

Interactive comment on “Bayesian integration of flux tower data into process-based simulator for quantifying uncertainty in simulated output” by Rahul Raj et al.

T. Wutzler (Referee)

twutz@bgc-jena.mpg.de

Received and published: 26 October 2016

The study of Raj et al. presents a successful Bayesian calibration of the biogeochemical model Biome-BGC to Flux-Tower derived Gross Primary Production (GPP). The success of the calibration is shown by several diagnostics and trace plots and by a validation to independent data. Although such Bayesian calibrations of similar models against flux data have been performed before, the aspects of usage of correlated residuals in the cost function and time-varying parameters as well as GPP instead of net ecosystem exchange (NEE) can help for further research. The paper reads well and all the information is given. In order to follow the conclusions some parts are missing, as explained below.

C1

In summary

To my opinion this study presents several aspects that can add to the insight already present by previous studies. But for all of these aspects some more work is required to draw valid conclusions.

1 Major concens

1.1 On conclusion that temporal correlation matters

A control case without the correlation is missing. How do the results and implications change between accounting versus not accounting for correlations?

1.2 On conclusion about time varying parameters

I do not agree with the applied approach. In the presented study several independently simulated time series are mixed together. Each series includes the impact of changed parameters on the previous state. The parameter set valid for July was applied already to April, May, and June and affected the starting states of July. In my opinion one cannot conclude on time-varying parameters with this approach. The simulator needs to be run for the previous months also with the previous parameter set. The model state of the end of the month must be the starting state for the run of the next month with changed parameters. In an ideal case the entire time series would be run as one forward model and the combined (larger) parameter set would be estimated. A more feasible approach is to calibrate each month separately. For the next month calibration continues from a state of the previous month. This starting state needs to be drawn from the distribution of state vectors from the previous month posterior of states for each run with a new parameter sample. For the currently used method, at

C2

minimum, the forward runs that produce the predictive posterior and the fit statistics need to change the parameters across months in each single forward run to discuss seasonally changing parameters.

1.3 On using GPP to calibrate the mechanistic model

Net-ecosystem (NEE)-Flux-partitioned GPP is already the output of another statistical model – here the nonrectangular light response curve. This model already makes some strong process assumption e.g. on relationship of respiration with temperature. In effect the mechanistic model is calibrated against the output of another model. This makes it difficult to interpret the estimated parameters, their distribution and their meaning and process understanding. This needs to be discussed.

Biome-BGC also computes respiration and NEE. You can compare these predictions to observations to gain additional insight into the model and the calibration. The flux partitioning also provides seasonally changing respiration at reference temperature and temperature sensitivity. Comparing these quantities to BIOME-BGC predictions lends further insight, which however, may also reveal sub-optimal calibration.

A more direct way would be to include the respiration parts of the Biome-BGC model in the simulation and fit the simulated, i.e. predicted NEE to the NEE observations. Probably, this will introduce correlations in the joint posterior parameter estimates. But the weaker correlations in the presented GPP fit, are only resolved by the assumptions of the NEE-partitioning model that was used to derive GPP.

While the presented GPP calibration has its own ground, those aspect needs to be addressed. The study would greatly benefit from a comparison to a calibration that uses NEE instead of GPP.

C3

1.4 On hitting the prior bound of residual uncertainty

Fig 1f clearly shows that the calibration tries to increase the residual variance and that high residual variances yields lower cost. In the current inversion, the residual variance is only bounded by the prior. This hints to deficiencies in the inversion. I sometimes experienced the same effect because a single equation of the cost (eq. 5) may in some cases not prefer the best fitting variance but the larger variance together with sub-optimal parameters. Prescribing an upper bound is to my opinion not a good solution for this problem. Even fixing the residual variance would be a better option. My recommendation is to use several parameter blocks in a Metropolis within Gibbs sampling (Chib S Greenberg E (1995) Understanding the Metropolis-Hastings algorithm): One block to fit the model parameter conditional on the parameters of the residual statistical distribution and another block to fit the residual distribution parameters conditional on the current sample of model parameters.

2 Further Concerns

- The cut of the posterior by the edge of the prior distribution of LFRT and FRC:LC (Fig 1) shows inconsistency in the combination of the model, the prior knowledge, and the observations. This hints to deficiencies of the calibration. It also makes it difficult to interpret the parameter estimates and process understanding. This needs more discussion. The introduction of bias parameters in model drivers or model predictions could help to resolve the inconsistencies and, moreover, the bias parameters then can be interpreted.
- How were the initial states of the model prescribed?
- Do you have correlations in the posterior parameter distribution, and how to you interpret them?

C4

- Please discuss your finding in the context of other studies that already performed a Bayesian calibration of BGC-models against Flux data. E.g. there is big body of studies using the DALEC model also looking at multiple constraints, model error, and different sources uncertainties.

3 Technical comments

Fig 1: Shows a very slow mixing. One chain needs more than 1000 steps to become uncorrelated with its previous state. Before computing the Gelman-Rubin criterion you should thin the chains by a factor so that autocorrelation or spectral density of the chain gets small.

Fig 1: shortly explain phi and SD in the figure caption, e.g. “parameters describing variance and correlations of the distribution of model-data residuals (eq. 5)”

Fig 1: Maybe mention, that only the end of the chains after the burnin are shown.

Fig 3, . . . : Legends are missing. Please, use a different line type so that model and observations can be distinguished without color. Readers would benefit if you indicate months at the time axis instead of or in addition to Julian day.

Fig 6 and associated discussion: For a model with state variables or pools this result is trivial. I suggest omitting or explicitly elaborating on the magnitude of the impacts of state versus drivers on the model output and with witch conditions the one or the other becomes important.

P9L25ff: More discussion needed on hitting the upper prior boundaries and its consequences.

P10L18: typo percentile

P13L25ff: I cannot agree to the discussion because of the method that actually did not

C5

alter parameters across seasons during a single simulation run.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-216, 2016.

C6