24/08/2016

Dear Editor,

The following document contains the initial responses to both reviewers that were submitted in June this year, followed by the updated response to reviewer #2 based on the fact we needed more time to complete the additional experiment that reviewer #2 suggested.

We have tried our best to address all concerns raised by both reviewers, and have provided detailed responses below. However here we wish to highlight two main points:

1) Based on the response of both reviewers (but particularly reviewer #1) we realised that we had not been clear enough that this paper is aimed at demonstrating issues with *multiple* data stream assimilation for parameter optimisation, which is a topic that has not yet been extensively covered in the carbon cycle (or terestrial ecosystem model) data assimilation literature. Therefore the literature review and experimental sections are geared towards this objective alone, and are not an examination of all the issues and options faced when performing a data assimilation experiment. We have tried to clarify this point throughout the text and we have tried to answer all of the reviewers comments pertaining to this point (please see the detailed responses below). We believe this a very useful addition to the literature because, as just mentioned, this is not a topic that has received a lot of attention to date.

2) Reviewer #2 asked us to add an additional experiment to broaden the experimental results section, and in particular they suggested an experiment investigating the impact of correlated errors between the data streams. We strongly agreed with reviewer #2 on this issue, and indeed have been working hard on these tests since we received their comments. However, this is not a topic that has been investigated in the literature so far in relation to terrestrial model optimisation, even more so than the topic of multiple data stream assimilation itself. Therefore it has taken more time than expected because the results we were getting were new to us and not what we initially expected. We apologise for asking for several extensions to the deadline for author response as a result. As you can see in the more detailed Updated Response to reviewer #2 at the bottom of this document, this is still a topic that we think needs further investigation (as with any topic and especially one that is relatively new to the literature). However, we hope that reviewer #2 will be satisfied with what we have done so far, keeping in mind that it is a relatively new topic and could be a whole study in itself. For more detail we refer you to that updated response at the end of this document.

We have tried to fully address the concerns and suggestions made by both reviewers, which we felt were very useful, and we hope that we have improved the manuscript so that it is useful addition to the literature on this subject.

Yours sincerely,
Natasha MacBean

Interactive comment on "**Consistent assimilation of multiple data streams in a carbon cycle data assimilation system**" ***by*** **Natasha MacBean et al.**
Anonymous Referee #1

Received and published: 8 April 2016

General comments:
The paper addresses the question of the assimilation of multiple data streams to es- timate model parameters and initial conditions together with their uncertainty using variational method for simple C cycle models using synthetic observations.
The paper is organized around two parts. A first part presenting a variational data assimilation (DA) experiment for a simple, yet non trivial, quasi-linear model for the carbon cycle, and a non-linear toy model using multiple (2) data streams. The DA method, 4DVAR, and the experimental setup are succinctly but clearly described. The results of the experiments are extensively exposed, but - while VAR provides (via adjoint techniques) a set of tools to analyse the DA problem - not explained. The second part is devoted to a rather long but factual literature review of the studies using multiple streams to constrain LMS in general and their carbon component in particular.
While the paper illustrates some of the challenges of the model-data fusion problem, it does not describe any new idea, concept or tool, and thus does not represent a sufficiently substantial advance in modelling science. The advices presented at the end of the paper describes the golden rules for any DA experiment and the manuscript would benefit from a strict application of these advices. As it stands the paper only reproduces what other studies have done: performing the assimilation of multiple data streams following different scenarii.

>> RESPONSE
We thank the reviewer for their comments and detailed review.

i) Firstly we would like to address the assertion that the paper does not describe any new idea, concept or tool, and therefore does not represent a sufficiently substantial advance in modelling science. While indeed there is no new idea, concept or tool, this is a model experiment description paper to elucidate the concepts and problems for multiple data stream assimilation, particularly in reference to large-scale, complex land surface models (LSMs) that are included in earth system models (ESMs), and that may have to rely on a variational data assimilation method (more on this below in the comments about the VAR framework). To our knowledge this is the first paper that has brought together a investigation of the issues surrounding multiple data stream assimilation for LSMs from a methodological point of view. Many papers have used multiple data streams, as described in the literature review section, but very few (or none – to our knowledge) have highlighted the impacts and challenges around this topic. With the growing increase in the number and length of data streams, we tend to think that adding more data streams will definitely be beneficial for optimising a model. This paper aims to show that whilst that may be true, it is not the magic "black box" that researchers may

hope for, and many factors must be considered when carrying out this type of assimilation experiment, particularly when assimilating several different data streams of a different nature (i.e. flux and stocks, satellite data) or density (number of measurements).

We agree however that the structure of the major sections of the paper (with the literature review after the experimental section) does not help to illustrate that this is indeed an open issue in land surface modelling and that we try to solve this by looking at two simple but representative models (more on this below).

We tried to provide such a justification for the work in the introduction with paragraph starting:

P3 line 22: "Increasingly, researchers are attempting to bring these sources of information together to constrain different parts of a model at different spatio-temporal scales within a multiple data stream assimilation framework (e.g. Richardson et al., 2010; Keenan et al., 2012; Kaminski et al., 2012; Forkel et al., 2014; Bacour et al., 2015). However, whilst the potential benefit of adding in extra data streams to constrain the C cycle of LSMs is clear, multiple data stream assimilation is not as simple as it may seem"

This paragraph goes on to explain the issues further, but as mentioned we agree that we could make this more clear, particularly its relevance to optimisation of large-scale, complex LSMs, therefore we have added the following sentence after the sentence above, which we hope will strengthen the justification for this work:

"This is particularly true when considering a regional-to-global scale multiple data stream, multiple site optimisation of a complex LSM that contains many parameters, and which typically takes on the order of minutes to an hour to run a one year simulation."

ii) Secondly, we are not completely sure what the reviewer wants to say with this comment: "The results of the experiments are extensively exposed, but - while VAR provides (via ad- joint techniques) a set of tools to analyse the DA problem - not explained".

We have tried to explain the results we see carefully. Please could the reviewer provide further details on how they would like to see the results explained more? However, we would also like to emphasise and to remind the reviewer that the paper was not intended to explore the strength of variational methods, or the data assimilation set-up as a whole, but to demonstrate the issues with multiple data stream assimilation (e.g. issues of non-linearity when considering a step-wise or simulteanous simulation – we have given more details on this below). Therefore we do not want to add too many prior or posterior diagnostic tests as we already feel the paper is long enough. Furthermore, reviewer 2 has suggested adding additional tests that are more related to the specific issue of multiple data streams (e.g. the impact of having correlated errors between the observations). We are going to do the test requested by reviewer 2 as we feel it more directly fits in with the

aims of the paper, whereas although diagnostic tests are a nice addition they may make the paper too long.

Having said the above, we do feel that the aim of the paper, to specifically explore the issues of multiple data stream assimilation, has not been brought out enough in the text. In paricular, as the reviewer notes, the advice presented at the end of the paper are golden rules for any DA experiment, and thus they are too general for the stated aims of the paper. We agree this muddies the manuscript focus. Therefore we can understand that the reviewer was misled by this, which led to their comment about using "VAR as a set of tools to analyse the DA problem (in general)". We have therefore changed the text as stated in point i) above, and in addition we have changed the advice to land surface modellers section (4) to remove any points that are general to a DA experiment, so as not to confuse the reader, and to reflect more on the results (see comment below). We have put the revised text for this section at the bottom of this review.

Lastly we would like to insist that this is not a simple replication of what other studies have done. We believe our work is different in the following ways:
- No study (to our knowledge) has specifically investigated the difference between step-wise and simultaneous optimisation for multiple data streams, which is important in the context of large-scale LSMs using variational methods due to the heavy computational burden, as a step-wise optimisation may be preferable (as discussed from P3 line 27 onwards in the introduction).
- Similarly, no study has investigated the impact of bias/incompatibility/inconsistency in the observations or model on the assimilation results, again which is particularly important in the context of large-scale LSMs using variational methods due to the heavy computational burden
- No study has used synthetic observations and simple models to demonstrate these issues of multiple data stream assimilation when using complex large-scale LSMs that necessarily require the use of a variational schemes for land surface modelers
- Whilst some studies have noted that using one data stream can degrade the fit to the another, only a couple have fully explored why this is the case. In this study we explore issue this with the aid of synthetic observations and simple toy models, making it much easier to see the potential negative impact of such a situation.

