Carlo (MCMC) simulations are useful for approximating properties of the posterior distribution based on a sample.”
10. p. 5141, line 25: sentences modified to “In general we expect this percentage to be larger than 95% as our uncertainty bands describe our (lack of) knowledge about future predictions. This combines Bayesian parametric and bias uncertainty with the uncertainty due to the observation error. These three components of predictive intervals are thus systematically more uncertain than the observation error alone.”

11. p. 5145, line 10: paragraph emended: "In general, accounting for model bias produced substantially wider predictions and separated uncertainty intervals in three components. The bias error models also reduced substantially the magnitude of the frequentist part of the residuals and decreased their autocorrelation. The different formulations of model inadequacy display, however, a considerable variability in terms of predictive distributions and behavior of the observation residuals."
12. p. 5146, line 9: "successfully" deleted.
13. p. 5151, line 17: sentence modified: "Third, in Honti et al. (2013) the log-sinh transformation was not implemented."
14. Conclusions shortened: "In this study, we proposed different strategies for obtaining reliable flow predictions and quantifying different error contributions. We adapted a Bayesian description of model discrepancy to urban hydrology, making the bias variance increase during wet weather in five different ways. The most reliable and sharp prediction intervals were obtained with a bias having a constant variance in a log-sinh transformed space. From the experience gained in this modeling study and theoretical considerations, we conclude that: (...)"

References

- Frey, M. P., Stamm, C., Schneider, M. K., and Reichert, P. (2011). Using discharge data to reduce structural deficits in a hydrological model with a bayesian inference approach and the implications for the prediction of critical source areas. *Water Resources Research*, 47.
- Honti, M., Stamm, C., and Reichert, P. (2013). Integrated uncertainty assessment of discharge predictions with a statistical error model. *Water Resources Research*, Accepted.
- Reichert, P. and Mieleitner, J. (2009). Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. *Water Resour. Res.*, 45(10):W10402.

C3341

- Schoups, G. and Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-gaussian errors. *Water Resour. Res.*, 46(10):W10531.
- Sikorska, A., Scheidegger, A., Banasik, K., and Rieckermann, J. (2012). Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models. *Hydrology and Earth System Sciences*, 16(4):1221.
- Sikorska, A. E., Scheidegger, A., Banasik, K., and Rieckermann, J. (2013). Considering rating curve uncertainty in water level predictions. *Hydrology and Earth System Sciences Discussions*, 10(3):2955–2986.
- Wang, Q., Shrestha, D., Robertson, D., and Pokhrel, P. (2012). A log-sinh transformation for data normalization and variance stabilization. *Water Resources Research*, 48(5):W05514.
- Willems, P. (2012). Model uncertainty analysis by variance decomposition. *Physics and Chemistry of the Earth, Parts A/B/C*, 42:21–30.
- Yang, J., Reichert, P., and Abbaspour, K. (2007a). Bayesian uncertainty analysis in distributed hydrologic modeling: A case study in the Thur river basin (Switzerland). *Water Resources Research*, 43(10):W10401.
- Yang, J., Reichert, P., Abbaspour, K. C., and Yang, H. (2007b). Hydrological modelling of the Chaohe basin in China: Statistical model formulation and Bayesian inference. *Journal of Hydrology*, 340(3–4):167–182.

C3342