Our responses are in Times New Roman following the individual comments in `courier`.

Interactive comment by Scheidegger et al., May 27, 2013

We think the comparison of ABC and GLUE is interesting, because on first glance both methods might appear to be similar. However, this is a topic with many subtleties and therefore requires a careful treatment with great attention to detail. Here, we would only like to comment on the use of ABC, because we got the impression that was not applied as it was supposed to.

Response: We would like to thank the reviewer(s) for their comment on our paper. We agree that the application of ABC involves many subtleties. Some of the subtleties have been addressed in our earlier work published in Water Resources Research (Vrugt and Sadegh, 2013). This paper addresses the main comment of the referee(s). The ABC approach is not limited to stochastic models, but can easily be used with a deterministic model. Our next response will provide a more detailed reply to this matter.

It is probably a misunderstanding that all the cited "likelihood-free" approaches have been developed "for cases when an explicit likelihood function cannot be justified". Instead, they were developed for situation where the likelihood is intractable, too expensive to be evaluated, or an explicit formulation is not available. In such cases, the numerical technique ABC offers a possible solution: Instead of evaluating the likelihood function, we only have to be able to sample from the likelihood function. Thus, bypassing the evaluation of the likelihood function widens the class of models for which statistical inference can be performed. Nevertheless, we must be willing to make assumptions about the distribution of the errors, i.e. “the data generating process” must be known (see e.g. the first paragraph of Diggle and Gratton, 1984; third paragraph of Marjoram et al, 2003; Sisson et al., 2007). In consequence, ABC requires a stochastic model (see the cited ABC literature).

In our view, this important fact has been overlooked in this paper, because the presented algorithms, apparently do not generate a random sample (e.g. the deterministic model output +

random error). Instead, only the output of the deterministic model $H(\theta|j)$ is computed and compared to the observations (fourth line of Algorithm 1, fifth and 18th line of Algorithm 2). This is not valid for ABC, and in contrast to those algorithms cited in the ABC literature.

Response: We highly appreciate this comment of the reviewer, yet rest assured our implementation of ABC is statistically correct. If indeed, the model is deterministic and the simulation not perturbed with an error, then the ABC derived posterior would collapse to a single point in the limit of epsilon going to zero, and hence the results would be inconsistent. To resolve this problem, two approaches can be implemented, also referred to as “noisy ABC”. The first one is to perturb each of the model simulated values with a prescribed measurement data error, whose properties are similar to those of the measurement data error in a traditional likelihood function. The consequence of this perturbation is that the model simulation is no longer deterministic but stochastic. A second approach is not to perturb the model simulated values, but rather to directly corrupt the corresponding summary metrics themselves. For a simple linear (deterministic) model, we have shown in Vrugt and Sadegh (2013) how the first approach results in a stable posterior distribution that does not continue to shrink with decreasing epsilon (Figure 4), and whose posterior moments are in agreement with those derived from a residual-based likelihood function (Figure 3). In the present paper, we elude to these results. Yet, we understand that this might not have been sufficiently clear.

We deliberately did not include a stochastic error term in our presentation of the ABC sampling algorithms as we assume that the stochastic perturbation is either part of the model itself or represented in the summary statistics. We obviously can make this more obvious in the revised manuscript.

We do like to highlight however, that the effect of perturbation is rather minimal in the present case. The reason for this is that the value of epsilon used herein is much larger than deemed appropriate within the ABC framework (see discussion section). To be consistent with GLUE we assign values of epsilon that are derived from the limits of acceptability, but are a manifold of what is theoretically acceptable. The dispersion of the posterior is thus so large (as evident in the size of the simulation intervals), that additional perturbation will have little effect. Numerical simulations prior to submission have demonstrated this in more detail, and hence we deliberately left out the stochastic component in the present paper. Simply because the measurement error is small compared to the limits used by GLUE.

In summary, we think that it should be clearly stated and discussed that ABC does not free the modeler of making explicit distributional assumptions about the errors. This is a fundamental difference to GLUE. In our view, such a comparison should rather highlight the theoretical and numerical differences between statistical and informal approaches instead of "proofing" equivalence of (modified) algorithms.

Response: We appreciate this comment of the reviewer. In our revision we will address this issue and highlight the formal need of an explicit stochastic perturbation of the data within the ABC framework. But this is not going to affect our conclusions, simply because the size of the measurement error is much smaller than the size of epsilon used in the GLUE limits of acceptability approach.

Minor points that you might want to consider

First sentence of Section 2: The classical Bayesian approach does not only consider model parameter uncertainty but also uncertainty represented by the error model, for example due to measurement uncertainty. The likelihood function describes the "remaining stochasticity" for given parameter values.

Response: Agree.

Line 3, page 4748: The normalization constant is required to analytically calculate the mean, variance, etc. However, samples from the posterior can be obtained without it.

Response: Yes.

Tables 4–6: It is surprising that the coverage of the prediction intervals obtained from Bayesian inference with likelihood evaluation are so overconfident and unreliable while the results with ABC are much better. One would expect, that ABC gives approximately the same results as Bayesian inference.

Response: We actually are not surprised at all. Application of classical Bayes in hydrologic modeling often results in very narrow prediction (simulation) intervals. The likelihood function is so peaked that the posterior distribution exhibits insufficient coverage. This is an artifact of input data errors (errors in the precipitation data), an issue that is difficult to resolve within the context of a formal likelihood function. Schoups and Vrugt (2010) have introduced a generalized likelihood function that better treats individual error sources, yet this function uses an AR-model of the error residuals to correct for model bias and residual autocorrelation. This is not ideal in simulation, nor is the multiplier approach of Kavetski et al. (2006) and Vrugt et al. (2008). The ABC results exhibit a much better coverage because the limits of acceptability used throughout the paper circumvent convergence to an unrealistically small posterior distribution. This is actually the underlying premise of GLUE, and why we should adopt a limit of acceptability approach rather than conventional Bayesian paradigm.

Throughout the paper: Diggle and Gratton (1984) proposed a “likelihood-free” approach for frequentist maximum likelihood estimation, so it is not a Bayesian approach. Therefore, strictly speaking, it cannot be classified as ABC.

Response: OK. We will re-assess our wording. The point that we were trying to make is that likelihood-free inference and ABC have similar underlying ideas, that is, to use a set of summary metrics rather than a formal likelihood function to quantify the distance between the model and corresponding data.

In summary, we greatly appreciate the comments of the reviewer, and will use those to our advantage when preparing our revised manuscript for publication in HESS.