Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-160-AC2, 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,

Thank you for your review and for your interest in this study. Below we provide the answers to your general and specific comments in sequential order.

Referee: There are a couple of key results that could use more discussion (either in the appropriate place in Section 3 or at the end in Section 4). For example, why do certain parameters have different responses to different optimization methods? What are the limitations of multiple-site optimizations? Adding some

discussion could really solidify the take home points of this paper and provide relevance to other land surface models and parameter optimization studies.

Answer: The major factors influencing the parameter estimates are related to the technical implementation of the optimization methods – whereas the gradient-based method mostly looks for the optimal parameter set in the vicinity of the prior parameter values, the random search algorithm may jump to a completely different parameter state in one step. This is the main reason why we obtain significant differences in the estimated parameters between L-BFGS-B (BFGS) and Genetic Algorithm (GA), no matter which case, single-site (SS) or multi-site (MS) is selected. The differences are pronounced for specific parameters such as C_{Topt}, LAI_{max}, SLA, $L_{age,crit}$, $K_{LAIhappy}$, K_{GR} and K_{rsoil} mainly. In the particular case of using pseudo-data, the GA manages to find the true values for these parameters much more precisely than the BFGS algorithm. In most cases, with real data we also see differences between BFGS and GA for these parameters, and we can thus speculate that the GA would provide more optimal posterior estimates in this case as well.

On the other side, if we compare SS vs MS, we do not observe specific patterns in the posterior parameter values, but the range of parameter values obtained with multiple first guesses are significantly lower for MS than for SS. This comes from the fact that for SS cases each site is optimized separately, so we can end up with a parameter value that is highly specific of each site, whereas for MS cases we optimize all the sites together, so the final estimate has less variability for the multi-site optimization. This is illustrated more specifically by the parameters K_{wroot}, C_{Tsen}, Q_{10}, HR_{Ha}, HR_{Hb} and K_{z0}.

Concerning the limitations of the multiple-site optimization, we would like to raise the following points. First as discussed in the paper the benefit of assimilating multiple sites of a given PFT follows from the need to neglect site peculiarities and to find an optimal set of parameters describing the PFT in general. However, the optimization usually does not work efficiently (i.e. does not lead to a large decrease of the cost function) if the different sites have very different behaviors in terms of carbon/water cycle responses to climate forcing. This informs us on the need to reconsider the PFT geographical description (with possible further regional split). This is slightly the case for TropEBF and C3 grass. Additionally, the use of multiple-site optimization requires more computing time and is slightly more complicated to set up with the need to have coherent observation errors between the sites, i.e. with no site that dominates the cost function because of a too low error (measurement and model errors grouped in the **R** term) and thus prevents the optimization to fit all sites together.

Overall, we agree that these two points were not detailed enough in the manuscript and we have thus included the points discussed above in the main text of the paper (in section 3.2.2 – second and fourth paragraphs for the parameters discussion and at the end of the section 3.1.2 for the limitations of the multi-site optimizations).

Referee: Though the focus is parametric uncertainty, I think the paper would also benefit from a brief discussion of model structural uncertainty, with relevant details specific to ORCHIDEE. This addition would provide useful context for discussing the results (e.g., Page 10, Line 25).

Answer: We agree that model structural uncertainties are also a critical part of any data assimilation experiment, but they are rather difficult to assess properly. However, from the existing knowledge on the different processes that control the land surface carbon, water and energy budgets we can list potential missing processes in ORCHIDEE that may have a direct impact on the parameter retrieval. For instance, the version used in this study still lacks a description of the nitrogen cycle and its potential limitation on photosynthesis (in the context of increasing atmospheric CO2), which may bias the retrieval of V_{cmax} parameter. We also do not describe properly forest stand and canopy structure (forest gap, age dependent effects, etc.), which is a limitation on the computation of the absorbed light for photosynthesis and of the turbulent fluxes exchanged with the atmosphere. The main risk is indeed to over-tune some parameters for wrong reasons (i.e., because of missing or incorrect process description) – that

some of the structural error will be aliased onto the model parameters. However, this is not the focus of this study so we do not go into too much detail here. It is described in depth in MacBean et al. (2016).

