Responses to referee #3' comments on

Inverse modeling of hydrologic parameters using surface flux and runoff observations in the Community Land Model

Y Sun, Z Hou, M Huang, F Tian, and LR Leung

We greatly appreciate the constructive comments and suggestions from the anonymous referee, which help us improve the quality of this paper. Our responses to the comments are provided item-by-item as follows.

General comments

The paper applies an inverse modelling approach based on MCMC for estimation of parameters in a land surface model. Inverse modelling is a well-established discipline within hydrological modelling, and the applied MCMC approach has been previously introduced for parameter estimation in this area. Thus, methodology-wise the paper does not make any new significant contributions to the research area. The paper is not fully to the point and lacks technical soundness for some parts in the presentation:

Response: *Thanks for the comments. We agree with the reviewer in that inverse modeling is a well-established discipline within hydrological modeling and the MCMC approach has been tested in hydrologic models in previous studies. However, our study focuses on applying inversion approaches to a highly complex land surface model, the version 4 of the Community Land Model (CLM4), as the land component of an Earth System model. This model, compared to its earlier versions and other land surface model, is much more computationally demanding and complex in terms of processes represented. This work, and our previous work on applying global sensitivity analysis approach to CLM (Hou et al., 2012; Huang et al., 2013), are the first few attempts to apply uncertainty quantification and inversion techniques to CLM4, to our knowledge.*

In addition, MCMC inversion represents a very broad category of methods; each specific problem corresponds to a particular graphical model and a new setup of parameterization, proposal selection, and sampling and rejection design, and so on. The methods are not directly transferrable to different problems or between models. So even if MCMC methods have been applied to other models, specific applications such as described in our manuscript can still make important scientific contributions by demonstrating vigorous use of the method and insights that can be gained about modeling uncertainties in the research modeling area. The purpose of our study is not to develop new approaches to inversion modeling, but rather sound applications of an established approach to improve understanding of the sources of model uncertainty and guide model development efforts.

1. The authors state that they ‘compare the performance of two different inversion strategies, including deterministic least-square fitting and a stochastic Bayesian inversion approach’ (page 5080, line 26-27). However, the paper only presents results of application of the Bayesian inversion approach.

   Response: *Thanks for pointing it out. We performed the deterministic least-square fitting using PEST to do the inversion with the default parameter values as initials, and found that simulations of heat flux and runoff using the calibrated parameters* *show little improvements. We have expanded the discussion on the PEST inversion results, although this is not the focus of our study. The least-square fitting approach targets a single optimum set of parameter values, and works well for more linear systems with monotonic relationships between unknown parameters and observable variables and weak interactions between the unknowns. For a complicated system, it is very likely that a single optimum set of parameter values does not exist, and a probabilistic description of all possible solutions is more reasonable. MCMC-Bayesian allows exploration of all ranges of possible solutions and avoids convergence to local minima. There is a single objective function in PEST, but in MCMC-Bayesian, we assign proper weights (i.e., likelihood) to all possible solutions of parameter sets. Since our focus is on the MCMC approach, we have de-emphasized the comparison of approaches in the revised manuscript.*

2. In the formulation of the Bayesian approach multiple data types are considered. In addition, the paper discusses multi-objective calibration in the Introduction. However, only one data type is considered for calibration in the three test cases.

   Response: *Thanks for pointing it out. Our formulation (Equation 2) is able to deal with different data types, as the residuals of each data point can be normalized by its own variability (e.g., σ_{ij}). In our study, we have two types of data: heat fluxes and runoff. The LH flux can be measured by flux tower, which is representative of a small area, while runoff is a composite response of a drainage basin, which is a large area. These provide complementary information for model calibration. However, we did not perform integrated inversion because energy flux observations are available only at the flux tower sites (US-ARM and US-MOz), and only runoff data are available at the MOPEX basins, which are the best available observations at each scale, respectively. Therefore, in this study, we focused on evaluating the potential of improving CLM simulations using the best available observations at appropriate scales. However, our study clearly demonstrated the potential of multi-objective calibration, which will be attempted in future studies.*

3. The results seem to be very sensitive to the acceptance probability of the MCMC algorithm. The reasons for this behaviour are not fully explored in the paper, and are partly neglected in the discussion of the results. The behaviour indicates that the inversion method is not very robust.

