Replies to Referee #2 Comments on Manuscript bg-2017-41

We would like to express our sincere gratitude to the anonymous reviewer for her/his insightful and constructive comments and suggestions. All comments have been addressed below and corresponding revisions will be made in the revised manuscript.

General Comments:

Comment 1:

This paper describes the testing of the performance of the Differential Evolution Adaptive Metropolis (DREAM) MCMC algorithm versus the Adaptive Metropolis (AM) algorithm in two benchmarking exercises and with the Data Assimilation Linked Ecosystem Carbon (DALEC) model using Harvard Forest flux tower data.

The manuscript is clear and well written, and highlighting the good performance of the DREAM algorithm is of interest to others addressing the issues associated with parameterizing ecosystem models.

Response:

We appreciate the reviewer for the concise and nice summary and positive assessment of the manuscript.

Comment 2:

This is a highly technical manuscript, detailing the implementation of two algorithms, and I note the interactive comments of Vrugt and Laine, both far more qualified than me to assess the technical aspects of this study. Therefore, I'll concentrate on my concern that this manuscript is too technical, or at least focused in the wrong area, for the scope of Biogeosciences.

For publication here, I would suggest some major revisions are required, shifting the focus of the manuscript to make it more relevant to this audience. This would involve: (i) relying more on referencing previous work when discussing the technicalities of the algorithms and their implementation and testing; (ii) bringing in an observing system simulation experiment (OSSE) approach; and (iii) concentrating more model and ecological insights these implementation of DREAM/AM and DALEC provide.

Response:

We thank the reviewer for the great suggestions. The manuscript has been substantially revised to be interested to the general audience of Biogeosciences. In summary:

(i) We have substantially reduced the technical Section 2 and completely removed the functional comparison study in Section 3 by providing related references and discussing previous work.

(ii) The OSSE study with known parameters and pseudo-observations has been added in the revised manuscript. Results indicated that (1) the problem can be well constrained by the NEE data along, and (2) the approximated parameters posterior distributions can enclose their true values very well. For more information about this study, please see our responses to the specific comments below.

(iii) More discussion about the model and ecological insights has been added in the revised manuscript. For example, we added the OSSE study and did not observe the bimodality of the posterior distributions, which provided a good evidence to support the statement that the bimodality was caused in part by model structural uncertainty (i.e., incomplete representation of the senescence process). In addition, we added the residual analysis and investigated the impact of observation error assumptions on the parameter estimation and model performance. More details please see our responses to the specific comments below.

Specific Comments:

Comment 1:

Given the large literature and other information there is already available describing DREAM, and the DE-MC Section 2.4 is overly long, and repetitive of much existing work.

Response:

We thank the reviewer for the suggestion. The Section 2.4 has been substantially reduced in the revised manuscript by providing related references and discussing previous work.

Comment 2:

For the benchmarking exercises described in Sections 3, similar tests have been carried out in the extensive existing literature on both DREAM and AM, and it doesn't seem that further benchmarking like this is relevant to the Biogeosciences audience.

Response:

We thank the reviewer for the suggestion. The Section 3 has been entirely removed in the revised manuscript; instead, we wrote a new section called strategies and capabilities of AM and DREAM in sampling complex problems. In this new section, we briefly summarized previous work of AM and DREAM performance.

Comment 3:

Section 4, the application of the MCMC algorithms to an ecosystem model seems to be more pertinent. Given the nature of the comparison between algorithms, I would perhaps prefer to see an OSSE-type experiment using the model with known parameters to generate pseudo-observations with realistic uncertainties that are then used to try estimate the (known) values, rather than the more standard benchmarks described in Section 3.

Response:

We appreciate the reviewer for the constructive suggestions. A synthetic study with known parameters and pseudo-observations has been added in the revised manuscript. The estimation results based on both AM and DREAM were presented in following Figure 1. As we can see, the estimated parameter PPDFs can enclose their true values very well. Moreover, the bimodality identified in the real-data study has disappeared in this synthetic case, which once again suggests that the bimodality may be caused by the model structural uncertainty. In addition, for the single-modal problems, with proper initialization, AM can produce the similar results as DREAM. We added a new section to analyze and discuss the results, which we believe is interesting to the Biogeosciences audience.

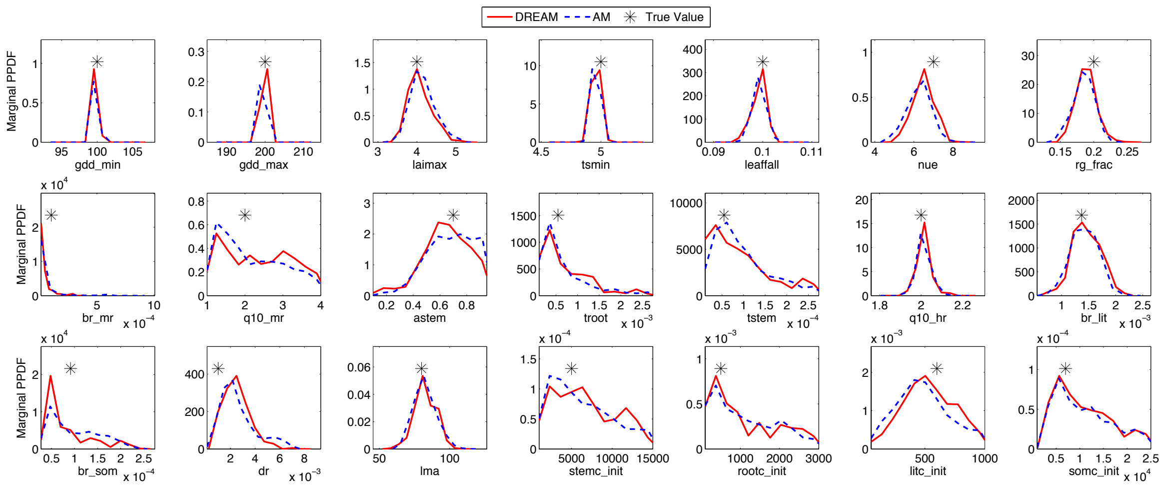

Figure 1. Estimated marginal posterior probability density functions (PPDFs) of the 21 parameters using the AM and DREAM algorithms and their true parameter values in the synthetic case with pseudo data.

