Hydrol. Earth Syst. Sci. Discuss., 10, C782–C787, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C782/2013/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Approximate Bayesian Computation in hydrologic modeling: equifinality of formal and informal approaches" *by* M. Sadegh and J. A. Vrugt

K. J. Beven (Referee)

k.beven@lancaster.ac.uk

Received and published: 14 April 2013

This is a useful demonstration of the equivalence of GLUE and ABC results, and of the utility of PMC sampling as a way of increasing the efficiency of sampling an ensemble of behavioural models. But it leaves a number of questions unaddressed.

1. In what sense is ABC more generic than GLUE. In fact GLUE, even as set out in 1992 has a much wider range of options (and has been used since with formal likelihood functions). So why is ABC not considered as a special case of GLUE rather than the other way round (ABC is not actually even formally Bayesian in that it makes no use of Bayes equation)?

2. After all the past criticism of GLUE "lumping all uncertainty into parameter distributions", why is the implicit/explicit error treatment issue totally ignored here, to the extent that formal likelihood results are presented (wrongly) only in terms of the posterior parameter predictions.

3. The authors do not clearly separate the two issues of defining criteria for choosing behavioural models and sampling the resulting model space. As noted below, efficient sampling methods help, but it is the choice of criteria that will control how complex the space is to be searched. It is also possible that the convergence of more efficient sampling methods will fail to identify local areas of behavioural models even given the random steps of MCMC type methods. I agree that efficiency is an issue – but the choice of criteria is much more important.

4. The results reveal that the calibration/validation process is subject to epistemic errors (as discussed for these same data sets by the Beven, 2009 comment on Vrugt et al. 2008). The method of estimating a reasonable range of acceptability used here, reveals something about the errors at least in calibration. But these are not then used in prediction (as formal error should be for the full Bayes approach), the results presented are based only on the posterior parameter distributions. Why not? Surely this is important in deciding whether the resulting ensembles should be considered fit for purpose or not (the authors make no comment as to whether bracketing 68% of the observations is fit for purpose – is it not indicating something about errors in either model or data?).

I conclude that the paper needs major revision in terms of both the presentation and discussion of the results.

Some specific comments

4740/9 Abstract. In this paper we introduce an alternative framework, called Approximate 10 Bayesian Computation (ABC) that summarizes the differing viewpoints of formal and informal Bayesian approaches. – what does this sentence mean? ABC does no such thing. Its only claim to resolve the different viewpoint is that for certain toy problems it can be shown to converge asymptotically (but not necessarily quickly) to a formal posterior.

4749/15 The use of such "insufficient statistic"promotes equifinality, and makes it unnecessarily difficult to find the preferred parameter values. - of course, but ABC gives equivalence to all all samples within the threshold of acceptability so does not eliminate equifinality. Indeed, given the types of epistemic error you demonstrate later you would not want to, since otherwise you might be overconditioning based on a particular realization of epistemic error in your calibration data.

4749/24 The premise behind ABC is that _0 should be a sample from the posterior distribution as long as the distance between the observed and simulated data, . . . is less than some small value. – and how is this different from GLUE then?

4751/16 For illustrative purposes we start with the mean of the actual data, - but why not use NSE to make similarity more obvious (and reduce the impact of using such an inefficient statistic)?

4752/13 search. Our sampler therefore adaptively determines the next value of _j ; j > 1 from the cumulative distribution function of the _(Âů) values of the N most recent accepted samples – this might be fine for simple surfaces but would appear to exacerbate the danger of not sampling areas of behavioural models in more complex spaces?

4754/8 The distance function specified in Eq. (5) has many elements in common with the triangular, trapezoidal or beta fuzzy-membership functions used in the limits of acceptability approach of GLUE – ???? surely has much more in common with "classic" GLUE thresholding of informal measures, especially since ABC as applied here uses no such weighting function

4754/19 Latin Hypercube sampling strategy used in GLUE to find behavioral solutions.

– err. . .LHS has been used in GLUE but not that commonly – and the original 1992 GLUE paper used a nearest neighbor MCMC-type sampler so it is not limited to either uniform or LHS sampling.

4755/19 The adaptive updating strategy of _ in PMC not only guarantees a more efficient search strategy than ABC-REJ (GLUE), but also automatically determines the maximum attainable coverage of the discharge observations within the limits of acceptability. – Here you should differentiate between search strategy and defining behavioural simulations. The "true" ensemble of behavioural simulations is not dependent on search strategy – efficiency helps but might also not find the complete sample if there are multiple local areas of behavioural simulations as identified by multiple criteria (also 4758/28 ff).

4756/7 The simulations nicely track the observed data with uncertainty intervals that appear relatively narrow and encompass about 90% of the data. – If I have understood correctly you are plotting only the ensemble of behavioural models in this plot. So you are saying that the implicit handling of model errors in the original GLUE formulation works (at least for this toy example – this was also demonstrated for the Mantovan and Todini toy example in Beven et al., 2008). But surely you cannot just report this, after all the past argument about the "subjectivity" of using an implicit error model (or as some people put it lumping all the error into the parameter distributions) without at least some comment?????

4758/10 This provides further support for our claim that the limits of acceptability approach of GLUE can be interpreted as a special case of formal Bayes. – No, surely not – up to now you have shown that ABC produces similar results to GLUE – it is in fact a special case of GLUE since GLUE is more general than the ABC approach you describe.

4760/11 The 95% uncertainty ranges derived with both methods encompass about 70% of the discharge observations. This coverage is significantly larger than the approximately 12–17% derived from a classical likelihood function. (also 4761/19) - Whoa!! You are comparing different things here. ABC/GLUE are using an implicit error model, formal Bayes has an explicit model that should be included in the outputs. This would normally cover more than the 70% for ABC/GLUE. You would surely use this if, for example, you were interested in flood forecasting – e.g. Romanowicz et al WRR 2008). Coverage of the parameter uncertainty is a totally inappropriate measure for the formal likelihood.

4761/5 Because of sampling inefficiency the GLUE calculations were terminated after 100 behavioral samples were identified – err, why? It surely does not take much to continue to run while preparing the paper. . .. Again efficiency is helpful, but it is not generally that much of a problem for this type of model.

4761/26 – but this issue – and the limitations of rainfall correction have already been discussed in my SERRA 2009 comment on your 2008 paper. It is really a bit naughty not to mention that.

4762/2 This is simply the effect of an increased rainfall intensity during the evaluation period. – how do you know that? Why could it not be some other sort of epistemic error, a consistent increase over such a period would be hydrologically rather strange would it not? Certainly not simply!! The issues of such non-ideal cases are discussed in Beven, 2006, 2010, 2012 and Beven and Smith HESS 2011 – and are even more important for formal likelihoods. Should surely be part of the discussion.

4764/14 The effective observation error remedies this problem, but the magnitude of this value is typically much larger than the theoretical value of _ to guarantee converge to the true posterior parameter distribution.

No, this is totally the wrong argument. There can be no true parameter distribution for this type of non-ideal problem, only for toy problems (and if you believe you have a toy problem then why not use a formal likelihood approach). The limits you are defining are related to the all the observational uncertainties you mentioned earlier (and you should

really be also wanting to guard against future unexpected uncertainties in prediction such as the increased rainfall intensity mentioned earlier).

4764/18 more generic ABC approach – No again! It is surely GLUE that is more generic in its possibilities (including using a formal likelihood as an option where the modeller is prepared to make strong assumptions about the error structure – as demonstrated before ABC started to be more widely used in Romanowicz et al. 1994!!)

Keith Beven

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 4739, 2013.