   Response: *Thanks for the point. We added relevant discussions in the text. In practice, most of the inverse problems are ill posed and involves nonlinearity, many parameters, especially the insignificant parameters, may not be uniquely determined. Those parameters can be reduced through sensitivity analysis or parameter screening. Theoretically, when there are an infinite number of runs, the inverse results should be* *the same, no matter what inversion approach is used, but in practice the results are limited by computing resources and the convergence criteria used. In our study, we found that some posterior distributions were sensitive to small changes in the reference acceptance probability. This issue can be addressed by reducing the parameter dimensionality with only the most identifiable parameter kept in the inversion parameterization. The most significant/identifiable parameters can be easily captured and their posterior distribution conforms more to expectations, while the insignificant parameter may vary widely within the feasible range and appears random due to non-uniqueness issues that are common to inverse problems. Another method to deal with non-uniqueness is to replace the nonlinear numerical CLM simulator with simplified surrogates, but some complexities in the input-output relationships may be lost. In this study, although the inversion results are subject to non-uniqueness (multi-modality), we demonstrated that with reduced parameters set, the calibration converges much faster and yield much more consistent results. This study, together with our previous studies on parameter screening across flux tower sites and MOPEX basins, can provide useful guidance on best practices of calibrating hydrological model parameters in land surface modeling.*

Detailed comments

1. Page 5082, line 18. Are these the parameters that were found to be sensitive in the previous studies by the authors? Were they all sensitive with respect to both flux and runoff ?

   Response: *Our previous studies (Hou et al., 2012; Huang et al., 2013) have analyzed the sensitivity of hydrological parameters to heat flux and runoff. The selected parameters affect heat flux and runoff with different sensitivity levels. Among them, some parameters (e.g., Fdrai, Qdm, and Sy) are of particular importance and are more identifiable from the heat flux and runoff data. Therefore, in this study, we performed inversion on the corresponding subset of parameters to take advantages of the previous parameter screening, which helps to reduce the issues of non-uniqueness.*

2. Page 5083, line 5. Very brief reporting of the PEST application results, which does not make any sound contribution to comparison with the MCMC approach. Did the authors further analyse PEST, e.g. by tuning algorithmic parameters, check of convergence, and using different initial parameter sets?

   Response: *We used the PEST to do the inversion with the defaults as initial values, and found that simulations of heat flux and runoff using the calibrated parameters show only small improvements. We did not show the results since it is not central to* *this paper. There is a set of tuning parameters in the PEST inverse setup, but we adopted the widely used default settings as given in the user guide. We added more discussion about the differences between the deterministic and Bayesian inversions in the text.*

3. Page 5083, line 17. PEST also provides an estimate of parameter and model predictive uncertainty.

   Response: *It is true that PEST can provide an estimate of parameter and model predictive uncertainty. But a probability distribution of the posterior estimates in the Bayesian inversion would be a more complete representation of the prediction uncertainty, and with MCMC-sampling integrated, it also helps avoid convergence to local minima.*

4. Page 5084, line 6. Typically, in hydrological model calibrations residuals are dependent, which should theoretically be included in the likelihood function. What is the impact on the results of the independence assumption?

   Response: *In this study, we adopted the Metropolis-Hasting sampling, which uses individual proposals to initiate new points, which are either rejected or accepted according to the likelihoods and acceptance probability. The inclusion of residual dependence might affect the likelihoods a bit but not much on whether a new sample set is accepted or not, particularly compared to the effect of other tuning parameters such as acceptance probability, which has been explored in the study. Also, an appropriate measure of such dependence is the covariance matrix of the residuals, the training of which is subject to great uncertainty with limited number of realizations, particularly at the early stage.*

5. Page 5084, line 18. Provide a reference to the Metropolis-Hasting sampling method.

   Response: *We added several references.*

6. Page 5085, line 17. This statement seems to contradict the previous discussion of the PEST results.

   Response: *We have modified the statement to avoid confusion.*

7. Page 5086, line 20. The purpose of the Bayesian modelling averaging is not clear. Model results using different sampling parameters are averaged. Theoretically this can be done, but does it make any sense?

