Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2018-62-AC2, 2018

Biogeosciences
Discussions
Open Access

Interactive comment on "A Bayesian Ensemble Data Assimilation to Constrain Model Parameters and Land Use Carbon Emissions" *by* Sebastian Lienert and Fortunat Joos

Sebastian Lienert and Fortunat Joos

lienert@climate.unibe.ch

Received and published: 6 April 2018

We want to thank the reviewer for the time and effort for the careful and very insightful review. In the following, we respond to the reviewer point by point, with our **responses in bold** and *quotations from the updated manuscript in cursive*. Please also consider the updated manuscript with track changes and a high-resolution figure for this reply in the supplementary. Also note that we expanded the discussion section to include a paragraph on a potential bias in the fossil fuel emissions used for the deconvolution by including non-fuel uses.

Printer-friendly version

Discussion paper

Point-by-point response

The study presents an approach to constrain a DGVM with multiple observational streams of carbon stocks, gross and net fluxes. The authors rely on a latin hypercube stratified sampling to perturb model parameters and create several 1,000-member ensemble simulations of the terrestrial carbon cycle for the historical period. Results focus on the estimation of land-use and land-cover change emissions. This study is quite innovative in the context of the global terrestrial carbon cycle as model parameters are constrained globally.

Thank you

I have found several similarities between the method described here and the Generalised Likelihood Uncertainty Estimation C1 method used in hydrological sciences (Beven and Binley 1992).

We have added a reference to Beven and Binley in the introduction:
"Other approaches have also been investigated, such as using generalized likelihood function for model calibration and uncertainty estimation (Beven and Binley, 1992)"

First, my main criticism targets the description of the sampling method. It is very unclear how the prior probability distribution in Figure 1 and the new best-guess values in Table 1 have been obtained, and how the posterior distribution of the parameters is calculated. Is it based on the selection criterion used to exclude the less skilled model parameters (p7 l5-8)? If Figure 1 and Table 1 present results from the current manuscript they should be described in the corresponding section.

We restructured the method section and introduced a new subsection describing the explorative approach used to obtain the prior distribution. Additionally, we clarified the procedure to arrive at the posterior distribution in Section 2.3.2 (p.6 l.9-p.8 l.7 in the manuscript with track changes).

Figure 1 in this reply shows the evolution of the median parameter values and

ranges of the ensembles with 200 and 300 members (T1-T6) and large ensembles with 1000 members (E1-E3), discussed in the new section in the manuscript. Only the parameters used in the final ensemble E3 are shown. In the small ensembles, different model configuration and parameters were tested. For instance, in ensemble T2 nitrogen limitation was not considered and thus the nitrogen cycle related parameters were not sampled. The factorial simulation and small ensembles informed the choice for the prior of ensemble E1, which is then iteratively improved to arrive at the prior of E3.

Second, I struggle to understand what experiments were actually undertaken. From section 2.6, it seems that three simulations are performed for each parameter set. These three simulations differ in the representation of LULCC: none (reference), gross or net transitions. Then, the results section reports the three different model configurations Mnet,net, Mgross,net, Mgross,gross while these are first described as three alternative skill weighted median.

To clarify we added the following text to section 2.5 (Former 2.6):
"For each of the parameter sets 4 transient simulations over the industrial period are performed: (i) a simulation with prescribed net transitions ($M_{net,net}$ and $M_{gross,net}$), (ii) a simulation with prescribed gross transitions ($M_{gross,net}$ and $M_{gross,gross}$), (ii) a run with landuse area fixed at preindustrial levels and (iv) a run with landuse including shifting cultivation held at preindustrial levels. The last two simulations are used purely diagnostic to determine E_{LUC}."

Third, I am unclear about the skill-weighted mean method. Simulations with either net or gross land-use configuration are likely to yield different results so it is hard for me to justify Mgross,net. I understand that the Mgross,net skill-weighted mean provides the best results compared to benchmarks (Table 3) but it could be an artefact, couldn't it?

As now explained in the revised MS (see our answer above), we did not perform the procedure for optimizing the prior distribution for $M_{gross,gross}$. The prior and

posterior distributions of this configuration do not converge and as such we feel not confident in using it as the basis for our estimates for E_{LUC}. However, it is clear from literature that processes such as shifting cultivation and wood harvest are an important component of landuse change. As a compromise we use the optimized $\mathbf{M}_{net,net}$ ensemble with the additional gross transition processes added, without the retuning of the model parameters.

Also, some parameter sets are likely to perform better in some regions and worse in other. Therefore, would a spatially-explicit weighting scheme (Schwalm et al., 2015; Exbrayat et al. 2018) be more suited to constrain the ensemble?