We have tried to bring these points out more in the introduction, by putting the literature review before the experimental section as advised, by building the case for the following experiments with the simple models more within the literature review, by altering the advice section to be more focused on multiple data stream assimilation and finally by adding an extra introductory paragraph at the start of the experimental results section (see response *** below).
>>

I would recommend to shorten the literature review and to insert it before the experimental study and to perform a thorough analysis including sensitivity analysis, non-linearity issues, conditioning of the problem, information content.

>> RESPONSE

We agree with the reviewer about the order of the sections, and have put the literature review before the experimental study as discussed above. We have also shortened it by taking out any references to the experimental results (P21 lines 22 to 27 – which was added to the advice section as the literature review is now before the results) and P22 lines 1 to 8 has also been added to the advice section as these sentences accompanied P21 lines 22 to 27).

In addition, as recommended by the reviewer, we have deleted a few other sections in both the literature review that were superfluous (e.g. P21 lines 14-15 and lines 27 to 32, P22 lines 9 to 14) as well as sentences/bullet points in the advice section that were more related to general issues in data assimilation, i.e. even just with one data stream (all noted in the revised document, for example P28 line 21 to 30). However, reviewer 2 asked for additional discussion of two key papers that had not yet been discussed, so we have added this in to the literature review.

We have not included a preliminary sensitivity analysis for the following reason. Normally we would perform a sensitivity analysis either a) when optimising a more complex model with many different processes and parameters with data streams that only correspond to one part of the model, or b) if the computational time is long enough that excluding non-sensitive parameters becomes important for computational efficiency. Neither is the case here because the toy models in this paper have so few parameters (2 and 4), all of which we want to optimise because the model variables are sensitive to all parameters. All observations constain all parameters. As we have seen from a preliminary analysis of the Jacobian, all observations constrain all parameters, the relative constraint from each observation may change within the time window and will depend on the specific DA set-up including the trajectory within parameter space, the specific noise realisations (added random noise) etc. Investigating all this is beyond the scope of this paper because these are general DA issues, not pertinent to a multiple data stream assimilation paper. Therefore we do not wish to include a specific sensitivity analysis in this paper. However, to explain this to the reader we have added the following sentence in (now) Section 3.1.5 (P9 line 11 in the original manuscript):

"We have not performed a prior sensitivity analysis to decide to which parameters are important to include in the optimisation, as the model variables are sensitive to all of the (small set of) parameters. However, in the case of a more complex, large-scale LSM it is advisable to carry out such an analysis, particularly given the computational burden of optimising many parameters."

We agree that showing the non-linearity (or lack thereof, for the first model) of the two models would be beneficial for the reader. Therefore we have plotted 3D plots with the pairs of parameters on the x-axes and the cost function on the y-axis to show this. We have put this figure in the supplementary material and refer to it when describing the non-linear toy model in Section 3.1.2.

Conditioning of the problem is an important issue in general for DA, but as mentioned before we are not aiming to provide a general demonstration of all options in a DA set-

up. Here we did not condition the problem (normally done by scaling the parameter values to a certain range or normalising by the parameter error covariance) as this was not necessary with the small set of paramters.

Analysing the conditioning of the problem and information content (or observation influence) are interesting and useful aspects of a DA set-up, but we would do not want to specifically address this in this paper with different tests for the reasons mentioned above – that it is not the focus of a multiple data stream assimilation paper (therefore will dilute the message) and that these types of additional analyses will make the paper too long, especially given that we will add tests related to the correlated errors between data streams. However given the reviewer's comment below ("Page 18, line 20") on how we define to information content, we have further clarified in the text.

>>

Due to its apparent complexity and because of "the burden" of coding and maintaining an adjoint, VAR is not the most popular method within this field, however it offers a framework where diagnostic and prognostic tools can be clearly (and sometimes analytically) defined, the capabilities of VAR deserve to be fully exploited in the scope of this paper.

>> RESPONSE

Side note: Variational methods (VAR) refers to the method of adjusting the intial state of a system via assimilating over an assimilation time window, as opposed to sequential methods which update the system at the time of the analysis. IN this context, using derivative-based methods (where the tangent linear model or adjoint is needed) or global search methods could both be used in the context of a variational scheme. In this paper, we were careful to distinguish between methods, derivative (gradient)-based and global search. But we see the reviewer is using VAR as it is commonly used, referring to derivative methods for minimising the cost function, which is the method we use in this study.

*i) We might agree with the reviewer that "VAR is not the most popular method within this field" when all optimisation studies are taken into account, particularly for site-scale studies. However, VAR **is** the most popular (or only – to our knowledge) method used when optimising land surface models with multi-site datasets at regional to global scales, because of the computational load of running these larger-scale more complex models. Global search methods simply take too long when doing a large-scale, multi-site optimisation with a land surface model (although new approaches in ensemble methods and updated computational resources will hopefully allieviate this problem). In the literature review (now Section 2), we detail the global scale studies with LSMs that have been carried to this point – to our knowledge - (although we have added a reference to a paper that has been published since our submission – Raoult et al., 2016). All of them use a variational (derivative-based) scheme.*

It is for this reason that we have chosen to demonstrate these issues using a variational (we refer to this in the paper as a derivative-based for the reason given above) algorithm instead of a global search method (such as the genetic algorithm or MCMC methods). We specifically wanted to show land surface modelers the implications of using

*variational methods, because whilst many site-scale optimisations have been performed, we are increasingly (and necessarily) moving towards larger scale optimisations with multiple data streams, and in these cases it will likely that they have to use a variational scheme. This is also the reason we have chosen to do the tests with the non-linear model (which is more representative of a more complex LSM), because as a result they **may** (unknowingly) violate some of the assumptions associated with variational methods given their complexity and possible non-linearities. But we agree that the reasoning for using variational (derivative-based) methods and for doing the non-linear test was perhaps not explained clearly enough, and therefore we have added the following paragraph to the beginning of the experimental section (now Section 3) which now comes after the literature review section (2):*

*** "The three sub-sections in Section 2 highlight examples within a carbon cycle modeling context of the three main challenges faced when performing a multiple data stream assimilation, namely, i) the possible negative influence of including additional data streams into an optimization on other model variables; ii) the impact of bias in the observations, missing model processes or incompatibility between the observations and with the model, and iii) the difference between a step-wise and simultaneous optimization if the assumptions of the inversion algorithm are violated, which is more likely to be the case with non-linear models when using derivative-based algorithms and least-squares formulation of the cost function. The latter point is important because derivative methods (compared to global search) are the only viable option for large-scale, complex LSMs given the time taken to run a simulation."

This (and the rest of the paragraph inclded here, which links the sections in the review to the experimental sections) also answers reviewer's 2 request to build the case for the experiments in the literature review and to link the literature review to the experimental section.

ii) We do not understand the last part of the reviewer's comment above that "VAR specifically offers a framework where diagnostic and prognostic tools can be clearly defined". This is the true for any Bayesian DA framework, including those that use global search methods. In that sense we do not know what the reviewer is referring to when he states that "particular capabilities of VAR (itself) should be exploited in this paper". But we take it to mean that we have not explored included all diagnostic tools that are available within a Bayesian DA framework. As we have discussed above, we think this reflects the reviewer's misunderstanding of the aim of this paper. The assimilation methodology (including exploiting the capabilities of VAR) is not (or at least should not be) a relevant issue for the message of the paper, and therefore we should not include many extra diagnostic tests. Though as discussed we take responsibility for the misunderstanding over the focus of the paper, and admit that the focus (issues of multiple data stream assimilation, particularly in relation to LSMs) was not defined clearly enough and we have taken steps to address this (see above).
>>

Specific comments:
- Page 2: "Observations allow us to understand the system up until the present day, but they cannot tell us about the future (...). They also cannot distinguish between the complex interactions that may occur between different processes". I strongly disagree with this statement, observations do carry information about the future through the deterministic processes that, we believe, govern our world.

