Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-160-AC3, 2018

Geoscientific Model Development Discussions

Interactive comment on "Land surface model parameter optimisation using in-situ flux data: comparison of gradient-based versus random search algorithms" *by* Vladislav Bastrikov et al.

Vladislav Bastrikov et al.

vladislav.bastrikov@lsce.ipsl.fr

Received and published: 29 October 2018

Dear Referee,

We would like to thank you for the review and for your interest in this work. Below we provide the answers to your general and specific comments.

Referee: Nevertheless, the introduction could also elaborate on further parameter estimation techniques which also could show a better performance than the ones addressed here, in particular concerning uncertainty characterization (ensemble methods) and handling non-Gaussianity (e.g., particle filter methods or

Markov Chain Monte Carlo methods for example the work by Post et al., 2017, JGR-Biogeosciences).

Answer: Indeed it is possible that other methods may provide a better performance than the ones addressed here. However, these studies have not been performed, so we do not know which method performs the best. It would certainly be a good extension to this study to test other methods, and it would be good for the DA community at large to perform these kinds of method sensitivity tests across a range of model complexities. We have discussed the differences in methods extensively – including MCMC and particle filter methods – in the text (original submission P3 lines 7-125). However, we have added a reference to Post et al. (2017) earlier in the text when we talk about assumptions of Gaussian PDFs (P3 Line 4) and references to Post et al., (2017), Richardson et al. (2010), Pinnington et al. (2016) in this paragraph to highlight, as suggested by the reviewer, that there are other methods out there to consider. We further add these sentences before listing the key questions we are investigating:

"Note that this study does not aim to provide an exhaustive examination of all methods belonging to both classes of inversion algorithms (as previously described), nor do we presume to have chosen the best method belonging to each class. We simply choose to test two methods belonging to each class that have already been used to estimate parameters of the ORCHIDEE model. A further examination of the benefits of all methods would be beneficial to the LSM and DA community, but is outside the scope of this study."

Referee: P2, L12, L13, L15: These error sources (vegetation and soil spatial information versus parameter values in model) are partially the same thing? Please clarify.

Answer: We believe that these error sources relate to different aspects of the model. The vegetation and soil spatial information represents the so-called forcing data sets (like the meteorology); they correspond to global maps derived partly from satellite earth observation missions and contain the necessary information to cluster the different types of soils and vegetation all over the globe. The parameter values correspond to specific and chosen formulations of the different processes controlling the carbon, water and energy budgets in the model. They are thus internal to the model while the forcing data are external and may easily vary depending on the region where the model is applied. It is thus helpful to distinguish these different error sources and to classify them into separate groups.

We have thus only slightly changed the text to be more precise.

Referee: P5, L7: Could you provide more details? How is the sensitivity study conducted? Reference?

Answer: The choice of the model parameters was done based on sensitivity tests of the data used in optimization (net CO2 ecosystem exchange (NEE) and latent heat (LE) fluxes) with respect to all related ORCHIDEE parameters. This was done in previous works, as already cited in the paper and based on the so-called Morris method (Morris, 1991), which ranks the variability of "elementary effects" of the sampled parameters with respect to their impact on the model outputs. This information together with the new reference has been added to the reviewed text.

Referee: P5, L22: Can you provide more details? How many sites were disregarded?

Answer: Based on an original list of 252 sites from the La Thuile dataset (Baldocchi, 2001), we conducted a screening as described in the paper resulting in the selection of the 78 sites used in this study. Thus, we have disregarded around 70% of the sites. The total number of the sites in the original database has been added to the reviewed text. Additional reference to the article devoted to the PFT refinement subject is also added: "Note that Peaucelle et al. (in review) explored with the same model how to account for plant functional trait variability, refining the PFT distribution".

Referee: P6, L31: How automatic was the automatic differentiation? Can you provide more details on the additional coding and time which was required?

Answer: The differentiation has been generated by the TAF automatic differentiation tool from the FastOpt company (see http://www.fastopt.com/). However, the success of automatic differentiation largely depends on the cleanliness of the model code and to a certain extent on the structure of the code. Our group had spent a large effort on cleaning and making the initial ORCHIDEE model code, suitable for the TAF software. Note that a specific document with recommended coding guidelines has been built as the result of this work. Additionally, some input/output peace of code was also inserted to handle the tangent linear variables that are differentiated through TAF. Overall, this work required a strong investment of one software engineer and it took us around two years to have a working tangent linear model.

We thus added in the text one sentence to resume the committed effort: "For ORCHIDEE the corresponding TL model has been derived with the automatic differentiation tool TAF (Transformation of Algorithms in Fortran; see Giering et al., 2005), following code cleaning and structural adjustments (without changing the physics) to allow the use of TAF (a significant effort that took around two years)"

Referee: P8, Eq. 2: How critical is the Gaussian assumption? Non-Gaussianity of parameters can be expected.