Overall we do agree a brief discussion of this issue would be useful in the text, so we have slightly extended the model description (section 2.1), to mention the importance of model structural uncertainties and listed potential effects of missing processes on the parameter retrieval.

Referee: I think it would be useful to include some background on choice of model parameters and how their sensitivity was assessed, as this is a key step to narrowing the parameter space.

Answer: The choice of the model parameters was done based on sensitivity tests of the data used in the optimization (net CO2 ecosystem exchange (NEE) and latent heat (LE) fluxes) with respect to carbon and water cycle 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). As nothing new was introduced in this study, we had limited the background description of this choice in section 2.2. However, we agree that more information on the subject is useful, so the following sentences are added in section 2.2 together with the new reference:

"Among all ORCHIDEE parameters we selected the ones that primarily drive net CO2 ecosystem exchange (NEE) and latent heat fluxes (LE) variations on synoptic to seasonal time-scales, excluding those impacting preferentially decadal time scales (i.e., like tree mortality). A preliminary parameter sensitivity analysis was conducted, as in Kuppel et al. (2012), based on the "one-at-a-time" Morris method (Morris, 1991), and we restricted the selection to the 28 most influencing parameters controlling photosynthesis, respiration fluxes, leaf phenology and evapotranspiration."

Referee: Minor note, but it would be preferable to have the line numbers continuously increasing throughout the document so identifying page numbers in the

specific comments is not necessary.

Answer: We agree that using a continuous line numbering would facilitate the referencing, we will follow this advice in the next revision.

Referee: Page 3, Line 2: "probability distribution function"

Answer: The typo is corrected.

Referee: Page 3, Line 6: What are the limitations of the Gaussian assumption? Could some parameters have different PDFs?

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. We have examined the issues that may arise when assuming Gaussian PDFs in MacBean et al. (2016). 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. MacBean et al. (2016) examined the issues that may arise when using Gaussian assumptions in gradient-based minimisation algorithms; however, they found that the algorithm used in this study could account for quasi non-linearity. Moreover, in the case of ORCHIDEE we have shown in an earlier study (Santaren et al., 2007) that most parameter errors follow Gaussian distributions."

Referee: Page 3, Line 19: What defines "excessive" here? Can you give examples of the number of parameters explored in these studies, and how they compare to the dimensionality of your problem?

Answer: In the cited study (Chorin and Morzfeld, 2013), it was shown that the effective problem dimension (defined as the Frobenius norm of the steady state posterior covariance) can remain moderate for realistic models even when the state dimension (i.e. the number of parameter in our case) is large (asymptotically infinite). The precise value of the excessive effective dimension varies from one problem to the other and depends on the level of accuracy required. However, obviously the effective dimension has to remain bounded. In our study the dimensionality of the problem is limited to the few tens of parameters and it can be considered to be small as compared to the cited study (going up to a thousand), which supports the main idea that the numerical data assimilation can be successful.

We slightly modified the text to include the definition of the problem dimension as it is meant in the cited study and changed the phrasing "not excessive" to "finite" for a clearer readability:

"With idealised models, Chorin and Morzfeld (2013) have shown that the assimilation can be optimal with particle filters or variational methods, under the condition that the effective dimension of the problem (defined as the Frobenius norm of the steady state posterior covariance) is finite".

Referee: Page 3, Line 33: L-BFGS_B should be L-BFGS-B?

Answer: The typo is corrected.

Referee: Page 4, Line 9: Word choice "exploited" could be changed to "utilized".

Answer: The word "exploited" is changed to the word "used".

Referee: Page 4, Line 13: Change to "a few".