The use of spatially dependent parametrization offers numerous advantages, which include the potential to yield better performance with regard to observational data. However we believe that assessing the performance of an individual model using global parametrization, can still provide valuable insight in the terrestrial carbon cycle, as a potential caveat of regional parametrization are the additional degrees of freedom which could potentially lead to an over-fitting of the problem. We have added the following text to the discussion:
"An other avenue of increasing model performance is to introduce spatially explicit parametrization, as recently used in multi-model averaging studies (Exbrayat et al., 2018; Schwalm et al., 2015). A caveat of using this approach with a single model is a potential overfitting of the parameters."

Hereafter are some more specific comments p4 l6: CRU TS3.23 covers 1901-2014, so how are simulations performed for 1800- 2014 (or is it 1800-2016 like in the abstract?) please clarify throughout the manuscript

Simulations are performed from 1800 to 2016 with recycled climate data from 1901-1930. Corrected wrong period 1901-2014 to 1901-2016 and changed the wrong reference from CRU TS3.23 to CRU TS3.25 (1901-2016). The recycling of the climate data is described at the end of section 2.2.

p4 l21: please define what model metrics

Added specification in sentence:
"..,the sampling is independent of the metrics used to assess model performance,.."

p5 l6: how have these distributions been chosen?

Please see update to sampling description.

p7 l1: please write MSEi rel

Done

p9 l28: 'LULUC'? please correct here and in several other places

Done, corrected LULUC to LULCC throughout the text

p10 l13-23: please include some information about the uncertainty displayed in the Figures here and throughout the text

We now report the skill weighted 90% confidence interval throughout the text, except for differences between different ensemble configurations.

p11 l1: please quantify 'slight'

The uptake from 1980 to 2016 amounts to 2.6 PgC. We now report the interval 1990-2016 and revised the sentence to read:
"The resulting total change in land carbon is negative, with a slight uptake of carbon at the end of the century, amounting to 9.3 (-0.9,22.2) PgC between 1990 and 2016"

p14 l11: see previous comment on the study period

The simulation spans 1800-2016, however spatial output was only saved after 1901 due to storage limitations.

p16 l3: an informative figure would a covariance matrix of the parameter sets' scores for each criterion

Figure 2 in this reply shows plots of the skill in individual observational targets for all parameter sets. A striking feature is the high correlation of the skill in total carbon map with skill in soil carbon map, which is not unexpected. There is no scatter plot with a lack of points in the upper right corner, i.e. no observational constraints are mutually exclusive. While we agree that this figure is informative, the sheer size and number of subplots make an inclusion in the manuscript or supplementary difficult. Please note that a version of the figure in vector graphic format is included in the supplementary of this reply.

p16 l12: according to Figure 8b and d, the model captures the seasonality but not the interannual variability. This is worth reporting (and explaining).

The interannual variability is not captured because the transport model used does not feature winds with interannual variability. Added sentence:
"As expected, the interannual variability in seasonal amplitude of CO_2 is not captured as the atmospheric transport model TM2 does not represent interannual variability in mass transport."

Fig 1: Mnet,net is not defined

Added:
"... ensemble with net land-use ($M_{net,net}$)"

Fig 3: please explain the sign convention as it seems at odd with figure 4 (ELUC in particular)

We updated Figure 4 to show a release of carbon to the atmosphere due to LULCC as positive, which is consistent with Figure 3 (And the rest of the text). Updated the figure caption and text to be consistent with this change.

Fig 7: this figure is very complicated. Why is it important to look at the whole ensemble, and the constrained one? Constraining the ensemble uncertainty is not a major point in the rest of the manuscript and uncertainties are not reported in most of the text.

We expanded the section explaining the constraining process (See answer above) and added additional confidence intervals for the numeric results. We revised the text in the first paragraph of 3.2 to better explain the figure:

In Fig. 7 a mapping of the MSE_{rel} *to an individual skill score is displayed for the observational data-sets with a spatial structure, to demonstrate how well the median of the ensemble and the new version LPX v1.4 are able to simulate individual observations. The figure also demonstrates the success of the assimilation process: the skill scores for many individual targets are improved in the ensemble median and LPX v1.4 compared to LPX v1.2, the starting point of our work. As a consequence of our iterative prior selection (section 2.3.2) the median skill for an individual constraint is similar in the constrained ensemble compared to the unconstrained ensemble. In all but the fAPAR benchmark the skill is consistently higher than the minimum skill criterion. With the exception of the biomass measurements by (Keith et al., 2009) and the fAPAR benchmark, the maximum skill in the constrained ensemble is identical to the full ensemble. The reduced maximum skill in those benchmarks is due to an exclusion of singular runs excelling at this benchmark but performing badly in others.*

Fig 8: please move the legend

Done

Please also note the supplement to this comment:
https://www.biogeosciences-discuss.net/bg-2018-62/bg-2018-62-AC2-supplement.zip

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-62, 2018.