>> RESPONSE
We agree with the reviewer – we did not intend to make this point so strongly. Therefore we have changed this sentence to:
"Observations allow us to understand the system up until the present day and provide inference about how ecosystems may respond to future change. However, their use in estimating model state variables and boundary conditions has limited use beyond diagnostic purposes, and they can be limited in their spatial coverage. They also do not contain all the information we may need to distinguish between the complex interactions that may occur between many different processes"
>>

- Pages 5-6, lines 12-18: I found the input/output terminology on page 5, line 17-19, a bit misleading. A brief summary of dynamics of the model as described in the work of Raupack 2007 could be useful. The models and the dynamic variables they describe try to encompass different time scales from diurnal to potentially much longer time scales, and the variables themselves are likely to differ by several order of magnitudes. A discussion about the implication of the different typical scales could enlighten some of the challenges. In the description of the experiments details concerning the time step size, observation window and observation frequency could be useful.

>> RESPONSE
We agree with the reviewer for the most part and have changed the following in Section 3.1.1 ("Simple carbon model"):

i) "The litterfall is an output of s1 and an input to s2 and is a fraction of the aboveground carbon reserve as represented by k1s1."
→ "The litterfall is an output of s_1 (aboveground biomass) and an input to s_2 (belowground biomass) and is calculated as a constant fraction of the aboveground carbon reserve, defined by k_1s_1".

ii) We do not want to repeat the description of the dynamical behaviour of the model as detailed in Raupach (2007) because it has already been described in depth by Raupauch (2007) and for the sake of brevity. But we have changed the words "model behaviour" to "dynamical behaviour of the model" (P5 line 22).
We have also changed this sentence to be clearer about what this model :
"It is based on two equations that describe the temporal evolution of two carbon pools, s1 and s2:"
to:

"It is based on two equations that describe the temporal evolution of two living biomass (carbon) pools, s_1 and s_2, and the biomass fluxes between these two pools"

iii) And finally we have added a part to the following sentence to detail that the model operates at a daily time step:
"It is based on two equations that describe the temporal evolution (on a daily time step) of two living biomass (carbon) pools, s_1 and s_2, and the biomass fluxes between these two pools".

The reviewer is also right that we have not defined the observation window or frequency. We have chosen to add these details into the section on the generation of synthetic observations (now Section 3.1.5 – Optimisation set-up: parameter values and uncertainty, and generation synthetic observations). Therefore after the first sentence of this section we have added the following:

"We optimised a ten-year time window for the simple carbon model, in order to capture the dynamics of the s_1 and s_2 pools over a time period compatible with typically available observations. For the non-linear toy model, which did not correspond to physical processes in the terrestrial biosphere, we ran a simulation over a window of 100 integration (steps) of the equations. The observation frequency was daily, corresponding to the time-step of the simple carbon model (a value of 1 for the non-linear toy model), and the observation error ... (as above)"
>>

- Page 6, line 28: "including measurement and model errors", how to include model error without a weak constraint formulation?

>> RESPONSE
Weak-constraint formulations allow explicitly reducing the model error by including some of its drivers (e.g., a bias in a prognostic equation) in the control vector, provided we have some knowledge of the statistical properties of these drivers. The latter requirement has dramatically reduced the use of this formulation and therefore we only use the standard formulation where the full model uncertainty is simply represented in the observarion error covariance matrix R. In the field of ecosystem model parameter optimisation, this is the standard.
>>

- Page 7, line 6: "strong linear dependence of the model to the parameters", 4DVAR is the perfect framework where this issue should and could be investigated as advised in the section "advice for LMS modellers".

>> RESPONSE
Here the reviewer is referring to the description of the assumptions of using the quasi-Newton algorithm (a derivative method) for finding the minimum of the cost function.

So indeed, we have explored this issue of the impact of violating the assumption of "strong linear dependence of the model to the parameters" specifically by including a section where we use this framework to try to optimise a non-linear model. We stated this objective in the description of the Non-linear toy model description (now Section 3.1.2), but we have further emphasised this point by changing this sentence:
"In order to illustrate the challenges associated with multiple data stream data assimilation for more complex non-linear models, we defined a simple non-linear toy model based on two equations with two unknown parameters"
to
"In order to illustrate the challenges associated with multiple data stream data assimilation for more complex non-linear models, especially when using derivative methods, we defined a simple non-linear toy model based on two equations with two unknown parameters".

We have also emphasised the point of this test with the paragraph at the beginning of the experimental section (see *** above).

And we have now added in an introductory sentence at the beginning of (now) Section 3.1.6 (Experiments) which says the following:
"The specific objective of the following experiments is to test the impact of a bias in the observations that is not accounted for in the R matrix, and the impact of using derivative methods with non-linear models (as may be necessary with large-scale LSMs), particularly with reference the differences that may arise between step-wise and simultaneous optimisations."

These results of this test about using non-linear models with derivative methods were described in now Section 3.2.3 (Difference between the step-wise and simultaneous approaches in the presence of a non-linear model) and discussed further in Section 3.2.5 (Lessons to be learned when dealing with non-linearity).

Therefore we strongly feel we have addressed this issue, albeit that we did not emphaise this point enough.

As above however we are confused by the fact that the reviewer says that "4DVAR is the perfect framework where this issue could be investigated", as this issue could be investigated within any DA framework. And again as above we do not think that this should be the focus of this paper, except in a context of multiple data stream assimilation. Accordingly we have re-ordered and deleted many points from the advice section to make it much more specific to multiple data stream assimilation, and not about general DA issues relevant to only one data stream.
>>

- Page 7, line 2: statement page 7 line 2 requires the model/observation operator to be linear as is discussed on page 17 line 20-31.

- Page 9 : concerning the experiment where only one observation for s2 is considered, worth mentioning that it corresponds (does it?) to the situation where only one estimation, say for soil C stock, is available. In this case is it used as a prior for s2 or as an observation later in the time window thus allowing the model to create correlation with other variables and parameters?

>> RESPONSE
Indeed it does correspond to this situation, and in this experiment we have taken the first observation (comparable to optimising the initial condition). We have added this point to the sentence the reviewer is referring to (end of P9):
"An additional test was included for the simultaneous assimilation in order to test the impact of having a substantial difference in the number of observations for the data stream included in the optimisation, as may be the case for belowground (e.g. soil) biomass observations in reality. Therefore in test case 3b, only one observation was included for data stream s_2."
>>

- Page 11, lines 14-17: discussion about "good or moderate reduction in RMSE for variables not included in any assimilation (...)" why is the reduction so poor for this flux? can this be expected from a model sensitivity analysis.

>> RESPONSE
Not necessarily, but we may expect that the fit is not as good for variables not included in the assimilation. We could see from a sensitivity analysis if changing the parameters included in the assimilation would change the model variable (i.e. if the model variable is sensitive to those parameters) but we could not know HOW they would affect it, or if the result of the assimilation is closer to the observations or not. In this study we know that all variables are sensitive to all parameters. This is also a general optimisation issue that may be faced when only one data stream is included.
>>

- Page 18, lines 16-19: "Rather if the model sensitivity to the parameters is very non-linear, multiple combinations of parameter values may exist that result in a similar reduction of the cost function (multiple minima), but provide a different fit to each data stream". This is exactly a crucial aspect that the paper should focus on, simplified and toy models are meant for this.

>> RESPONSE

We feel we have investigated this in the paper. In particular we did test multiple first guesses precisely to see if we had an issue with multiple minima, or indeed with parameter equifinality. This is detailed at the beginning of the results section (now Section 3.2) with a figure on the reduction in cost function from all twenty first guesses in the supplementary material. We did not have a problem with multiple minima, we find that in general the same reduction in the cost function is found (as described). However as already discussed at length above we disagree that the paper should focus on this, given it is a paper about multiple data stream assimilation. This is an issue that could arise with one data stream, therefore although we briefly describe the twenty first guess results at the beginning, in order to demonstrate that we have found the global minimum (or at least close to the global minimum), we do not consider that we should go into more depth on this topic in this paper. We do agree it is a key topic, but would be more appropriate for a general DA tutorial, which is not the purpose of this paper. However, we also agree that given we do not focus on the paper, this section is rather speculative and superfluous, and therefore we have removed it (P18 lines 16 onwards).
>>

- Page 18, line 20: information content not defined, and more generally the expression "enough information" appear twice in the text but never made explicit.

