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

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

Anonymous Referee #2

Received and published: 7 July 2013

Summary: This paper describes and compares ABC and GLUE next to each other. The paper is well written and I recommend publication if the authors thoroughly address the following comments.

1) Please do not claim that ABC is "introduced" in this paper. Your recently accepted paper in WRR "introduced" ABC for hydrological applications (Vrugt and Sadegh, 2013). It is fine to have some duplication (in fact, even 'needed' to understand the paper by itself), but I would be more explicit about the fact that the ABC method is already introduced. This paper for HESS has the nice feature of connecting ABC with GLUE. Yet, that is only mentioned at the bottom of the abstract and end of introduction. I would suggest to bring the connection with GLUE front and center in the HESS-paper to avoid a vague feeling that this is a repeat of the WRR-paper (which is not the case, I checked). Also maybe add on l.161 that this paper is also a follow up of Vrugt and Sadeg 2013, rather than just Vrugt et al (2008c).

Response: At the time of submission, we anticipated that the present paper would be published earlier than the Vrugt and Sadegh (2013) manuscript that was submitted to Water Resources Research. This turned out to not be the case. We can therefore make the appropriate revisions needed to highlight the main scope of the present paper (ABC and GLUE) in comparison to our other paper (ABC for diagnostic model evaluation). We will revisit the abstract and introduction, and see whether it would be advantageous to bring forward the goal of the present paper.

2) Could you please address the following conceptual issue: The traditional 'calibration' approaches aim at minimizing squared residuals (mean square errors, MSE). This paper instead minimizes differences between means (and standard deviations), i.e. summary statistics. However, the MSE is nothing else but a difference between means, plus some additional terms, including standard deviations (sigma). The additional information in the MSE is a correlation (r) between the obs and obs predictions.

MSE = sigma_obs^2 + sigma_model^2 - 2.r.sigma_obs.sigma_model +(mean(obs)-

mean(model))^2

So, it comes as no surprise that with inclusion of more terms that are 'like' the MSE/likelihood function, the ABC method will become better. And it is also no surprise that the 'DREAM' (better 'formal Bayesian') results (table 4-5-6) yield a more accurate (lower RMSE) result and with less simulation uncertainty. The 'DREAM' -calibration simply included more constraints than the ABC-approach, leaving less wiggle room for the posterior parameter estimates. In short: I think that the comparison of the ABC and DREAM results could perhaps be improved by adding more constraints to the ABCalgorithm, so that is more 'like' the likelihood function used in DREAM: e.g. I think that it would be better to use the first 3 terms instead of (std(Y) - std(Y(\theta)) to do a fair comparison.

Response: We appreciate this comment from the reviewer. The goal of the present paper is to show that ABC and GLUE have many elements in common. This requires the use of a summary metric similar to that of the limits of acceptability approach. We did include the DREAM results as comparison to illustrate what RMSE and simulation uncertainty one would get with a (formal) residual-based likelihood function. It is indeed true that if one uses more and more summary metrics (mean, std. + measure of correlation) that the ABC posterior becomes closer and closer to what one would derive form a standard Bayesian procedure with DREAM. We are currently investigating this in more detail and expect to report our results in due course. But this is not the goal of the present paper. The present paper simply shows that GLUE (limits of acceptability) and Bayesian approaches (ABC) have more elements in common than the current debate in the literature might suggest.

3) Text around Line 19: something is confusing dimension-wise. If n is not the number of time steps, but really the dimension of an observation vector (multiple obs) at one time step, then n

cannot be used as the dimension of forcings (e.g. precip, ET) at one time step. The index t is used for time. Please clean up. Also: one system has one evolving state, consisting of multiple state variables, so line 19 should read 'x_0 signifies the initial state' (or state variables, not states). Similarly, take out the "(s)"at number 3 and 6 in Fig. 1. Finally, number "7" (observation error, mentioned on line 30) is not in the Figure 1 (also missing in the WRR paper).

Response: We appreciate this comment. The variable "n" denotes the length of a vector and is by no means intended to suggest multiple observations at a single time step. We will carefully assess the text, and resolve the issues the reviewer addresses. This includes Fig. 1.

4) Eq. 5: how about changing the 'nrho'-symbol in a capital 'nDelta'-symbol? 'nrho' is often associated with correlation. (suggestion)

Response: We appreciate this suggestion. The main reason we use "rho" is that this is the variable that is used in statistics papers to denote the distance between the observed and simulated data (or summary metrics). We therefore prefer to leave it as is, although we understand the concern of the reviewer.

5) Case studies: could you please comment what to do if there is bias, rather than only random error, in either data or simulations? Would you simply inflate the epsilon?

Response: Indeed – that is what Beven et al. has recommended within the context of the limits of acceptability.

6) Line 489: I like this bridging idea and would like to see it more stressed in the paper, but please correct the typo (cap – gap).

Response: Thanks. We will do in revision.

7) Last sentence and the use of 'DREAM' throughout the paper: it is confusing to think of ABC using simulations with DREAM (after presenting these methods apart in this paper).

Response: We appreciate this comment. The context of this remark is that standard Rejection and Population Monte Carlo samplers receive a very low acceptance rate. Moreover, it is very difficult to handle many different parameters with these methods and still find acceptable solutions (within epsilon of measured data/statistics). We have therefore developed a MCMC simulation procedure with DREAM that uses a continuous acceptance/rejection kernel rather

than a discrete one (epsilon) to derive the posterior parameter distribution. This significantly enhances search efficiency. We can remove this remark, but just wanted to highlight that the sampling efficiency can be improved with orders of magnitude by using better sampling methods such as $DREAM_{(ABC)}$.

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.