Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-34-RC2, 2017

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Assimilating solar-induced chlorophyll fluorescence into the terrestrial biosphere model BETHY-SCOPE: Model description and information content" *by* Alexander J. Norton et al.

Anonymous Referee #2

Received and published: 7 June 2017

This study evaluates the benefit of assimilating satellite-retrieved chlorophyll fluorescence into a mechanistic land surface model, to reduce the uncertainty in model parameters and simulated gross primary production (GPP). This study indeed tackles a critical issue in the current efforts towards making the most of diverse data infromation content when building efficient carbon cycle data assimilation systems.

There are, however, a few important issues in this manuscript, some of them critical. They are listed in the general comments below, followed by specific remarks/corrections.

Printer-friendly version

Discussion paper

General comments

First, while the manuscript is often fairly written, on numerous occasions sentences are redundant, strangely formulated, thus logical progression of arguments is hard to follow. Frankly, it sometimes feels as if the authors did not read themselves again before submitting the manuscript. It could be just be a matter of style, but in some occasions it simply results in a lack of clarity. While I tried to list specific parts in the *Specific comments* and *Technical comments* section, I suggest a strong effort of rewriting in general. That will also make the manuscript much more accessible to modellers/data experts outside the field of CCDAS or even data assimilation at large.

Second, and perhaps more importantly, the way the observation uncertainty used in Eq. 1 is defined is quite vague. Judging from the elements presented in Sect. 2.4, it seems that only the 'measurements' uncertainty of GOSAT retrievals of SIF is accounted for in C_D, neglecting the structural uncertainty (C_T, using the notation of Tarantola (1987)) of the BETHY-SCOPE model. If structural uncertainty is considered, that should be detailed in Sect. 2.4. If C_T is not taken into account, this would bear important consequences. While C_T is hard to estimate explicitly (although some diagnostic methods exist, e.g. see Desroziers et al. (2005), applied to land surface models by Kuppel et al. (2013)), its magnitude and structure might be comensurate or even dominant over measurement uncertainties when building C_D. Not including it in Eq. 1 would then largely underestimate the posterior uncertainty of parameters and, by propagation that of modelled GPP. As noted for another reviewer, this would constitutes a serious theoretical flaw in the scope of this study and make it unsuitable for publication.

Specific comments

P2, L11-12: This sentence is rather vague, can the authors be more precise and add

references to support this assertion?

P2, L27-28: Data assimilation is not only used with mechanistic models nor for terrestrial carbon cycle modeling. I suggest to reformulated, for example: "In the case of mechanistic models, this is done by constraining the simulated underlying processes.".

P2, L28-32: In this review of the state of the art, efforts from other groups to build "mechanistic" CCDAS might deserve to be cited as well, e.g. (Peylin et al., 2016) and the discussion/review by MacBean et al. (2016).

P3, L4-6: Some references would be necessary to back these assertions.

P5, L8: The last sentence of this paragraph feels rather clumsy, it should reformulated.

P5, L9: Table A1 is rather long and that is fair game given the number of parameters, yet to make it more reader-friendly I would suggest to:

- include a description column for each type of parameter,
- add the corresponding PFT between brackets for all PFT-dependent parameters, as is done for Vc_{max},
- add "subsection rows" with parameter categories (leaf growth, ecophysiology etc.).

P6, L2-3: It is because the PDFs of parameters and observations is treated as Gaussian that it can be described by their first two moments, mean and standard deviation (taken here as the metric of uncertainty, that might need to be specified here already well), not the other way around.

P6, L1-4: The definition of observations here should be more precise; the reader (especially if not familiar with the data assimilation vocabulary) would assume it relates to *measured* observations (as the previous paragraph uses "SIF observations" to designate measurements), while in a rigorous probabilistic framework it should refer to quantities in the observation space (including measurements and model outputs, see *General comments*).

P6, L12-13: I guess that the authors meant with this sentence that a) in a linear world H is independent from x, but b) this is an oversimplication, therefore c) bringing limitation in accuracy to a method relying on $H(x_0)$ to approximate $H(x_{post})$. It is not clear at all from the current formulation, which even almost suggest that because of linearity the choice of x_0 can influence the results (through a changing H)...

P6, L13-21: I am not sure how "the use of prior knowledge limits the effect of this problem": is it because we assume that the posterior parameters values will be close enough to the prior set, so that $H(x_0)$ is anyway similar to $H(x_{post})$ even if the model is not linear? In addition, the authors should give a reference for Eq. 1 (e.g., Tarantola, 1987) and explictly state that because linearity is assumed it takes the formed expressed in this manuscript (while the general equation is $C_{x_{post}}^{-1} = C_{x_0}^{-1} + H(x_{post})^T C_d^{-1} H(x_{post})$).