>> RESPONSE
The reviewer is right that information content here is not defined and we have not been explicit when saying "enough information". We have changed this to agree with how we refer to information elsewhere (e.g. in the introduction we do make several references to what we mean by information at those point, with the sentence: "These data bring information on different spatial and temporal scales,", with the bullet points following that detailing the temporal and spatial scales each data stream contains) so this is now: "spatio-temporal information content" and "enough spatio-temporal information" [to constrain the parameters].

We also agree that in some other places in the text we have not been explicit as to what we mean by information . We have changed this where the word "information" is not clear in the text. For example we have changed P4 line 11 to be: "information on the error covariance", and P22 lines 22 to 24 to be: "The study of Keenan et al. (2013) was particularly notable in its aim to quantify which data streams provide the most information (in terms of model-data mismatch) and how many data streams are actually needed to constrain the problem".
>>

- Page 18, lines 23-29: how to find the "troublemaker" and "peacemaker"?

>> RESPONSE
Given the re-structuring of the advice section (see bottom of the review), and the addition of some aspects of the literature review to the advice section, given the literature review

is now before the experimental section, we see that this whole section runs the risk of repeating what is in the advice section, and the discussion provided at the end of this section, including the words "troublemaker" and "peacemaker" (e.g. P18 lines 16 onwards) is somewhat speculative and superfluous (see comment "Page 18, lines 16-19" above). Therefore given another request to streamline and cut down the paper, we have removed this section, and added the following sections to the advice and perspectives section (Section 4):

"Most optimisation studies with a large-scale LSM use derivative methods based on a least-squares approach, and therefore rely on assumptions of Gaussian probability and linear model sensitivity. However,"

"it may not be possible to find the true global minimum of the cost function and the characterisation of the posterior probability distribution will be incorrect. This is a particular problem if the posterior parameter error covariance matrix is then propagated in a step-wise approach, although these issues are relevant to both step-wise and simultaneous assimilation."

"Note that performing a number of tests starting from different random "first guess" points in parameter space can help to diagnose if the global minimum has been reached, as outlined in Section 2.1.6 and discussed at the beginning of the results (Section 2.2)."

We have put the sentence: "An important finding of the results presented for the non-linear toy model in Section 2.2.3 is that degradation in another data stream is not necessarily the result of a bias or incompatibility between the observations and the model" at the end of Section (now) 3.2.3 (the non-linear toy model section).

Finally we have removed latter part of the section altogether (P18 lines 16 onwards), given the reason above. However, we have added a sentence into the advice and perspectives section that mentions there are several diagnostic tests that can be used if you want to determine the relative influence or constraint brought about by different data streams (e.g. the observation influence metric and the degrees of freedom of signal). We have also added that we have not investigated these metrics here as they are not useful for such simple models with so few parameters, and therefore were beyond the scope of this paper.
>>

- Page 26, lines 16-17: biases and inconsistencies, and other problematic features, could be addressed prior to optimisation in the context of the linearisation of the model.

>> RESPONSE
That is true in an ideal case. But in practice it is rarely done unless there is a very obvious bias or inconsistency between the model and the observations (hence why we try to demonstrate its importance in this paper) because in most cases it is not obvious that there is a bias. For example, clearly in the studies we have reviewed, it has not been

possible to see the bias in FAPAR data prior to the optimisation. The assimilation revealed this bias. It is not easy to validate satellite data so it is unclear how these biases may be revealed (for this particular example which we have highlighted).
>>

- Page 29, lines 25-26: "it is crucial to understand the assumptions and limitations related to the inversion algorithm used" yet I feel that the paper did not provide the analysis, though possible with VAR, that would have helped understanding these as- sumptions and limitations in the case of the "simple" models presented here.

>> RESPONSE
We feel we have provided this analysis related to the assumptions of linearity requrired by the inversion algorithm, as described above in the reviewer's comment above that begins with "Page 7, line 6".
Again, these issues are general DA issues therefore we were not aiming to do a full exploration of all the assumptions related to the inversion algorithm, but have highlighted those that are particularly pertinent to multiple data stream assimilation, e.g. the results in Section 3.2.3 (Difference between the step-wise and simultaneous approaches in the presence of a non-linear model), which were discussed further in Section 3.2.5 (Lessons to be learned when dealing with non-linearity).
We welcome further suggestions from the reviewer on how we can improve the description of the non-linear model section results, which aim to explore these issues in detail. We have attempted to make the point of these experiments clearer, as described in the above comment (Page 7 line 6).
>>

Technical corrections: - On page 1 line 25: "data stream" instead of "data steam". - On page 3 line 30: "matrices" instead of "matrixes".

>> Corrected.

- In Table 1 : for the non-linear toy model the observation uncertainty for s2 is set to 0.5 whereas it is set to 5 for the simple carbon model, shouldn't it be 5 instead of 0,5?

>> RESPONSE
No, the s1 and s2 observations are very different entities for the different models. The uncertainty was set as a defined 10% of the mean value over the whole timeseries for each pseudo-observation (derived from multiple first guesses of the model). As the magnitude of the s2 observations is larger for the simple carbon model, the associated uncertainty was larger. However the magnitude of s1 and 2 was about the same for the non-linear toy model, so the uncertainty is the same. However, the reviewer has highlighted that this was not defined in the text, therefore we have added the following

sentence in (the new) Section 3.1.5 (Optimisation set-up: parameter values and uncertainty, and generation of synthetic observations) and in the caption of Table 1:
"The observation error was set to 10% of the mean value for each set of pseudo-observation derived from multiple first guesses of the model".
>>

****

(NEW ADVICE SECTION 4 – remains in Section 4 after the toy model experiments)

4   Perspectives and advice for Land Surface Modellers

Although it is clear that in many cases, increasing the number of different observations in a model optimisation provides additional constraints, challenges remain that need to be addressed. Many of these issues that we have discusssed are relevant to any data assimilation study, including those only using one data stream. However, most are more pertinent when considering more than one source of data. Based on the simple toy model results presented in here, in addition to lessons learned from existing studies, we recommend the following points when carrying out multiple data stream carbon cycle data assimilation experiments:

- If technical constraints require that a step-wise approach be used, it is preferable (from a mathematical standpoint) to propagate the full parameter error covariance matrix between each step. Furthermore, it is important to check that the order of assimilation of observations does not affect the final posterior parameter values, and that the fit to the observations included in the previous steps is not degraded after the final step (e.g. Peylin et al., 2016).

- 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, though note that this is not possible in a step-wise assimilation.

- The presence of a bias in a data stream, or an incompatibility between the observations and the model, will hinder the use of multiple observation types in an assimilation framework. Therefore it is imperative to analyse and correct for biases in the observationsand to determine if there is an incompatibility between the model and data. Alternatively, it may be possible account for any possible bias/inconsistency in the observation error covariance matrix, R, using the off-diagonal terms or inflated errors (Chevallier, 2007), or by using the prior model-data RMSE to define the observation uncertainty.

- Most optimisation studies with a large-scale LSM require the use of derivative-based algorithms based on a least-squares formulation of the cost function, and therefore rely on assumptions of Gaussian error distributions and quasi model linearity. However, if the these assumptions are not met it may not be possible to find the true global minimum of the cost function and the characterisation of the posterior probability distribution will be incorrect. This is a particular problem if the posterior parameter error covariance matrix is then propagated in a step-wise approach, although these issues are relevant

to both step-wise and simultaneous assimilation. Therefore it is important to assess the non-linearity of your model, and if the model is strongly non-linear, use a global search algorithms for the optimisation – although at the resolution of typical LSM simulations (≥0.5x0.5°) this will likely only be computationally feasible at site or multi-site scale.

