

# ***Interactive comment on “Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation” by Istem Fer et al.***

**I. Fer**

fer.istem@gmail.com

Received and published: 9 April 2018

We are grateful Dr. Van Oijen’s expertise and comments. We thank him for pointing out the relevant literature. While we mainly focused on use of emulators or surrogate models in ecosystem modeling studies, he is right that our citations should widen outside of the literature in ecosystem modeling as we aim to further the use of these techniques in the ecosystem modeling community. We would be happy to include his recommendations and revise the novelty statements in the light of these references.

**The Introduction mentions that “Parameter error refers to the uncertainty about the true values of the model parameters”, which is quite wrong. Parameter error**

Printer-friendly version

Discussion paper



means assigning a value to a parameter which differs from reality, e.g. stating that the light- use efficiency is 1 g MJ<sup>-1</sup> when in reality it is 2 g MJ<sup>-1</sup>. Not knowing whether it is 1 or 2 or anything else is uncertainty. It is therefore also incorrect to state, as the authors do, that "parameter error asymptotically goes to zero with enough data". It is the conditional uncertainty that goes to zero, not the error. Every experimentalist knows that having any number of biased measurements makes no parameter converge to its correct value - and all measurements have their hidden or unhidden biases. There is no safe way to "estimate observation error from data".

We share Dr. Van Oijen's concerns about the consistency of concepts. It is important for us that discussions of these concepts, and methods for their analyses, become more common practice in ecosystem modeling studies. We completely agree with reviewer's definitions of error and uncertainty: "Parameter error is the difference from [its correct] value, and parameter uncertainty is not knowing what that value is" (Van Oijen, 2017). We intended to refer to "uncertainty about parameter errors", thank you for catching it. Therefore, we also agree with that it is the parameter uncertainty that goes to zero. We would be happy to revise these statements to the satisfaction of the reviewer, including additional citations.

**The treatment of the subject matter in the Introduction is further hampered by poor terminology regarding parameters. Terms like "parameter", "parameter vector", "parameter set[s]" are used arbitrarily and inconsistently. [As an exercise for the reader: show that lines 98 and 147 cannot both apply.] Note that a set is unordered and a vector is ordered, so a point in parameter space can not be a "parameter set".**

We thank Dr. Van Oijen for pointing out the usage of parameter vector vs parameter set. We were using "sets of parameters" for a number of parameter combinations that went in for a particular LHC ensemble, as in [nKNOTS x nPARAMS] (being n rows x n cols). For fitting the GP, the order among rows is not important, whereas each row (a

parameter set in that sense) is a(n ordered) vector of parameters (e.g. usage on L189). However, the reviewer is right that their current usage is confusing, if not wrong. We will refer the latter [1 x nPARAMs] as “a parameter vector” and the former [nKNOTS x nPARAMs] as “a parameter set” and will not use “sets of parameters” unless we refer to multiple parameter sets (e.g. multiple parameter sets for multiple iterative emulator rounds). We will go through the text and make sure these are introduced and referred consistently.

### **And how exactly does PEcAn calculate the contributions of different parameters to overall uncertainty, i.e. what was the screening algorithm?**

The uncertainty analysis in PEcAn uses a simple one-at-a-time (OAT) approach. OAT approach involves multiple model runs while holding all parameters at their median except one each time, and evaluating how it translates to differences in model outputs. The parameters are varied at their parameter data assimilation (PDA) analysis priors’ (which could be original priors or if the parameter was constrained by the meta analysis, they could be meta analysis posteriors in PEcAn) median and at six PDA prior quantiles equivalent to  $\pm[1,2,3]\sigma$  in the standard normal. More details are given in previous papers as cited (LeBauer et al., 2013; Dietze et al., 2014).

### **How is the "Euclidean distance between confidence intervals" determined?**

“Euclidean distance between confidence intervals” were determined simply by calculating the mean Euclidean distance between 2.5%- 97.5% CIs of post-emulator and post-bruteforce PDA ensembles at each time point. For example, for half-hourly time step and two years (2005-2006) of flux outputs, we have 35040 points in our model output time series. Then there will be 35040 values of  $(CI_{E,L} - CI_{B,L})^2$  where E stands for emulator, B stands for bruteforce ensemble and L stands for lower CI limit. The same is calculated for upper CI limit and sum of their mean is used as a score for relative confidence interval (RCI) coverage per variable:

$$RCI_{VAR} = \text{mean}((CI_{E,L} - CI_{B,L})^2) + \text{mean}((CI_{E,U} - CI_{B,U})^2)$$

