Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-25-AC3, 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.

Natasha MacBean et al.

natasha.macbean@lsce.ipsl.fr

Received and published: 25 August 2016

UPDATE (24/08/2016) to: 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

In the first response to reviewer #2 (see above) we responded to the following comment by saying we agreed but needed more time to complete the experiments, and had agreed this with the Editor. This response therefore concerns our update to the point made below, and our "UPDATED RESPONSE" appears after the original response at the end of the document.

Printer-friendly version

Discussion paper

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.

» RESPONSE

We agree that we could, or should, have added more experiments. Indeed we thought of such experiments from the outset of this work but ended up not including such experiments for fear the paper was too long or the message too complex. We agree that the most obvious, and hopefully most informative, experiment would be one investigating the impact of having correlated observations and populating (or not) the off-diagonal elements of R. We considered other additional tests such as the impact of non Gaussian errors (although we have effectively done this by including an unccounted for bias as described in Section 3.2.2), and we considered examining temporal autocorrelation, but as we want to focus on the multiple data stream aspects we have just considered the correlation between the two data streams. We have implemented this test, but the results we obtained were not what we expected (little impact). As we think this is the most useful extra experiment to include, we have asked the editors for more time (from the 24th June) to investigate this issue and we will provide a further update to this response within the next month. However we will upload the response to the rest of the comments now so the reviewer has more time to look and reply should they wish. The results from these experiments will be presented in a separate section at the end of the experimental section (now Section 3).

»

» UPDATED RESPONSE

We have implemented the experiment the reviewer suggested, that is to test the impact of correlation between observation errors of the two data streams. We implemented a correlation between the observation errors for each time step of the model following the method of Trudinger et a.l (2007). We could have examined a temporal correlation as well, but as we want to focus on aspects related to multiple data stream assimilation we chose to only look at the cross-correlation between the data streams. We then tested the impact of both accounting for these covariance (correlation) terms in the prior covariance matrix, and ignoring them (i.e. not included them in R). We performed these tests using simultaneous case for both models.

To describe this additional experiment we have changed the text of the manuscript in the following sections:

âĂć Abstract We have added the following sentence: "In addition, we perform a preliminary investigation into the impact of correlated errors between two data streams for two cases, both when the correlated observation errors are included in the prior observation error covariance matrix, and when the correlated errors are ignored."

âĂć Introduction to the experimental results section (now Section 3) We have added this sentence to the introduction to the experimental results section, which itself is an addition to the original submission. The following sentence is an addition to the initial response to the reviewers posted at the end of June 2016.

"In addition to the above three challenges we have performed a preliminary investigation into the impact of correlated errors between the two data streams, which is a topic that has not yet been studied in the context of carbon cycle models"

âĂć Methods section 3.1.6 ("Experiments" – note previously Section 2.1.6 in the original submission) We have added the following paragraph to the end of this section: "For all the above tests wee assumed independence (i.e. uncorrelated errors) for both the parameters and observation covariance matrices, thus the R and B matrices were diagonal. In a final test we performed a simultaneous optimisation to examine the impact of having correlated errors between the s1 and s2 observations. Thus the random

Gaussian noise added to s1 for each time step was correlated to the noise added to s2. The correlated observation errors were generated following the method used by Trudinger et al. (2007 – paragraph 22). The added noise was time invariant, i.e. there was no correlation between one time step and the next as we were specifically looking at correlations between the observations. We tested both accounting for the correlated errors by populating the corresponding off-diagonal elements of the R (observation error covariance) matrix, and ignoring the correlated errors by keeping R diagonal. The reason for performing both tests was to demonstrate the possible real world scenario where correlated observation errors exist, but this information is not included in the optimisation due to a lack of knowledge as to how to characterise the errors. For both tests we performed optimisations using a combination of different of observation error and correlation magnitudes (observations errors between 0.05 and 20 in 9 uneven intervals, and observation correlations between -0.9 and 0.9 with an interval of 0.4). As in the above experiments, twenty random first guesses in the parameter space were used and 15 iterations of the inversion algorithm were performed."

âĂć Finally we have added a whole section to describe the results of this additional experiment – now Section 3.2.5. We will not repeat the text here as it is a clear new standalone section.

We initially found that the model set-up we had used for the set of experiments included in the original submission did not result in any difference when we included the off-diagonal (covariance) terms (accounting for correlation in the observation errors) in the observation covariance matrix (R) compared to when we did not include the off-diagonal terms. This is because the observation errors were small enough to accurately find the minimum of the cost function and the true value of the parameters, and therefore accounting for the correlation in the observation errors had no discernible effect. This was true for any magnitude of observation correlation (postive and negative). We hypothesised that accounting for observation covariance terms (or not) would be an issue if the observation errors were larger (note that larger observation error can be considered a proxy for anything that would result in lower information in the assimilation system). Therefore we then implemented a test with a range of observation errors and observtion cross-correlation. Indeed above a certain observtion error we did then see a difference between accounting for the off-diagonal terms in R matrix.

These results are described in Section 3.2.5 (entitled “Impact of accounting for correlated observation errors in the prior observation error covariance matrix”) and summarised in plots in Figure 7. We highlighted the key finding that at low observation error there is not a discernible difference if you do or do not account for correlated observation errors; however, at higher observtion error (or when the information content of the observations is reduced by another means) it does become important to accurately characterise the correlated errors. We feel this is an important point to make as correlations between observations are largely ignored by the modeling community in parameter optimisation studies, in part because we do not yet have an idea how to characterise the correlations between observations. We have also made the further point, relevant to Section 3.2.3, that accounting for correlation between observations is not possible when performing a step-wise assimilation.

âĂć Perspectives and advice section 4: We updated one bullet point in the advice section from the previous reviewer responses about correlation between observation errors: “Devote time to carefully characterising the parameter and observation error covariance matrices, including their correlations (Raupach et al., 2005), although we appreciate this is not an easy task (but see Kuppel et al., 2013 for practical solutions). In the context of multiple data stream assimilation, this should include the correlation between different data streams, particularly with higher observational uncertainty, though note that this is not possible in a step-wise assimilation.”

âĂć Conclusions: Finally we have the following sentence to the Conclusions: “We further note that the consequence of not accounting for cross-correlation between data streams in the prior error covariance matrix becomes more critical with higher observation uncertainty.”

Having made all these changes, we also wish to highlight to the reviewer that these experiments have taken some time, in part because this is a new topic that has not yet been fully investigated in any multiple data stream assimilation associated with terrestrial carbon models. As such although we knew what to expect in theory, the detail of results we have obtained beyond the "key finding" discussed above, have puzzled us slightly in that the pattern does not always correspond to our hypotheses. We have made tentative suggestions in the text as to why this is the case, related to non-linearity in the models resulting in inaccurate calculation of the posterior error covariance matrix (as well as higher observation error). We thus feel this topic merits further investigation, a point we have also made in the text. We ourselves plan to continue this investigation topic by starting from scratch and laying out fully our theoretical understanding from a mathematical standpoint using linear model equations. However for this work, given that it is a big topic that may merit a whole study in itself, and given this was suggested as an additional test and we feel we have at least gained one key insight, we hope that the reviewer feels it is a useful addition to this paper. Therefore we are submitting the results of this experiment as they stand for now, despite the fact we would like to (and will) investigate further. We hope that the reviewer now thinks the experimental section is broad enough. Indeed we have tried to further clarify the point of all the experiments (in three main "challenges") by linking them more to the issues related to multiple data stream assimilation in the literature review section (as detailed in the additional response to the reviewer).

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