In addition to the above points, we have investigated the impact of a difference in the number of observations in each data stream in this study. Test case 3b, in which only one observation was included for the s_2 data stream instead of the complete time-series, shows that a substantial difference in number of observations between the data streams can influence the resulting parameter values and posterior uncertainty (compare test cases 3a and b in Fig. 2 for the simple C model and Fig. 4 for the non-linear toy model) as each data stream will have a different overall "weight" in the cost function. Xu et al. (2006), among others, have mentioned the possible need to weight the cost function for different data sets. Different arguments abound on this issue. Some contend that the cost function should not be weighted by the number of observations because the error covariance matrices (**B** and **R**) already define this weight in an objective way (e.g. Keenan et al., 2013), and we would agree with this assertion. It should not be necessary to weight by the number of observations in the cost function if there is sufficient information to properly build the prior error covariance matrices (**B** and **R**).

It is always useful to investigate the issues such as those discussed here by setting up synthetic experiments, as in this study, to understand the possible constraint brought by different data streams, and the impact of a possible bias and observation or observation–model inconsistency. Note also that performing a number of tests starting from different random "first guess" points in parameter space can help to diagnose if the global minimum has been reached, as outlined in Section 2.1.6 and discussed at the beginning of the results (Section 3.2). Furthermore, several diagnostic tests exist to help infer the relative level of constraint brought about by different data streams, including the observation influence and degrees of freedom of signal metrics (Cardinali et al., 2004). Performing these tests was beyond the scope of this study, particularly given that the simple toy models contained so few parameters, but such tests may be instructive when optimising many hundreds of parameters in a large-scale LSM with a number of different data streams.

Aside from multiple data stream assimilation, other promising directions could also be considered to help constrain the problem of lack of information in resolving the parameter space within a data assimilation framework, including the use of other ecological and dynamical "rules" that limit the optimisation (see for example Bloom and Williams, 2015), or the addition of different timescales of information extracted from the data such as annual sums (e.g. Keenan et al., 2012). Finally we should also seek to develop collaborations with researchers in other fields who may have advanced further in a particular direction. Members of the atmospheric and hydrological modelling communities, for example, have implemented techniques for inferring the properties of the prior error covariance matrices, including the mean and variance, but also potential biases, autocorrelation and heteroscedasticity, by including these terms as "hyper-parameters" within the inversion (e.g. Michalak et al. 2005; Evin et al., 2014; Renard et al., 2010; Wu et al. 2013;). Of course this extends the parameter space – making the

problem harder to solve unless sufficient prior information is available (Renard et al., 2010), but such avenues are worth exploring.

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 work.

>> RESPONSE

We thank the reviewer for their clear and constructuve review of our manuscript. We understand all his/her concerns about the different parts combining to make a complete piece of work, and we have tried to address these concerns by following all of their suggestions, as detailed below.

>>

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.

>> RESPONSE

We agree with the reviewer on this. We have re-ordered the advice section accordingly taking into the suggestions above. However, we have also deleted many of the points related to general DA issues (e.g. conduct preliminary sensitivity analyses) as we felt, particularly after reading reviewer 1's comments, that this was confusing the focus of the paper on multiple data stream assimilation. Although many of the issues we raise are indeed general issues related to the assimilation of only one data stream, we tried hard in the manuscript to show how they affected an assimilation with more than one data stream. However we can now see that some of the points made in advice section were counteractive to this goal and could confuse the reader. We hope that in the process we have also tidied up the list, but given the list is now shorter and (hopefully) more focused, we have not broken the points up into sections. However we would be happy to have sub-sections instead of bullets for each of the points. We have put the new advice section at the bottom of this response.

>>

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.

>> RESPONSE

We agree with the reviewer about the structure of the paper and so have put the literature review before the experimental component and slightly re-ordered it to better fit as an introductory section (see point (2) below). We have addressed the suggestions for additional papers below.

>>

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 exam- ple 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 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).

>>

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

>> RESPONSE

We have moved the literature review before the experimental section and have removed some sections, either those that described the experimental results (P21 lines 22 to 27 and P22 lines 1 to 8 – which have now been included in the "advice and perspectives section 4 – see the end of the response), and we have deleted sections that we felt were superfluous, in order to shorten the length as requested by reviewer 1 (e.g. P21 lines 14-15 and lines 27 to 32, P22 lines 9 to 14).
We have added in more reference to the Trudinger and Fox et al. papers but we have not discussed these in too much detail because we want the emphasis of the literature review to be on multiple data assimilation. In this context the Fox et al paper is perhaps more relevant, so we were wrong not to include it in the original text. It is now included in Section 2.1 – Extra constraint from multiple data streams (P21 line 9 before "Thum et al."). The focus of the Trudinger paper is on testing the assimilation set up more than testing issues related to multiple data stream assimilation, and therefore we have not discussed the paper in too much depth. However, given we do want to emphasise the focus on multiple data stream assimilation (please see the response to reviewer 1 for

further comments and changes to the manuscript in this regard), we have expanded the last paragraph in the introduction so it starts with the following:

"This tutorial-style paper highlights some of the challenges of multiple data stream optimisation of carbon cycle models discussed above. Note that we do not aim to explore all possible issues related to a DA system, for example the choice of the cost function, minimization algorithm, or the characterization of the prior error distributions; indeed previous studies have investigated such aspects at length (e.g. Fox et al., 2009; Trudinger et al., 2007), therefore we refer the reader to these papers for more information. Section 2 reviews recent carbon cycle multiple data stream assimilation studies with reference to some of the aforementioned challenges. Section 3…"

We hope that these additions are sufficient?

Finally we have moved the following section from the advice to the literature review (end of Section 2.2 – impact of bias) because we felt it was better placed there and gave more context to the discussion on bias in FAPAR data seen in previous studies:

"Aside from simple corrections, Quaife et al. (2008) and Zobitz et al. (2014) suggested that LSMs should be coupled to radiative transfer models to provide a more realistic and mechanistic observation operator between the quantities simulated by the model and the raw radiance measured by satellite instruments. This proposition followed the experience gained in the case of atmospheric models for several decades (Morcrette, 1991)."

We have tried to further build the rationale for performing the experiments with the toy models throughout the literature review, which we also hope was partly achieved by cutting out speculative and superfluous sections.

To further help link the literature review and the experimental section we have added an introductory paragraph to (the new) Section 3 ("Demonstration with two simple models and synthetic data") that summarises the issues raised in the previous section and introduces the experiments at the same time. Therefore the following paragraph has been inserted before Section 3.1 ("Methods"):

"The three sub-sections in Section 2 highlight examples within a carbon cycle modeling context of the three main challenges faced when performing a multiple data stream assimilation, namely, i) the possible negative influence of including additional data streams into an optimization on other model variables; ii) the impact of bias in the observations, missing model processes or incompatibility between the observations and with the model, and iii) the difference between a step-wise and simultaneous optimization if the assumptions of the inversion algorithm are violated, which is more likely to be the case with non-linear models when using derivative-based algorithms and least-squares formulation of the cost function. The latter point is important because derivative methods (compared to global search) are the only viable option for large-scale, complex LSMs given the time taken to run a simulation.

This section aims to demonstrate these challenges using simple toy models and synthetic experiments where the true values of the parameters are known. Most importantly this framework also allows us to investigate the impact of biases and violation of assumptions related to linearity (as discussed in Section 2.2. and 2.3), which are not always evident with real data and large-scale models. Thus the following sections include a description of the toy models together with the derivation of synthetic observations, the inversion

algorithm used to optimise the model parameters and the experiments performed, followed by the results for each test case."

Finally, we have included one sentence in the literature review regarding representativity – if I have understood the suggestion of the reviewer correctly ("The spatial distribution of each data stream is also important, especially for heterogeneous landscapes (Barrett et al., 2005; Alton, 2013)") but we did not discuss this or the localisation problem further as we would like to keep the focus on multiple data stream assimilation and not general DA issues. Indeed, we have modified and added certain sentences throughout to try to reinforce this main focus of the paper (see response to reviewer 1).

>>

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

>> We agree and have done this – please see the response above and the new text at the bottom.

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.

