

Interactive comment on “Bayesian calibration of terrestrial ecosystem models: A study of advanced Markov chain Monte Carlo methods” by Dan Lu et al.

J. Vrugt (Referee)

jasper@uci.edu

Received and published: 17 March 2017

This paper advocates the use of Bayesian inference to estimate the parameters of the data assimilation linked ecosystem carbon (DALEC) model. The proposed approach builds on the DREAM algorithm and uses a 14-year data record of daily net ecosystem exchange observations collected at the Harvard Forest Environmental Measurement Site. The DREAM parameter distributions are compared against those obtained using another MCMC method, namely the Adaptive Metropolis (AM) sampler. Results demonstrate a superior performance of DREAM with DALEC parameter estimates that outperform their AM derived counterparts during an independent evaluation period.

The paper is generally well-written and discusses an important topic in ecosystem

C1

modeling. The

1. Can you not estimate C_0 from the prior parameter ranges? Just create some samples in this space, as DREAM does, and then compute $C_0 = \text{cov}$ of these samples? Not to say that C_0 is correctly scaled this way. But it makes comparison with DREAM more fair. If you use prior information to construct C_0 then you should also use this for DREAM to sample the initial states of the chains. 2. No surprise that single-site Metropolis does not work well in case of correlated parameters - as correlated dimensions have to be updated together. These arguments have been made in previous DREAM related papers. 3. Page 11: The authors refer to the univariate R statistic to monitor convergence of the sampled chains. Indeed, this approach is often used in multi-chain methods such as DREAM. Nevertheless, I recommend the authors to look into the multivariate R stat of Brooks and Gelman. This statistic does not compare parameters one at a time (their between and within-chain variance) but rather assesses the entire posterior distribution. This multivariate R stat is a single convergence diagnostic and will suggest convergence of the sampled chains at a later time than the univariate R stat of the parameters. The latest DREAM toolbox in MATLAB returns the multivariate R stat. 4. Section 2.4: This section on DE-MC/DREAM has many similarities with published work; for instance, Vrugt (2016). Similar argumentation. I am not sure whether the authors should repeat all this or that a citation to this DREAM manual paper suffices at some places. 5. Case study 1: This study is a standard study that has been used in the DREAM literature. I think the authors should reflect this in their writing. They made some adaptations (50d, variance/covariance matrix of target), nevertheless, this type of study has been published before to illustrate DREAM and AM performance. I think the authors should properly discuss related examples in previous publications. As the authors seem to be very familiar with the DREAM body of work I do not think it is necessary that I provide references here. For example, Laloy and Vrugt (2012) do what the authors present in Figure 3 but then in substantially higher dimensions. 6. Overall the benchmark case studies illustrate performance of DREAM but similar studies have appeared in many other papers - not sure if they are needed

C2

in this work. Reference to those previous studies might suffice. This includes work in different fields, including the present field of application: biogeosciences. 7. Make sure that the math notation in your figures (and labels) matches exactly symbols used in text. This is not the case presently, for instance, Figure 2, $x_1 \rightarrow x$ should be italic. Fig. 4: $R_{\text{statistic}} \rightarrow \hat{R}$ as in text, etc. 8. The paper is technical - the main theme of this paper is a comparison of two different MCMC methods. This comparison is clear and results are fine. Yet, personally I would appreciate a little bit more focus on what we actually learn from using methods such as DREAM. For example a) the authors assume a Gaussian likelihood. We know that such likelihood function is often too simplistic, that is, the assumptions of normality, independence, and constant variance of the residuals can often not be justified. Indeed, a reader might wonder what the impact of these assumptions is on the final parameters and model behavior (behavior during evaluation period) b) The authors do not investigate the residual properties. Do they satisfy the residual assumptions made? For instance, a plot of residuals versus NEE (constant variance justified?), histogram of residuals (Gaussian?) and autocorrelation plot of residuals (no serial correlation?). c) Without an adequate check of the residuals we cannot conclude whether the parameters of DALEC are "correctly" estimated. Maybe a Gaussian likelihood is appropriate for the model and data at hand. I would suspect that a more flexible likelihood function, with nuisance variables, would be more appropriate. This would allow a better representation of the residual properties (tails, skew, nonnormality, heteroscedastic variance, etc.). d) With the use of a more complex likelihood function the bimodality of DALEC parameter t_{min} might disappear. This is interesting by itself. I do suspect though that the performance of the AM algorithm will further deteriorate (in comparison to DREAM) if a likelihood function is used with nuisance variables; for example the generalized likelihood function of Schoups and Vrugt. This is part of the DREAM toolbox (MATLAB) and DREAM Suite (Windows).

Indeed, I think some focus on the choice of likelihood function, and the properties of the residuals would significantly enhance this paper without too much additional work. Otherwise, the paper is merely an important demonstration for the need of robust MCMC

C3

methods in ecosystem modeling; simpler methods might get stuck in local minima. This is an important message for the ecosystem modeling community, yet similar studies/messages have appeared elsewhere, in other journals using different Earth system models.

Note: Figure 8 is very nice. An excellent demonstration of the effect of inadequate inference of AM and consequence of bimodality.

A few editorial suggestions Line 143: ...at similar sites...? Line 144: In the absence of prior information, ... Equation (2) \rightarrow min should not be italicized. Line 180 \rightarrow many studies have demonstrated this - way before Lu et al. (2014). In fact, this is justification why better MCMC methods have been developed in past two decades. Line 189 \rightarrow covariance matrix, C_t , should be bold. It is a matrix of size $d \times d$, where d is number of elements of x , the parameters to be estimated Equation (3) \rightarrow C should be bold, and function Cov as well. Also no need to place s_d in front of e^*I_d , as last term is just for small perturbation to avoid singularity of C_t Line 328: $x_1 \rightarrow x$ should be italic. Please carefully check your math notation. scalars italic, vectors lower case bold, matrices, upper case bold.

Altogether, I would recommend a major revision. Comments should be relatively easy to address - but will require more work (investigate residual assumptions) and DALEC simulations (to test another likelihood function).

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2017-41, 2017.

C4