Answer: The missing word is added.

Referee: Page 4, Line 16: ORCHIDEE should be defined at first mention (Page 3, Line 23 and in abstract).

Answer: The ORCHIDEE transcription as it exists (ORganizing Carbon and Hydrology In Dynamic Ecosystems) is more a play of words, than the real meaningful definition. So, we decide to mention it only in the devoted section (section 2.1) and not in the abstract and earlier mentioning of the model in order not to be misleading.

Referee: Page 4, Line 19: Why "possibly"? Has the use of ORCHIDEE on thousand-year timescales not been proven?

Answer: Indeed, the use of ORCHIDEE model is proven on the long timescales basis. The word "possibly" was only used to show that it is "possible" to run the model on such timescale. The word is now removed for a smoother readability.

Referee: Page 5, Line 5ff: A few lines about the choice of parameter ranges and sensitivity assessment would be useful here (even if in the supplemental to go along with Table S1).

Answer: The ranges of variation for the parameter values have been assigned based on literature analysis and parameter database such as the TRY database

(http://www.try-db.org) as well as following expert knowledge of the model equations. We added this point in the revised manuscript.

For the sensitivity assessment, we agree that additional information is useful as already discussed/provided in the answer to the third reviewer comment.

Referee: Page 5, Line 13: Repetitive here to again mention land use change in parentheses.

Answer: The typo is corrected, the first occurrence is deleted and the one at the end of the sentence is kept.

Referee: Page 5, Line 20: How useful is 1 year of station data? Only one observed seasonal cycle, especially relevant as optimization is on seasonal/annual time scales. Relevant also at Page 10, Line 23.

Answer: Even only one year of data already provides valuable information on the 'main' seasonal cycle and important information on the ecosystem response to synoptic weather events. In general, we thus tried to keep as much data as possible, even if we had only one year of data for a specific site. However, in the case of the multiple sites optimizations (MS), this may lead to some representativeness issues, with long-record sites dominating the cost function and the overall optimization. We have faced this problem and for some sites with a small amount of data the optimization could led to a degradation of the model – data fit in few MS cases. Although using sites of similar record length would be optimal, we believe that keeping short record sites is still crucial to account for the diversity of ecosystem within a given PFT.

We did not revise the text as this point is already mentioned in page 10 (section on multi-site optimization).

Referee: Page 5, Line 27: Please add a sentence explaining why you would expect the Bowen ratio to be constant.

Answer: The correction of the energy balance closure is a difficult task and experts in eddy-covariance flux measurement have not put into evidence that one of the turbulent heat fluxes (latent or sensible) is on average more impacted than the other one. Using a constant Bowen ratio is thus a conservative and natural choice that is applied in most studies (Lasslop et al. 2008, Twine et al. 2000). We slightly modified the text to include this point:

"Where possible, the LE fluxes have been corrected to close the energy balance, using measurements of the ground heat flux (G) and keeping a constant Bowen ratio to update the latent and sensible heat fluxes (i.e., conservative choice without strong evidences that one turbulent flux may be more impacted than the other one; Twine et al., 2000)."

Referee: Page 6, Line 19: Should be Tables S3-S4.

Answer: Corrected.

Referee: Page 6, Line 23: L-BFGS-B acronym should be defined at first mention, in introduction Page 3, Line 30 (and abstract).

Answer: Corrected in the text and added in the abstract.

Referee: Page 7, Line 1: Change "to threshold" to "with threshold".

Answer: Corrected.

Referee: Page 9, Line 13: Would be nice to include the equivalent of Figure 1 for LE flux. Figure S1 has it broken out by PFT but not a summary figure.

Answer: We chose initially not to include the equivalent of Figure 1 for the LE flux as we mainly focus on the carbon fluxes in this paper. However, we now follow the reviewer's suggestion and we have added it into the Appendix as the ending part of the multi-panel in Figure S1.