>> RESPONSE

We thank the reviewer for pointing out the lack of clarity here. Indeed it is not possible to say definitively that it is related to the model linearity and we also feel fairly sure that if we had a more complex linear model that not all parameters would be observable from one data stream. We have therefore clarified this in the text by changing "under this assimilation set-up" to "under this assimilation set-up with this model", and with the sentence at the end of the section:
"However, we cannot definitively say whether this is due to the simplicity or relative linearity of the model – it is possible that observations of variables in more complex linear model would not be able to retrieve the true values of all parameters."

\>\>

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.

\>\> RESPONSE

Thank you for pointing this out. We have tried to provide some references. This paragraph now reads:
"Mathematically, the optimal approach is the simultaneous, but computational constraints related to the inversion of large matrices or the requirement of numerous simulations, especially for global datasets (e.g Peylin et al., 2016), and/or the weight of different data streams in the optimisation (e.g. Wutzler and Carvalhais, 2014), may complicate a simultaneous optimisation. On the other hand, in a step-wise assimilation the parameter error covariance matrix has to be propagated at each step, which implies that it can be computed. If the parameter error covariance matrix can be properly estimated and is propagated between each step, the step-wise approach should be mathematically equal to simultaneous. However, many inversion algorithms (e.g. derivative based methods that use the gradient of the cost function to find its minimum) require assumptions of model (quasi-) linearity and Gaussian parameter and observation error distributions (Tarantola, 1987, p195)."

We have also changed this sentence to explain more what we mean:
"If these assumptions are violated, or the error distributions are poorly defined, it is likely that the step-wise will not be equal to the simultaneous, and that information will be lost at each step.",
to: "If these assumptions are violated, or the error distributions are poorly defined, it is likely that the step-wise will not be equal to the simultaneous, because information will be lost at each step due to an incorrect calculation of the posterior error covariance matrix at the end of the first step.".

\>\>

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).

\>\> RESPONSE

Indeed we have not described how the function F(t) is calculated at all! Thank you for pointing this out. We have added the following sentence in:

"The F(t) forcing term is a random function of time ("log-Markovian" random process) representing the effect of fluctuating light and water availability due to climate on the NPP (Raupach, 2007 – Section 5.3)."
Also, following some of the comments from reviewer 1, we have added in further clarifications in this section, including for example the model time step in this sentence:
"The first term on the right-hand side of Eq. (1) corresponds to the Net Primary Production (NPP) i.e. the carbon input to the system as a function of time, represented by F(t), weighted by factors (the two fractions in parentheses) that account for the size of both pools, in order to introduce a limitation on NPP.", and "It is based on two equations that describe the temporal evolution (at a daily time step) of two living biomass (carbon) stores, s_1 and s_2, and the biomass fluxes between these two stores".

>>

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.

>> We agree with the reviewer, this was a lax use of the word in this context. We have changed it to "additional".

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).

>> RESPONSE

Thank you for these corrections, we have changed them all.
Also we have changed P15 L4-11 from:
"Most step-wise test cases (particularly 2b-d) do not result in the same parameter values as the simultaneous test case 3a in which all the observations are included (Fig. 4a), highlighting that strong non-linearity in the model sensitivity to parameters together with the use of an algorithm that is only adapted to weakly non-linear problems, as well as the assumption of linearity in calculating the posterior error covariance matrix at the minimum of the cost function, can result in differences between a step-wise and simultaneous approach in multiple – data stream assimilation (see Section 1)."
To:
"Most step-wise test cases (particularly 2b-d) do not result in the same parameter values as the simultaneous test case 3a in which all the observations are included (Fig. 4a). This highlights that strong non-linearity in the model sensitivity to parameters, together with the use of an algorithm that is only adapted to weakly non-linear problems, can result in

differences between a step-wise and simultaneous approach in multiple – data stream assimilation (see Section 1)."

>>

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

>> Changed, thank you.

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.

****

(NEW ADVICE SECTION 4 – remains in Section 4 after the toy model experiments)
4   Perspectives and advice for Land Surface Modellers
Although it is clear that in many cases, increasing the number of different observations in a model optimisation provides additional constraints, challenges remain that need to be addressed. Many of these issues that we have discusssed are relevant to any data assimilation study, including those only using one data stream. However, most are more pertinent when considering more than one source of data. Based on the simple toy model results presented in here, in addition to lessons learned from existing studies, we recommend the following points when carrying out multiple data stream carbon cycle data assimilation experiments:

- If technical constraints require that a step-wise approach be used, it is preferable (from a mathematical standpoint) to propagate the full parameter error covariance matrix between each step. Furthermore, it is important to check that the order of assimilation of observations does not affect the final posterior parameter values, and that the fit to the observations included in the previous steps is not degraded after the final step (e.g. Peylin et al., 2016).
- 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, though note that this is not possible in a step-wise assimilation.
- The presence of a bias in a data stream, or an incompatibility between the observations and the model, will hinder the use of multiple observation types in an assimilation framework. Therefore it is imperative to analyse and correct for biases in the observationsand to determine if there is an incompatibility between the model and data. Alternatively, it may be possible account for any possible bias/inconsistency in the observation error covariance matrix, **R**, using the off-diagonal terms or inflated errors (Chevallier, 2007), or by using the prior model-data RMSE to define the observation uncertainty.
- Most optimisation studies with a large-scale LSM require the use of derivative-based algorithms based on a least-squares formulation of the cost function, and therefore rely on assumptions of Gaussian error distributions and quasi model linearity. However, if the these assumptions are not met it may not be possible to find the true global minimum of the cost function and the characterisation of the posterior probability distribution will be incorrect. This is a particular problem if the posterior parameter error covariance matrix is then propagated in a step-wise approach, although these issues are relevant to both step-wise and simultaneous assimilation. Therefore it is important to assess the non-linearity of your model, and if the model is strongly non-linear, use a global search algorithms for the optimisation – although at the resolution of typical LSM simulations ($\geq 0.5 \times 0.5°$) this will likely only be computationally feasible at site or multi-site scale.

In addition to the above points, we have investigated the impact of a difference in the number of observations in each data stream in this study. Test case 3b, in which only one observation was included for the s_2 data stream instead of the complete time-series, shows that a substantial difference in number of observations between the data streams can influence the resulting parameter values and posterior uncertainty (compare test cases 3a and b in Fig. 2 for the simple C model and Fig. 4 for the non-linear toy model) as each data stream will have a different overall "weight" in the cost function. Xu et al. (2006), among others, have mentioned the possible need to weight the cost function for different data sets. Different arguments abound on this issue. Some contend that the cost function should not be weighted by the number of observations because the error covariance matrices (**B** and **R**) already define this weight in an objective way (e.g. Keenan et al., 2013), and we would agree with this assertion. It should not be necessary to weight by the number of observations in the cost function if there is sufficient information to properly build the prior error covariance matrices (**B** and **R**).

It is always useful to investigate the issues such as those discussed here by setting up synthetic experiments, as in this study, to understand the possible constraint brought by different data streams, and the impact of a possible bias and observation or observation–model inconsistency. Note also that performing a number of tests starting from different random "first guess" points in parameter space can help to diagnose if the global minimum has been reached, as outlined in Section 2.1.6 and discussed at the beginning of the results (Section 3.2). Furthermore, several diagnostic tests exist to help infer the relative level of constraint brought about by different data streams, including the

observation influence and degrees of freedom of signal metrics (Cardinali et al., 2004). Performing these tests was beyond the scope of this study, particularly given that the simple toy models contained so few parameters, but such tests may be instructive when optimising many hundreds of parameters in a large-scale LSM with a number of different data streams.