Printer-friendly version

Discussion paper

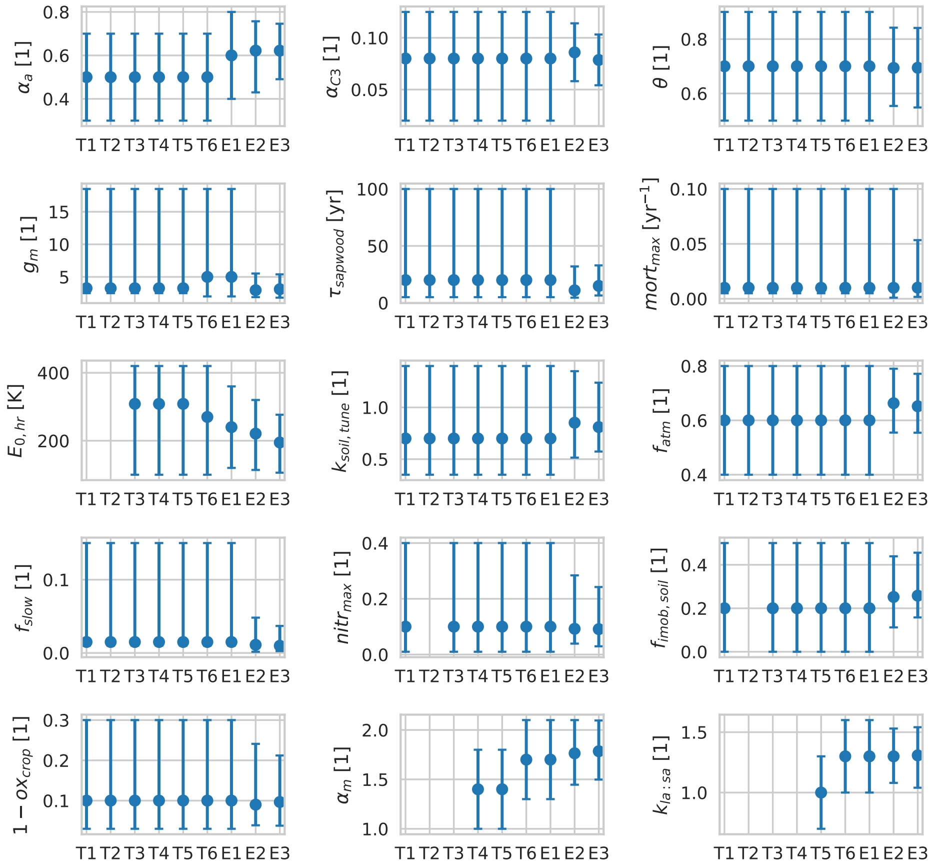

Fig. 1. Median and 90% confidence intervals used for the prior distributions of the parameters of 9 ensembles. T1-T6 are ensembles with fewer members and E1 and E2 were precursors of the final ensemble E3.

Printer-friendly version

Discussion paper

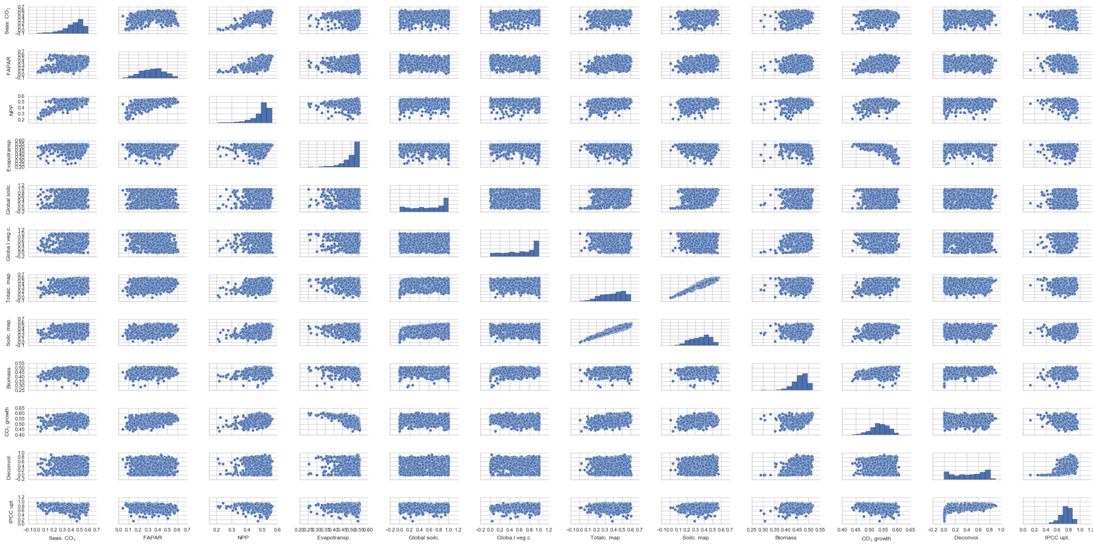

Fig. 2. Skill in observational targets for all parameter sets. The diagonal shows a histogram of the skills for the targets, the off-diagonal shows the skill of two observational datasets in a scatter plot.

Printer-friendly version

Discussion paper