P7, L6: "those observations" is at best vague and at worst confusing, since it seems to relate to "observational uncertainty" (rather than "uncertainties") but again, observational uncertainties normally also includes the model component.

P7, L27 - P8, L8: In this whole paragraph (and the derived results and discussion), it would be important to mention which uncertainty is dealt with (random or systematic). Since only the random error can be studied this kind of framework, the potential impact of a systematic error (a bias) should be discussed as well, or at least mentioned.

P8, L10-11: Any proof/reference this it is sufficient? Even if it is expert knowledge, the authors should at least state it.

P8, L22: "Effective constraint" rather than "constraint", might be more accurate.

P9, L9: Which global physiological parameters are the authors referring to? Rows 37-68 in Table A1? See earlier comment on making Table A1 clearer.

P9, L10-17: The values of constraints in the text do not correspond to those shown in Table A1. Please update.

P10, L3: Maybe add between brackets than the chlorophyll parameters are C_{ab} components.

P10, L3-4: "During the assimilation" comes a bit abruptly. I guess the authors are talking about prospective data assimilation efforts with BETHY-SCOPE and SIF, please expand to make easier for the reader to understand.

P10, L9: This is a somewhat confusing formulation to say that uncertainty (and its subsequent reduction) is quantified as one standard deviation. Maybe giving this reference metric already in the methods would be helpful.

P10, L10-15: I suggest to have Fig. 3 (not mentioned in the text, maybe already in Sect. 2.4.?) on the color same scale as Figs. 4 and 5.

P11, L3: A figure showing the uncertainty reduction Could the authors briefly detail how they assessed the relative contribution of covariances to the total uncertainty in GPP? By summing the non-diagonal terms in $H_{GPP}C_xH_{GPP}^T$?

P11, L7: Could the authors briefly detail how they assessed the relative contribution

of covariances to the total uncertainty in GPP? By summing the non-diagonal terms in $H_{GPP}C_xH_{GPP}^T$?

P11, L17: As the authors state in the discussion, the fact that GPP is relatively insensitive to C_{ab} derives from the lack of a mechanistic link in the model between chlorophyll content and carboxylation rate. I suggest therefore to remove the "discussive" end of this sentence here and leave for the discussion where it is explained.

P11, L23-24: I disagree with the last part of this sentence: it seems to me that the increase in relative uncertainty contribution of physiological processes only says that they are less constrained than other processes, therefore the stated "limitations" is just *relative* to other well-constrained parameters. Without looking at the *absolute* value of uncertainty in GPP arising from each group of parameters (from which is then calculated the relative contribution), no statement can be made about how really "limited" is the constraint of SIF in ultimately reducing the uncertainty of a given parameter to simulate GPP.

P11, L27 to P12: I feel that an additional figure would be needed here, to show how the constraints in GPP from given parameters groups changes across the year in Temperate and Boreal regions. It could be for example a monthly-binned boxplot, each box corresponding to the range of constraint GPP for a given group of parameters, using colors or panels to separate regions. That would help the reader to support all the description given in the main text.

P12, L4: "exaggerated" seems quite subjective.

P12, L8: The parameter Vc_{max} is mentioned, then "these parameters", I guess referring to the different PFT components Vc_{max}? Please specify.

P14, L10-14: This might be suited for the discussion section.

P15, L10: How did the authors get this number?

P16, L810: I would move this sentence to the next paragraph, where diurnal dynamics are discussed.

P15, L10: How did the authors get this number?

P16, L31-35: And addtional figure showing the relative contribution of each parameters to modelled GPP uncertainty would make the results clearer. Perhaps using the same barplot setup as Fig. 1, except that y-axis would relative contribution to GPP uncertainty, and prior and posterior results could be shown using mirroring bars (2 y-axis would be needed then, one going upwards and the other downards).

P16, L33: "Free" sounds a bit odd here, what do the authors want to say?

P16, L34-35: I assumes that by "[. . .] only other free parameter controlling leaf area index other [. . .]" the authors mean that the model is highly *sensitive* to this parameter (i.e., large values in H), so adding to little prior *parameter knowledge* results indeed in large propagated uncertainty. The first aspect is however not quite clear from the current formulation. Since this separate consideration of *sensitivity* and *parameter knowledge* is essential when considering output uncertainty, here in the discussion I suggest detailing a bit more these aspects. Useful supporting references are, e.g., discussions in Dietze et al. (2014) and Kuppel et al. (2014).