Then the sum over variables (in our case,  $RCI_{FINAL} = RCI_{NEE} + RCI_{LE} + RCI_{SoilResp}$ ) gives us the final RCI score. We expect this score to get smaller as the approximation error of emulator decreases with increasing number of knots and the post-emulator PDA ensemble CIs overlaps more with post-bruteforce CIs. While this metric is not perfect and certain scenarios can even result in misleading scores (e.g. a narrower -than bruteforce- emulator CI coverage can give the same score with an equally wider -than bruteforce- emulator CI coverage according to this calculation, which is an unlikely scenario but just to give an example), in our experience it contains information about how close the CI coverages of post PDA ensembles for both approaches are. We would be happy to include more details and discussion on this in the text and provide more supplemental figures as an example. Please below see a visualization of emulator CI coverage approaching to bruteforce CI coverage with increasing number of knots for NEE (Fig 1). Please note that this is a smoothed time-series for ease of visual inspection. Otherwise, these are from the exact same runs from our scaling experiment with 8 parameters that are used in RCI calculations reported in Figure 7a in the paper.

**Can you explain the results shown in Tables A2 and A5? How can posterior distributions for parameters following MCMC neatly fall into parameterised probability distributions (which also are often of different type than their priors)? And what were the posterior covariances?**

We would like to clarify that the results reported in Tables S2 and S5 are fitted parametric distributions to the MCMC samples. We wanted to provide an approximate parametric distribution for the reader for ease of use. Otherwise, all the raw MCMC samples are accessible via our workflow directories for more interested users/readers. (e.g. from the following url <http://pecan2.bu.edu/pecan/08-finished.php?workflowid=1000008503> -please note that this takes a while to load- the reader can first select an “mcmc.list.pda\*\*\*.Rdata” file under the “PEcAn Files” drop-down menu on the left frame. By clicking “Show File” button they can download the raw

[Printer-friendly version](#)[Discussion paper](#)

MCMC outputs to their own machines for further analyses.) We can include a more detailed explanation in the text and in the supplement, we thank the reviewer for pointing this out.

Correlation density plots were provided in the supplementary but were not discussed in the text. We will list the strongest correlations in the text more explicitly and add discussion accordingly.

Finally, we would be happy to include further discussion on limitations of our approach and address all comments of Dr. Van Oijen in a final response given the chance of revision.

Istem Fer

Dietze, M.C., Serbin, S.P., Davidson, C., Desai, A.R., Feng, X. Kelly, R., Kooper, R., LeBauer, D., Mantooth, J., McHenry, K., and Wang, D.: A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes, *J. Geophys. Res. Biogeosci.*, 119, 286–300, doi:10.1002/2013JG002392, 2014.

LeBauer, D. S., Wang, D., Richter, K. T., Davidson, C. C., and Dietze, M. C.: Facilitating feedbacks between field measurements and ecosystem models, *Ecological Monographs*, 83: 133–154. doi:10.1890/12-0137.1, 2013.

---

Interactive comment on Biogeosciences Discuss., <https://doi.org/10.5194/bg-2018-96>, 2018.

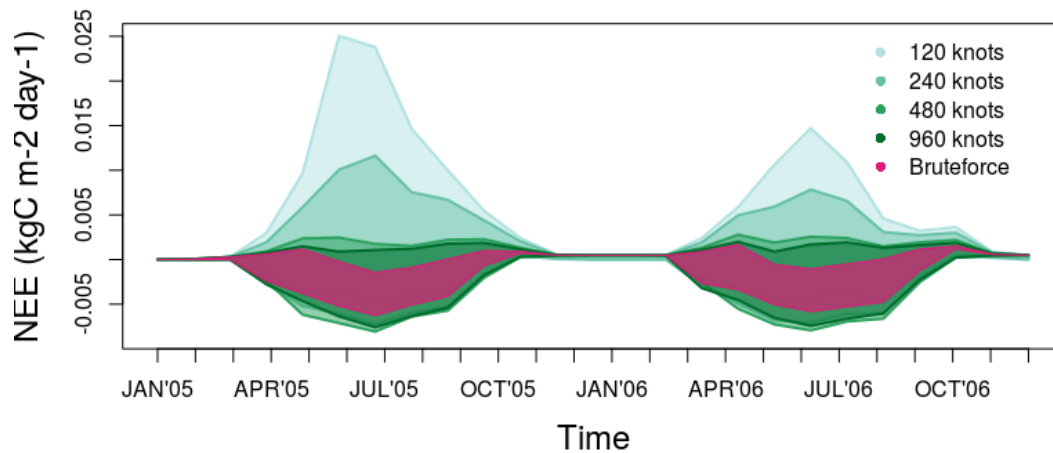
BGD

Interactive  
comment

Printer-friendly version

Discussion paper





**Fig. 1.** Post emulator-PDA ensemble CI converging to bruteforce CI with increasing number of knots

[Printer-friendly version](#)[Discussion paper](#)