Hydrol. Earth Syst. Sci. Discuss., 9, C3023–C3027, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C3023/2012/


Hydrology and Earth System Sciences Discussions

# *Interactive comment on* "Informal uncertainty analysis (GLUE) of continuous flow simulation in a hybrid sewer system with infiltration inflow – consistency of containment ratios in calibration and validation?" *by* A. Breinholt et al.

**K. J. Beven (Referee)**

k.beven@lancaster.ac.uk

Received and published: 14 July 2012

This is a study of uncertainty in the predictions of a simplified urban drainage model within a GLUE methodological framework. I think the presentation could be usefully revised to make more of the evident epistemic uncertainties in the modelling process for this study. It would then make a really interesting paper.

The authors currently start off by suggesting they are making a test of GLUE's ability to



provide uncertainty bounds that bracket observations in both calibration and validation but then it gradually becomes evident that there are a variety of epistemic problems with the data set that is being used that mean it might be difficult for any assessment of uncertainty to do so. They also suggest at the end that the limitations of the model used might also be important in requiring effective parameters outside the prior specified ranges. And yet, the stage 2 of their proposed methodology is to carefully check and leave out any periods of dubious data. It would seem from the author's comments that there are still significant limitations of the data that have been left.

It is, of course, this type of non-ideal simulation problem that GLUE is an attempt to deal with, but as noted in the comments below, one of its advantages is in making clear where there are clear differences (or non-stationary bias) between the model predictions and observations. This is a clear indication that something (model or data) needs to be improved. That seems to be the real lesson of this study, rather than a real test of GLUE as a methodology (interestingly there is rather consistent coverage of the observations between different periods for different behavioural thresholds – does this suggest that some form of non-stationary error correction might be worth investigating?).

More detailed comments

P5 add paragraph split

P6 1950s

P10 L5 is this residual error variance assuming a zero mean bias (as normally used in NSE, but worth saying explicitly)?

Equations 1 are not themselves likelihoods – should use proportionality signs not equals signs

Worth noting that Equation 2 is effectively a Bayesian updating of likelihoods – but again should be proportionality not equals

P11 *This could be necessary if the dotty plots show high likelihood values at the lower or upper end of any of the prior parameter ranges*

- but this is quite common, and ranges might be limited by physical considerations not just a fall in likelihood???

*We instead took a statistical approach to the acceptability criterion requiring a given prediction interval to bracket the proportion of the observations consistent with the chosen interval*

- should this really be described as statistical when your weights are not based on a formal statistical model and you are leaving error series implicit?

- Should also mention here that this "third" option has been used in the past – e.g. by Xiong and Connor paper cited. Might be better to say that there have been three methods used in the past (actually 4 because of the pre-defined limits of acceptability approach suggested in Beven manifesto paper for which there have now been a number of applications, and some previous applications can be interpreted in this way)

P17 *attributed to the inability of the GLUE methodology to fully describe the uncertainty of the system.*

- but one of the advantages of the GLUE method (relative to formal statistical methods) is to detect failures due to either model or data limitations. It is quite clear from some of the plots that – for whatever reason – the model cannot predict the observations but in a way that could also not really be reproduced by a stationary statistical error model (even allowing for heteroscedasticity). So this should not really be described an inability but it is rather informative about something that needs improving in the modelling process. You do, after all, discuss these sources of additional error later in the paper.

P20 *It is important to recognize that the GLUE methodology as applied here and in many other GLUE studies implies a transfer of all uncertainties to the model parame-*



*ters.*

- No!!!! You have stated in the introduction that you wanted to test the implicit error handling in GLUE, but you are now forgetting that each parameter set carries along with it an implicit(non-stationary) error series. Models that underpredict in calibration are expected to underpredict in similar circumstances in prediction etc. The uncertainties are NOT being transferred to the model parameters, but the error series are (implicitly) weighted along with the simulated outputs from that model. This does not, of course, ensure that the errors reproduced in this way will be similar in the predicted period – especially when the sources of error are epistemic as discussed here, but please do not perpetuate this misinterpretation (also in Conclusions P22 L22).

*This means e.g. that insufficient rain input will be compensated for by adjusting the size of the paved area, which adds a level of variation in addition to that caused by parameter correlation (see Table 7), and the posterior parameter ranges therefore lack physical interpretation and thus cannot be 25 used for e.g. inference about the relative size of infiltration area versus size of paved area, which otherwise would be desired knowledge.*

- but this has nothing to do with GLUE implying a transfer of uncertainties to model parameters. Such compensations will be apparent in any calibration exercise (they may even be worse in a formal statistical calibration because of the stretching of the likelihood surface) UNLESS you build in prior knowledge about what is acceptable or not acceptable for parameters and their interactions.

P21 *where the contributing runoff area and pipe network data can be estimated independently)*

- but you have already noted that this may not be possible because of lack of knowledge about what is actually connected to the network – another source of epistemic errors in addition to the nonstationarity in the input errors (and perhaps observed flows)

P23. *however 5 we call for further comparisons between formal and informal approaches in which both calibration and validation periods are included for performance comparison in real-world applications, and we suggest that users of formal approaches demonstrate that their error assumptions are valid.*

- surely you cannot really call for this when you have not made the effort to do so yourselves.

- Is it not better to say simply here that the evidence from the changing nature of the errors in this study between and within periods suggests that it might be very difficult to find a valid error model for use in a formal likelihood approach, and that we should therefore try to learn from the significant discrepancies between model and observations.

Keith Beven

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 8579, 2012.