Referee: Page 10, Line 25: This is a key point – how do model structural uncertainties get in the way of multiple site optimizations? What are the limits to

finding an optimal parameter set across multiple sites? (Also "degrade" should be "degrades"?)

Answer: As already discussed above in the response to the general comments, we agree that model structural uncertainties are crucial in the optimization process but difficult to assess. Multiple sites optimizations are likely to reveal more directly the impact of structural uncertainties as the optimization will not be able to fit simultaneously all data streams, while in a single site optimization some parameter changes may more easily compensate for model structural errors. However, given that the primary objective of the optimization of ORCHIDEE is to improve the model for large-scale applications (regional to global), the use of a multiple sites optimization is the only way to account for the 'within PFT' diversity. The limits arise when the objective is to study the response of a specific ecosystem (and not a generic PFT) to external drivers as the multiple site parameter set might be sub-optimal for the particular ecosystem. We have thus included few sentences in the text to highlight these issues (see end of first paragraph in section 3.1.2).

The mentioned typo is corrected.

Referee: Section 3.1.3: This section could be moved earlier in the paper as it is referenced in earlier parts of the results. Overall the flow of the results section could be improved.

Answer: We agree that this section could potentially be moved upfront. However, it would then somehow hide the first order message of the paper linked to the comparative performances of the two algorithms, gradient-based versus Genetic. We thus propose to keep the section where it is but to improve the overall flow of the results section. We have i) dropped the introduction in section 3.1.3 as it was redundant with the justification for multiple first-guess tests in the method section, ii) improved slightly the method, section 2.5, to better justify the use of several first-guess tests and iii) added a sentence at the beginning of the results section (first paragraph in 3.1) to explain the flow of results and the different sub-sections.

Referee: Page 11, Line 32: First use of "SG" – replace with S_Genetic? That is the abbreviation used in other parts of the text.

Answer: The typo is corrected.

Referee: Page 12, Line 27: Why does the pseudo-observation experiment perform poorly for $Z_{crit,litter}$? Why does it perform better for other parameters?

Answer: The poor performance for $Z_{crit,litter}$ is likely to be related to the relatively low sensitivity of the model outputs (NEE and LE) with respect to that parameter. This is by comparison to the other parameters. We have mentioned this reason in the text.

Referee: Page 13, Line 8ff: What drives different parameters to respond better or worse to different optimization methods?

Answer: Overall the main reasons that drive the differences in the parameter response to the different optimization methods are linked to the sensitivity of the chosen model output to each parameter, the prior parameter errors and error correlations and the overall shape of the cost function with respect to parameter sensitivity at any point in parameter space. The two algorithms explore parameter space in very different ways, therefore they deal with complicating issues (i.e. the existence of local minima) in different ways. We do have model equifinality in our results as the result of parameter error correlations and a lack of prior constraint (high prior uncertainty). As a result, we will not necessarily obtain the same posterior parameter vector but will achieve the same reduction in model-data misfit. The GA in particular is a random walk, therefore it is possible that it converges on a different parameter vector than the BFGS but with the same reduction in model-data misfit.

However it is difficult to investigate more deeply why one particular parameter is more sensitive to the Genetic or the Gradient-based algorithms in the context and scope of this paper. We thus believe that a general explanation of the reasons underlying these

differences is the appropriate level of detail. We have thus only slightly changed the text in section 3.2.1 to better highlight the reasons of these differences.

Referee: Page 13, Line 12: The error in $Z_{crit,litter}$ was mentioned on the previous page as 29%, please clarify.

Answer: It was rounded in the second case. Indeed, it makes no sense to put the numbers again here, where we compare the differences between the methods. The sentence has been changed to: "A few parameters are not well estimated by both methods, like $F_{stressh}$, HR_{Hb} and Z_{crit_litter}, having the largest difference with respect to the true value."

Referee: Page 13, Lines 27, 33: Same question as previous section; what drives different responses in different parameters?