Aside from multiple data stream assimilation, other promising directions could also be considered to help constrain the problem of lack of information in resolving the parameter space within a data assimilation framework, including the use of other ecological and dynamical "rules" that limit the optimisation (see for example Bloom and Williams, 2015), or the addition of different timescales of information extracted from the data such as annual sums (e.g. Keenan et al., 2012). Finally we should also seek to develop collaborations with researchers in other fields who may have advanced further in a particular direction. Members of the atmospheric and hydrological modelling communities, for example, have implemented techniques for inferring the properties of the prior error covariance matrices, including the mean and variance, but also potential biases, autocorrelation and heteroscedasticity, by including these terms as "hyper-parameters" within the inversion (e.g. Michalak et al. 2005; Evin et al., 2014; Renard et al., 2010; Wu et al. 2013;). Of course this extends the parameter space – making the problem harder to solve unless sufficient prior information is available (Renard et al., 2010), but such avenues are worth exploring.

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.

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 exam- ple 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:

• 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."

• 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"

*• **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 s_1 for each time step was correlated to the noise added to s_2. 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."*

*• **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.

• 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."*

• 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).

Consistent assimilation of multiple data streams in a carbon cycle data assimilation system

Natasha MacBean[1], Philippe Peylin[1], Frédéric Chevallier[1], Marko Scholze[2], Gregor Schürmann[3]

[1]{Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France}

[2]{Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden}

[3]{Max Planck Institute for Biogeochemistry, Jena, Germany}

Correspondence to: N. MacBean (nlmacbean@gmail.com)

Abstract

Data assimilation methods provide a rigorous statistical framework for constraining the parametric uncertainty of land surface models (LSMs), with the aim of improving our predictive capability as well as identifying areas in which the models need improvement. The increase in the number of available datasets in recent years allows us to address different aspects of the model at a variety of spatial and temporal scales. However, combining data streams in a DA system is not a trivial task. In this study we highlight some of the challenges surrounding multiple data stream assimilation for the carbon cycle component of LSMs. We give particular consideration to the assumptions associated to the type of inversion algorithm typically used for optimising the parameters of global LSMs – namely, Gaussian error distributions and linearity in the model dynamics. We explore the effect of biases and inconsistencies between the observations and the model (resulting in non Gaussian error distributions), and we examine the difference between performing a simultaneous assimilation (in which all data streams are included in one optimisation) and a step-wise approach (in which each data stream is assimilated sequentially) in the presence of non-linear model dynamics. 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

1

included in the prior observation error covariance matrix, and when the correlated errors are ignored. We demonstrate some of these issues by assimilating synthetic observations into two simple models: the first a simplified version of the carbon cycle processes represented in many LSMs, and the second a non-linear toy model. Finally we provide some perspectives and advice to other land surface modellers wishing to use multiple data streams to constrain their models.

Keywords: data assimilation, carbon cycle, biogeochemical cycles, land surface model.

1   Introduction

The carbon cycle is an important component of the Earth system, especially when considering the climatic impact of rising greenhouse gas concentrations from fossil fuel emissions and land use change. It is estimated that the oceans and land surface absorb approximately half of the CO_2 emissions due to anthropogenic activity, but uncertainties remain in the strength and location of sources and sinks, as well as in predictions of future trends (Ciais et al., 2013). Observations allow us to understand the system up until the present day and provide inference about how ecosystems may respond to future change. However, their use in estimating model state variables and boundary conditions is limited beyond diagnostic purposes, and they can be restricted in their spatial coverage. They also do not contain all the information we may need to distinguish between the complex interactions that may occur between many different processes. Incorporating our current knowledge of physical mechanisms of biogeochemical cycles, including carbon, C, dynamics, into land surface models (LSMs) represents a promising approach to analyse these interacting effects, to upscale observations to larger regions, and to make future predictions. However, the models can be limited by the lack of process representation, either due to gaps in our knowledge, or in our technical and computing capability. As a result, model evaluations reveal that not all variables are well-captured by the model under current conditions (Anav et al., 2013), and the spread between model projections is still very large (Sitch et al., 2015).

Aside from model structural and forcing errors, one source of uncertainty is related to the parameter (i.e. fixed) values of a model. Model-data fusion, or data assimilation (DA), allows the calibration, or optimisation, of these values by minimising a cost function (that quantifies the model-data misfit) while accounting for the uncertainties inherent in both the model and data in a statistically rigorous framework. The C cycle component of most LSMs is complex

2

and contains a large number of parameters; luckily however, there are an increasing number of in-situ and remote sensing-based data streams that can be used for parameter optimisation. These data bring information on different spatial and temporal scales, such as:

- Atmospheric CO_2 concentration data measured at surface stations at continental to global scales, which provide information from synoptic timescales to inter-annual variability (IAV) and long-term trends.

- Eddy covariance net CO_2 (net ecosystem exchange – NEE) and latent (LE) and sensible heat fluxes measured at half-hourly intervals at many sites across different ecosystems/regions, providing information at seasonal to inter-annual timescales.

- Satellite-derived measures of vegetation dynamics, including "greenness" indices (i.e. the Normalised Difference Vegetation Index – NDVI), fraction of absorbed photosynthetically active radiation (FAPAR) and leaf area index (LAI) at global scales and at daily time steps spanning more than a decade, thus capturing IAV and long-term trends (though usually with a trade-off between spatial and temporal resolution).

- Satellite-derived measurements of soil moisture and land surface temperature at the same temporal and spatial scales as the satellite-derived observations of vegetation productivity.

- Aboveground biomass measurements are currently taken at only one or a few points in time at plot scale up to regional scale from aircraft and satellite data, or are estimated from allometric relationships at each site.

- Soil C stock estimates usually are only taken at one point in time at plot scale.

- Ancillary data on vegetation characteristics such as tree height or budburst – only measured at certain well-instrumented sites.

Increasingly, researchers are attempting to bring these sources of data together to constrain different parts of a model at different spatio-temporal scales within a multiple data stream assimilation framework (e.g. Richardson et al., 2010; Keenan et al., 2012; Kaminski et al., 2012; Forkel et al., 2014; Bacour et al., 2015). However, whilst the potential benefit of adding in extra data streams to constrain the C cycle of LSMs is clear, multiple data stream assimilation is not as simple as it may seem. This is particularly true when considering a

3

regional-to-global scale, multiple data stream, multiple site optimisation of a complex LSM that contains many parameters, and which typically takes on the order of minutes to an hour to run a one year simulation. When using more than one data stream there is the option to include all data streams together in the same optimisation (simultaneous approach), or to take a sequential (step-wise) approach. Mathematically, the optimal approach is the simultaneous, but computational constraints related to the inversion of large matrices or the requirement of numerous simulations, especially for global datasets (e.g Peylin et al., 2016), and/or the weight of different data streams in the optimisation (e.g. Wutzler and Carvalhais, 2014), may complicate a simultaneous optimisation. On the other hand, in a step-wise assimilation the parameter error covariance matrix has to be propagated at each step, which implies that it can be computed. If the parameter error covariance matrix can be properly estimated and is propagated between each step, the step-wise approach should be mathematically equal to simultaneous. However, many inversion algorithms (e.g. derivative-based methods that use the gradient of the cost function to find its minimum) require assumptions of model (quasi-)linearity and Gaussian parameter and observation error distributions (Tarantola, 1987, p195). If these assumptions are violated, or the error distributions are poorly defined, it is likely that the step-wise will not be equal to the simultaneous, because information will be lost at each step due to an incorrect calculation of the posterior error covariance matrix at the end of the first step. An incorrect description of the observation (– model) error distribution could result from i) the wrong assumption about the distribution of the residuals between the observation and the model, ii) a poor characterisation of the error correlations, iii) an incompatibility between the model and the data (possibly due to a model structural issue or differences in how a variable is characterised), or iv) a bias in the observations that is not unaccounted for (i.e. is treated as a random error). As mentioned, whilst a simultaneous optimisation is mathematically more rigorous in the sense that the error correlations are treated within the same inversion, if the prior distributions are not properly characterised any bias may be aliased to the wrong parameters (Wutzler and Carvalhais, 2014), more so than in a step-wise approach.

