Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-25-RC2, 2016

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Consistent assimilation of multiple data streams in a carbon cycle data assimilation system" *by* Natasha MacBean et al.

Anonymous Referee #2

Received and published: 11 April 2016

This manuscript examines aspects of assimilating multiple data streams into carbon cycle models, includes discussion of the preceding literature and makes recommendations for the carbon cycle data assimilation (DA) community as to best practice when performing DA experiments. A real strength of this paper lies in the clarity of the description of the Data Assimilation problem.

Overall the work presented is well written, appears technically sound and should be easily reproducible. However the value of the individual parts of the manuscript feel somewhat limited, and as a whole I am not convinced they combine to make a complete piece of work. Although I don't doubt that setting up the DA system itself was technically complex, the experiments performed with it are rather limited in scope. My feeling is that it would have been easy to explore some further aspects of the carbon cycle DA problem and make the resulting manuscript much stronger with relatively little extra

Printer-friendly version

Discussion paper

work.

The "advice for land surface modellers" in section 4 is a good concept but could be better organised. For example the points "conduct preliminary..." and "set up experiments..." are very related. I think the list should be tidied up - perhaps broken into different sections, for example "understanding errors", "preliminary analyses" and so on. Each of these sections can then contain the smaller points.

The literature review section is reasonable but does not go into some of the preceding work in sufficient depth. In particular there are two studies I can think of that also look at carbon cycle DA problems with simple models that should have been dealt with in more detail. The Optic paper by Trudinger et al. (2007) is referenced, but a discussion of what experiments were performed and what they authors found is lacking. I think this is an important oversight given that this manuscript uses the same model. The Reflex paper by Fox et al. (2009) which looks at parameter estimation using a variety of DA techniques using a simple model and synthetic data isn't referenced. Furthermore the ordering of the manuscript feels a bit backward. One would normally expect the literature review to come prior to the experimental component and to set up the rationale for the experiments that follow.

I have the following major recommendations to make the manuscript publishable:

1) The experiments performed with the model need to be broader. There are several issues brought up later in the manuscript which could be easily examined. For example some simple experiments looking at populating the off-diagonal elements of the R matrix to set correlation between observations of S1 and S2 would seem to be an easy thing to do. I would be happy to see any sensible additional experiment though.

2) The literature review should be moved before the experimental section and modified so that it builds the rationale for performing the specific experiments undertaken. It should include greater discussion of the papers mentioned above. There are also classic problems in data assimilation which have not been well investigated in the car-

bon cycle to date such as localisation and errors or representativity and these have not been mentioned. They should be added into the discussion.

3) The "advice" list needs to be re-written to provide a bit more order. See comments above.

4) On page 11 at line 27 there is a statement suggesting that the data streams of s1 and s2 contain enough information to retrieve all the parameters individually for the quasi-linear model. This to me seems to be a flaw in the experimental design. Some of the conclusions from this part of the paper revolve around the linearity of the model, e.g. that differences between the step-wise and simultaneous experiments are minimal because of this. However given that the model is such that either set of observations can be used to determine both parameters it is not possible to say definitively that is the models linearity which is responsible for this. My hunch is that the authors are correct, but what would happen with a more complex linear model where not all parameters are observable from either one data stream? The only way to demonstrate this is by introducing a new model - which I do not recommend - however I think it is vital that the authors are clear about what can or cannot be deduced from these experiments.

I have the following minor comments:

1) The first paragraph of page 4 makes a lot of statements that are not referenced. It would be helpful to the reader who wanted to follow up on some of these aspects to provide references.

2) On page 5 I felt a bit more information was required about the model. How is the value of the functions F(t) being evaluated (possibly I have just misunderstood what is going on - so maybe just some clarification is needed).

3) Page 23, line 4, I am not sure what is meant by orthogonal here. Given that S1 and S2 are interdependent on each other in the quasi-linear model the observations of them (assuming the model is correct, which it is in these synthetic experiments) cannot

be not orthogonal. Perhaps the word "additional" would be better used here? Either that or I think the choice of "orthogonal" needs to be justified.

Typographic and small errors:

P03L10: step -> steps P03L19: one -> only P12L11: uniform -> constant (?) P13L11: than -> as P15L4-L11: this sentence needs to be broken up for clarity. P29L05: 2013.). -> 2013).

F2a: y-label should read "posterior" instead of "post"? F2b: y label should contain "%". F3caption: Equation should be 1-(RMSE_post/RMSE_prior)x100 F4b: as F2b

References:

Trudinger, Cathy M., et al. "OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models." Journal of Geophysical Research: Biogeosciences 112.G2 (2007).

Fox, Andrew, et al. "The REFLEX project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data." Agricultural and Forest Meteorology 149.10 (2009): 1597-1615.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-25, 2016.

Printer-friendly version

Discussion paper