Answer: The currently implemented data assimilation technique relies on the assumption that the errors on both the parameters and the observations have Gaussian PDF. In this case, the resolution of the inverse problem, following a Bayesian framework, is equivalent to the minimization of a quadratic cost function. This Gaussian hypothesis significantly simplifies the interpretation of the minimum of the quadratic cost-function (i.e. being the mean of the posterior parameter PDF). If some parameters would have other PDF, the L-BFGS-B minimization procedure would not provide a meaningful value to describe the posterior PDF. However, for the GA this restriction does not hold, as we could use the ensemble of model trajectories to describe the posterior PDF. Additionally, the Gaussian hypothesis is also central to calculate a posterior parameter uncertainty that fully describes (together with the mean value) the shape of the PDF, and this hypothesis also allows to compute the posterior uncertainty with a simple matrix formulation (see for instance Tarantola, 2005). Such hypothesis is used in many inversion problems.

However, we agree that non-Gaussianity may be the case for some parameters and that it could thus partially bias the overall parameter optimization with Gaussian assumption. It is nevertheless out of the scope of this paper to investigate non Gaussian errors. Moreover, in the case of ORCHIDEE we have shown in an earlier study (Santaren et al., 2007) that most parameter errors follow Gaussian distributions. We have inserted these points at the end of section 2.4.1:

"Note that using non-Gaussian errors would significantly complicate the application of one class (gradient-based) and is thus out of the scope of this study and that Santaren et al. (2007) have shown with a previous version of ORCHIDEE that most parameter errors follow Gaussian distributions."

Referee: P14, L30: How do you know whether this number (5) is not case dependent? In case of multiple sites, many tests could be carried out, as the parameters need to be determined just once (per PFT of course). Why would one need to impose restrictions, and would it not be better to use a larger number of initial guesses for cases with parameter estimation for multiple sites?

Answer: We agree with the reviewer that the larger the number of first-guess tests is, the more robust the results will be. However, this has to put in regards to the computing time that is increasing proportionally to the number of first-guesses. Note that the computing time does not vary substantially between the single and the multi-site cases as even for the multi-site case the model has to be run at all sites for each iteration. This chosen number (5) comes from a first order analysis of the results presented in

Figure 3 with our particular model and set of observations. It should be seen as a first order and prior suggestion. For any other model and parameter inversion exercise it could of course be different. Note that this summary point is primarily to stress the fact that using only one first-guess inversion is very risky with a gradient-based method.

Overall, we agree that this statement needs to be put into a more general context and perspective. We have thus changed the text to include the above elements.

Referee:

P1, L26: trapped instead of trap.

P10, L25: degrades instead of degrade

P10, L29: skip "a"

P10, L32: "reductions" instead of "reduction".

P11, L11: "at the same level as" instead of "at the same level than".

P11, L25: "maxima" instead of "maximums". The results are shown in?

P11, L31: "as" instead of "than".

P13, L5: "minima" instead of "minimum".

P13, L28: Change to: "the most constraint ones"

P14, L17: "dependent"

P14, L30: change to: "ensures".

P15, L5: change to "decreased".

Answer: All mentioned typos are corrected in the text. We thank very much the referee for the thorough and attentive reading.

Best regards,

Vladislav Bastrikov

References

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., and Evans, R.: FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, B. Am. Meteorol. Soc., 82, 2415-2434, 2001. Morris, M.D.: Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2), 161-174, 1991.

Giering, R., Kaminski, T., and Slawig, T.: Generating efficient derivative code with TAF, Fut. Gen. Comp. Syst., 21, 1345-1355, doi:10.1016/j.future.2004.11.003, 2005.

Morris, M.D.: Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2), 161-174, 1991.

Peaucelle, M., Bacour, C., Ciais, P., Peylin, P., Vuichard, N., Kuppel, S., Peñuelas, J.: Exploring plant functional traits variability with a terrestrial biosphere model, in review.

Pinnington, E. M., Casella, E., Dance, S. L., Lawless, A. S., Morison, J. I. L., Nichols, N. K., Wilkinson, M. and Quaife, T. L.: Investigating the role of prior and observation error correlations in improving a model forecast of forest carbon balance using Four Dimensional Variational data assimilation. Agricultural and Forest Meteorology, 228-229, 299-314, doi:10.1016/j.agrformet.2016.07.006, 2016.

Post, H., Vrugt, J.A., Fox, A., Vereecken, H. and Hendricks Franssen, H.J.: Estimation of Community Land Model parameters for an improved assessment of net carbon fluxes at European sites. Journal of Geophysical Research: Biogeosciences, 122(3), 661-689, 2017.

Richardson, A.D., Williams, M., Hollinger, D.Y., Moore, D.J., Dail, D.B., Davidson, E.A., Scott, N.A., Evans, R.S., Hughes, H., Lee, J.T. and Rodrigues, C.: Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint

constraints. Oecologia, 164(1), 25-40, 2010.

Santaren, D., Peylin, P., Viovy, N., and Ciais, P.: Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France, Global Biogeochem. Cycles, 21, GB2013, 2007.

Tarantola, A.: Inverse problem theory and methods for model parameter estimation, Siam, 2005.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-160, 2018.