P17, L1-2: This sentence ("The prevalence [. . .] global scale") is rather general and does not add much to the following one (which gives numbers). I suggest removing the former.

P17, L1-2: This sentence ("The prevalence [. . .] global scale") is rather general and does not add much to the following one (which gives numbers). I suggest removing the

former.

Technical comments

P2, L16: Definition of NDVI and EVI acronyms, first introduced here

P2, L23: *has* instead of *have*.

P2, L35: It is not the process that provides the constraints, rather the latter being constrained!

P6, L9: Replaces "equation 1" by "Eq. 1". It also applies to L17, to "equation [2,3,4]" on [P6;L26], [P7;L2-L4-L14] and [P10;L8].

P6, L10-11: Strange formulation, I would suggest: "[. . .] a Jacobian matrix (H), which is calculated around [. . .]"

P6, L10-11: Strange formulation, I would suggest: "[. . .] a Jacobian matrix (H), which is calculated around [. . .]"

P6, L26: "p. 71" instead of "pg. 71".

P7, L6: "its" instead of "it's".

P7, L10: "described" would be more accurate than "demonstrated"

P7, L27-29: "As might be expected" is quite subjective. I suggest to connect the two sentences: "[. . .] while uncertainty in forcing such as incoming radiation is not considered in the curret CCDAS setup, it is considered to be an important variable

driving SIF (Verrelst et al., 2015) and GPP (*rference needed*)."

P9, L2: "Table A1" instead of "Table 1".

P10, L7: If "as" refers only to the posterior uncertainty in GPP, it should then be replaced by "the latter being".

P11, L12: "stems" instead of "stem"

P12: "made up by" (L2) and "make up" (L8) are somewhat colloquial/vague here, it could be respctively replace by "arises from" and "contribute to".

P14, L4: Changing with "Second, we also increase [. . .]" might help the reader understand you are describing the other experiment.

P15, L7: I suggest "[. . .]SWRad, in both cases resulting in a relative reduction in the GPP uncertainty by about 78.6%".

P15, L17-18: "constraints" is repeated a lot here, I suggest: "[. . .] ultimately yields a global annual GPP estimate within $\pm$ 2.8 PgC.yr^{-1}.".

P16, L18: "however" seems somewhat redundant.

P16, L18: "PSII" should be defined on L11.

P17, L9: "feasible with" feels odd. Maybe "acheviable using"?

P17, L23-24: I suggest rephrasing as follows: "This in line with Koffi et al. (2015) who found limited sensitivity of simulated SIF to Vc_{max}."

P18, L7-8: The meaning is not clear, I assumed the authors meant "While including

this forcing uncertainty increases the prior GPP uncertainty, incorporating the former within SIF uncertainty itself mitigates the downstream effect on GPP."

P18, L16: Maybe replace "can also be" by "will also be".

References

- Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysisâĂŘerror statistics in observation space. Quarterly Journal of the Royal Meteorological Society, 131(613), 3385-3396.
- Dietze, M. C., Serbin, S. P., Davidson, C., Desai, A. R., Feng, X., et al. (2014). A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. Journal of Geophysical Research: Biogeosciences, 119(3), 286-300.
- Kuppel, S., Chevallier, F., & Peylin, P. (2013). Quantifying the model structural error in carbon cycle data assimilation systems. Geoscientific Model Development, 6(1), 45-55.
- Kuppel, S., Peylin, P., Maignan, F., Chevallier, F., Kiely, et al. (2014). Model–data fusion across ecosystems: from multisite optimizations to global simulations. Geoscientific Model Development, 7(6), 2581-2597.
- MacBean, N., Peylin, P., Chevallier, F., Scholze, M., & Schürmann, G. (2016). Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. Geoscientific Model Development, 9(10), 3569.
- Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., et al. (2016). A new stepwise carbon cycle data assimilation system using multiple data streams to

constrain the simulated land surface carbon cycle. Geoscientific Model Development, 9(9), 3321.

- Tarantola, A. (1987). Inverse problem theory: Methods for data fitting and model parameter estimation.
- Verrelst, J., Rivera, J. P., van der Tol, C., Magnani, F., Mohammed, G., & Moreno, J. (2015). Global sensitivity analysis of the SCOPE model: What drives simulated canopy-leaving sun-induced fluorescence?. Remote Sensing of Environment, 166, 8-21.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-34, 2017.

Printer-friendly version

Discussion paper