Comment 4:

This is in part motivated by being a little surprised NEE alone has allowed all the parameters to be "successfully" determined when using flux tower data from Harvard Forest. This seems to run counter to many (most?) studies that suggest constraining slow turnover rates and a large pool size from NEE data alone is problematic. With such an experiment you might hope to both demonstrate that this result is feasible (in the absence of model structural and initial condition error) and provide a tool to enable a more detailed analysis of why this seems to be case – simply saying the model is simple enough/doesn't have many parameters is insufficient. For example – how important are the data themselves to this conclusion? Is the length of the record and quality of the observations important?

Response:

We thank the reviewer for the excellent suggestions. As it is shown in above Figure 1, the parameters were well constrained by using the NEE for calibration alone in the synthetic study. In the revised manuscript, we analyzed the results in the following: "*These results were opposite from some previous conclusions that eddy-covariance observations along could not identify all the model parameters with their posterior distributions significantly smaller than their priors, as reported in Wang et al. (2007) and Keenan et al. (2012, 2013). Whether a parameter is identifiable depends on the model, model parameters, and the calibration data. When the parameter related processes are necessary to simulate the model outputs whose corresponding observation data are sensitive to the parameters, the parameters can usually be identified and sometimes well constrained. For example, Keenan et al. (2013) showed that in their FöBAAR model with 40 parameters, many parameters couldn't be constrained even with the consideration of several data streams together. They found that these unidentifiable parameters might be redundant in the model structure representation. Roughly speaking, for a simple model with a few number of parameters, the parameters can be more identifiable than the complex models with a large parameter size (Richardson et al., 2010, Weng and Luo, 2011). If the calibration data are sensitive to the parameters, even a complex model can sometimes be well constrained by using a single type of observations. For example, Post et al. (2017) estimated eight CLM parameters using one year records of half-hourly NEE observations at four sites, and found that for most sites the CLM parameters can be well constrained with their 95% confidence intervals close to the maximum a posteriori estimates. For the only site where the parameter uncertainties were relatively large, they concluded that the simulated NEE was less sensitive to these parameters. The DALAC model used in this study is a simple model with considering only six processes and five carbon pools, and all the 21 parameters were shown to be sensitive to the NEE data, despite that some are more sensitive than others (Safta et al, 2015). Therefore, it is not surprising that both AM and DREAM algorithms can constrain the parameters pretty well. In addition, the observation uncertainty of the 14 years NEE data was relatively small, where the standard deviations had values between 0.2 and 2.5 with the mean value of 0.7. The small observation uncertainty and a rather large observation size could be another reason for the well-constrained parameter PPDFs, given that parameter uncertainty is propagation from data uncertainty and inversely proportional to the data size (Hill and Tiedeman, 2007).*"

In addition, we considered another error model in implementing the MCMC simulation. The error model accounted for error correlations. The new calibration results indicated that the parameter uncertainty is larger compared to the uncorrelated error model used in the original manuscript. The reason can be that accounting for error correlation reduces the data information for calibrating parameters. Underestimation of parameter uncertainty using uncorrelated error model was also reported in Ricciuto et al., (2008), Schoups and Vrugt (2010), and Lu et al., (2013).

Comment 5:

Post et al, 2017 JGR-Biogeosciences used DREAM to optimize a set of parameters in the Community Land Model, an ecosystem model massively more complex than DALEC, using flux tower data. Given the similarities, you should draw analogies and make comparisons as appropriate.

Response:

We thank the reviewer for the insightful suggestions and reference. The discussion about the Post et al. (2017) has been added in several places in the revised manuscript. For example, in the Introduction, we added the following sentences: "*Recently, Post et al. (2017) reported a successful application of DREAM in estimation of the complex Community Land Model (CLM) using one-year records of NEE observations. They found that the posterior parameter estimates were superior to their default values in the ability to track and explain the measured NEE data.*" In addition, we discussed Post et al. (2017)'s work when we analyzed the synthetic study results. Please see our response to the above Comment 4.

Comment 6:

Parameter estimation using MCMC techniques remains very challenging for complex ecosystems models such as CLM for many practical reasons, including computational costs. Again, focusing on the readership of Biogeosciences, it would be useful to provide a comparison of the algorithms not just in terms of intrinsic performance given unlimited resource, but also most importantly their efficiency and also their ease of use and set up.

Response:

We thank the reviewer for the insightful suggestions. In several places of the revised manuscript, we compared the two algorithms' efficiency and ease of use. For example, when we described the implementation of the synthetic study, we added the following sentences: "*To facilitate the convergence of AM, we started the chain from the true parameter values and constructed the initial covariance from samples around the true parameter values. This setup can only be done in a synthetic case with information of true parameters available; practically it needs some test runs to get information of underlying distributions. In addition, this initialization of AM makes an unfair comparison with DREAM that launched chains blindly, but on the other hand, it suggests DREAM's ease of use and setup, its robustness and efficiency.*"