Answer: The response is similar to that for the case of the pseudo-data experiment (previous section). We have thus referred in this section to the previous one for the explanation of the causes of the different responses between the different parameters.

Referee: Page 14, Line 20ff: Some grammar issues in the bullet points, and throughout this section (e.g., Page 15, Lines 5, 16 and 23).

Answer: The bullet point text together with the Summary section has been proofread.

Referee: Page 15, Line 18: I think you should mention this point in Section 2.4.3.

Answer: Indeed, this could be considered as an additional feature of the minimization algorithm. The following sentence is added in the end of Section 2.4.3:

"Contrary to gradient-based methods, random search algorithms allow to use any form of probability distribution functions for the observation and parameter uncertainties, and thus to use non-Gaussian PDFs."

Referee: Figure 1: In legend, use dots instead of lines to help guide the reader.

Answer: Lines are changed to dots on the figure.

Referee: Figure 3: Add error bars here, the uncertainty on the RMSD reduction is referenced in the text (e.g., Page 11, Line 28).

Answer: It's not the uncertainty on the RMSD reduction that is referenced in the text (Page 11, Line 28), but the full range of variation for the single-site Genetic optimization which goes from 50.5% with only one first guess to 55% with 16 first guesses (see Fig. 3, red line). As the notation 52±2% happened to be misleading, we change the text and now refer to the range of variation, 50.5–55.0%. Besides, there is no obvious uncertainty that we can add on Figure 3, because the RMSD reduction that we plot corresponds to the average of the "maximum RMSD reductions" obtained across all possible groups of N first guesses (X axis). Such measure is thus not directly a stochastic variable. Although we could add for each number of first guesses (X axis) the standard deviation across all "maximum RMSD reductions", it does not change much with the number of first-guesses, so we choose not to add this information as it would complicate the description of the figure and will not add much to the overall message.

Referee: Table S1: Missing units for parameters, as applicable.

Answer: The missing units for parameters are added.

Referee: Tables S3, S4: PFTs are numbered but not specified by name.

Answer: PFT names are added in the tables.

Referee: Figure S1: Have the legend in one panel and get rid of them elsewhere, they are just distracting/overlapping data points.

Answer: The legend is now kept only on the first panel and dropped for the others.

Referee: Figures S2, S3: Where in the main text are these figures referenced?

Answer: These figures were indeed added to support the interpretation of the results

but not properly referenced in the text. The references have been added.

Referee: Figure S3: Add error bars, same comment as Figure 3.

Answer: We choose not to add any error bars following the response above for Figure 3.

Referee : Figure S4: If some parameters do not apply to certain PFTs (e.g., $K_{pheno,crit}$), why are they optimized for that PFT? Is this an error in Table 1 and/or Table S1?

Answer: This is a mistake in Table S1 and Figures S4. Indeed, there are two specific parameters that apply only to selected PFTs ($K_{pheno,crit}$ applies to deciduous PFTs only and thus TempDBF, BorDBF, C3 grass; C_{Tsen} applies to TempDBF and BorDBF). The corresponding table/figures are corrected.

Best regards,

Vladislav Bastrikov

References

Chorin, A. J., and Morzfeld, M.: Conditions for successful data assimilation, J. Geophys. Res. Atmos., 118, 11522-11533, doi:10.1002/2013JD019838, 2013.

Lasslop, G., Reichstein, M., Kattge, J., and Papale, D.: Influences of observation errors in eddy flux data on inverse model parameter estimation, Biogeosciences, 5, 1311-1324, doi:10.5194/bg-5-1311-2008, 2008.

MacBean, N., Peylin, P., Chevallier, F., Scholze M. and Schürmann, G.: Consistent assimilation of multiple data streams in a carbon cycle data assimilation system, Geosci. Model Dev., 9, 3569-3588, doi: 10.5194/gmd-9-3569-2016, 2016.

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

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.