Response: *The basic principle of the BMA method is to generate an overall forecast probability distribution function (pdf) by taking a weighted average of the individual posterior pdfs. The weights represent the model performance, or more specifically, the probability that a model will produce the correct forecast. Here, the weights are not determined arbitrarily, but rather based on the posterior probabilities of all the models sharing the same unknowns and observational data in the inverse setup. The results might have bigger spread but are more unbiased.*

8. Page 5087, line 6. This applies for all inversion methods, and if actually new information is included by the additional data. In general, it is better to include new data of another type (or from another location) rather than more data of the same type (from the same location).

   Response: *We agreed with this view, and added some more discussion accordingly.*

9. Page 5088, line 9. This seems to be a very small number of samples retained for estimation of the posterior distributions. And may have a large impact on the results. Was sensitivity studied applied to look on the sampling variability?

   Response: *Note that these are the 'posterior' samples during the burn-in period; it is normal to use the last several hundreds of samples to check convergence of posterior pdfs/statistics. Convergence of the posterior statistics has been reached.*

10. Page 5091, line 21. What does 'reasonable' mean? Compared to using monthly data?

    Response: *Compared to using monthly data, the big fluctuations disappear when using daily runoff. We clarified it in the text.*

11. Page 5092, line 3. Is it the same subset of parameters that are most sensitive to both flux and runoff?

    Response: *Yes, the most sensitive parameters are the same, although the secondary parameters are a little different.*

12. Page 5094, line 9. The reason for analysing the impact of temporal resolution of observation data is not well explained. If you are only interested in output on a monthly time scale (which is the results presented), then there is no reason to use data with a finer temporal resolution. Obviously, if results on a daily time scale are required, daily data should be used for the calibration.

    Response: *We agree with the comment. In this study, we are interested in improving* *both monthly and daily simulations through calibration. Figure 15b shows validation results at the daily scale. The study of temporal resolution is also to evaluate data worth, data redundancy, and their impacts on Bayesian updating results.*

13. Page 5095, line 25. Sentence not clear.

    Response: *The uncertainty of a model can stem from the input/output data (observations), model parameters, and model structures. In this manuscript, an inverse approach is used to calibrate the parameter sets based on a given forcing dataset and model structure. The results show that better estimated parameters can improve the runoff simulations; however, there are still observation points that cannot be fitted perfectly, which indicates the possibility of model structural errors related to the hydrological parameterizations or uncertainties embedded in observations such as related to the gap-filling procedure of flux tower measurements. We added clarifications in the text.*

14. Page 5096-5097, Sections 6.5 and 6.6. New results are presented. Should be moved to Results section.

    Response: *Thanks for the suggestion. The BMA results are not based on the MCMC-Bayesian procedure as explained in the methodology, we think the materials, together with model validation in section 6.6, indicate the possibility for further improvement and reliability. Therefore, we prefer to include them in the discussion section.*

15. Page 5104, Table 1. Something seems to be missing in 'Used equations from'. Explain STD.

    Response: *Thanks for pointing it out. "Used equations from Cosby et al. (1984)." STD means standard deviation. We have revised the text accordingly.*

16. Page 5113+5115, Figs. 9+11. Which acceptance probability is used?

    Response: *Reference acceptance probability is 1.0. This has been clarified in the captions.*

17. Page 5114+5116, Figs. 10+12. The 10-parameter solution could be included for a direct comparison.

    Response: *Figure 4 and Figure 8 are the result with 10 parameters at the US-MOz site and the MOPEX site, respectively.*

References

Hou, Z., Huang, M., Leung, L. R., Lin, G., and Ricciuto, D. M.: Sensitivity of surface flux simulations to hydrologic parameters based on an uncertainty quantification framework applied to the Community Land Model, J. Geophys. Res., 117, D15108, 10.1029/2012jd017521, 2012.

Huang, M, Hou, Z., Leung, L. R., Ke, Y., Liu, Y., Fang, Z., and Sun, Y.: Uncertainty Analysis of Runoff Simulations and Parameter Identifiability in the Community Land Model – Evidence from MOPEX Basins, J. of Hydromet., 10.1175/JHM-D-12-0138.1, 2013.