This tutorial-style paper highlights some of the challenges of multiple data stream optimisation of carbon cycle models discussed above. Note that we do not aim to explore all possible issues related to a DA system, for example the choice of the cost function, minimization algorithm, or the characterization of the prior error distributions; indeed previous studies have investigated such aspects at length (e.g. Fox et al., 2009; Trudinger et

4

al., 2007), therefore we refer the reader to these papers for more information. Section 2 reviews recent carbon cycle multiple data stream assimilation studies with reference to some of the aforementioned challenges. Section 3 demonstrates the issues related to multiple data stream assimilation with synthetic experiments designed around two simple models: one a simplified version of the carbon dynamics included in many LSMs, and the other a "toy" model designed to demonstrate the issues that arise with complex, non-linear models. Finally Section 4 provides some advice to land surface modellers wishing to carry out multiple data stream assimilation.

2   Review of existing multiple data stream carbon cycle data assimilation studies

2.1   Extra constraint from multiple data streams

Most site-based carbon cycle data assimilation studies have used eddy covariance measurements of NEE and LE fluxes to constrain the relevant parameters of ecosystem models. However, a few studies have also made use of chamber flux soil respiration data and field measurements of vegetation characteristics (e.g. tree height, budburst, LAI) or estimates of litterfall and carbon stocks as ancillary information (e.g. Fox et al., 2009; Keenan et al., 2012; Thum et al., in review; Van Oijen et al., 2005; Richardson et al., 2010; Williams et al., 2005). Two recent studies combined high-resolution satellite-derived FAPAR data and in-situ eddy covariance measurements to optimize parameters related to carbon, water and energy cycles of the ORCHIDEE and BETHY LSMs at a couple of sites (Bacour et al., 2015; Kato et al., 2013, respectively).

At global scales the number of studies that use multiple data streams from satellites or large-scale networks to optimise LSMs has been increasing in recent years, although this remains a relatively new area of research. CCDAS-BETHY was the first global carbon cycle data assimilation system (CCDAS) making use of the high-precision measurements of the atmospheric CO_2 concentration flask sampling network (Rayner et al., 2005; Scholze, 2003) to constrain process parameters of the prognostic terrestrial carbon cycle model BETHY (Knorr, 2000). Since its first application in which only atmospheric CO_2 concentration data were assimilated, CCDAS-BETHY has been further developed to consistently assimilate

5

multiple data streams both at local and global scales. In particular, Kaminski et al. (2012) optimised 70 process parameters plus one initial condition by simultaneously assimilating a satellite-derived FAPAR product derived from the Medium Resolution Imaging Spectrometer (MERIS; Gobron et al., 2008) and flask samples of atmospheric CO_2 at two sites from the GLOBALVIEW product (GLOBALVIEW-CO2, 2008) at coarse resolution. More recently, Scholze et al. (2016) demonstrated the added value of assimilating remotely sensed soil moisture data in addition to observations of atmospheric CO_2 concentration from the flask-sampling network. They used the same coarse resolution set-up of CCDAS as Kaminski et al. (2012) and CO_2 observations from 10 sites of the GLOBALVIEW product (GLOBALVIEW-CO2, 2012) together with the SMOS L3 daily soil moisture product (version 246; CATDS-L3, 2012).

Two other global CCDAS based on different LSMs have been developed in recent years (Peylin et al., 2016; Schürmann et al., 2016). Schürmann et al. (2016) optimized model parameters and initial conditions of the land component JSBACH (Raddatz et al. 2007) of the MPI Earth System Model (ESM) (Giorgetta et al. 2013) using atmospheric CO_2 concentration data and the TIP-FAPAR product (Pinty et al., 2007) as joint constraints over a 5 year period, in addition to evaluating the mutual benefit of each data stream in a fully factorial design. Peylin et al. (2016) used three different data streams as global constraints for the ORCHIDEE LSM (Krinner et al., 2005), which forms the land surface component of the IPSL ESM (Dufresne et al., 2013), in a multi-site step-wise assimilation approach. First, satellite-derived vegetation index data (NDVI) from the MODIS instrument were used to constrain the phenology parameters at 60 sites for the temperate and boreal deciduous PFTs, followed by NEE and LE observations at 78 FLUXNET sites for 7 PFTs to optimise all the carbon-related parameters, and finally atmospheric CO_2 concentration measurements from 53 sites in the GLOBALVIEW network (GLOBALVIEW-CO2, 2013), which predominantly provided a constraint on the initial magnitude of the soil carbon reserves in the model. These three global multiple data stream CCDAS have allowed an improvement in both the mean seasonal cycle as well as the trend of net land surface CO_2 exchange, especially with the inclusion of atmospheric CO_2 data (Kaminski et al., 2012; Peylin et al., 2016; Schürmann et al., 2016). Atmospheric CO_2 concentration observations are one of the most accurate, long-term data sets in environmental science and they provide important information about the global CO_2 sink capacity by land and ocean.

Natasha MacBean 23/5/2016 16:44
Moved down [1]: Atmospheric CO_2 concentration observations are one of the most accurate, long-term data sets in environmental science and they provide important information about the global CO_2 sink capacity by land and ocean.

Natasha MacBean 23/5/2016 16:44
Moved (insertion) [1]

Many of the aforementioned studies reported that adding extra data streams helped to constrain unresolved sub-spaces of the total parameter space. Scholze et al. (2016) found that simultaneously assimilating SMOS soil moisture data with CO_2 observations reduced the ambiguity in the solution space when assimilating CO_2 only, and the multiple data constraint was able to resolve a much larger sub-space in parameter space (about 30 parameters out of the 101 compared to 15 without SMOS data). Bacour et al. (2015) and Schürmann et al. (2016) both reported that the addition of FAPAR data bought extra information on the phenology-related processes in the model, and therefore retrieved different posterior C flux-related parameter values than when assimilating NEE or atmospheric CO_2 data alone. An interesting aspect of the Kaminski et al. (2012) study was that the inclusion of FAPAR in addition to atmospheric CO_2 concentration samples resulted in a particular improvement for the hydrological fluxes in the model, thus demonstrating the importance of assessing the potential benefit for model variables that may not have been the main target of optimisation.

Richardson et al. (2010) and Keenan et al. (2012) concluded that using ancillary information (e.g. woody biomass increment, field-based LAI and chamber measurements of soil respiration) as in addition to NEE data provided a valuable extra constraint on many model parameters, which improved both the bias in model predictions and reduced the associated uncertainties. The results of the REFLEX model-data fusion inter-comparison project also indicated that observations of the different carbon pools would help to constrain parameters such as root allocation and woody turnover that were not well resolved using NEE and LAI data alone (Fox et al., 2009). Thum et al. (in review) found that the addition of aboveground biomass stocks brought a longer-term constraint on allocation parameters and mortality/turnover processes. However, they noted an incompatibility when assimilating both annual increment and total biomass data, as the total stocks take into account losses related to disturbance and management (e.g. canopy thinning) – processes that were not included in that version of the model. On the other hand, Williams et al. (2005) observed that one-off, or rarely taken, measurements of carbon stocks were unable to constrain components of the carbon cycle to which they were not directly related. This raises the issue of the relative influence of different data streams in a joint assimilation, especially if the number of observations for each is vastly different, which will be the case when assimilating both half-hourly C flux data in addition to soil C stock observations that are typically available at an annual time scale. The spatial distribution of each data stream is also important, especially for heterogeneous landscapes (Barrett et al., 2005; Alton, 2013).

7

1 Although a number of multiple data stream assimilation studies exist at various scales,
2 very few studies have specifically investigated the added benefit of different combinations of
3 data streams in a factorial study, with a few notable exceptions (Barrett et al., 2005;
4 Richardson et al., 2010; Kato et al., 2013; Keenan et al., 2013; Bacour et al., 2015;
5 Schürmann et al., 2016). Kato et al. (2013) and Bacour et al. (2015) both evaluated the
6 complementarity of eddy covariance and FAPAR data streams at site level, i.e. the impact of
7 assimilating one individual data stream on the other model state variable, as well as when
8 both data streams were included in the optimization (see discussion in Section 3.2). The study
9 of Keenan et al. (2013) was particularly notable in its aim to quantify which data streams
10 provide the most information (in terms of model-data mismatch) and how many data streams
11 are actually needed to constrain the problem. They reported that of the 17 field-based data
12 streams available, projections of future carbon dynamics were well-constrained with only 5 of
13 the data sources, and crucially, not with eddy covariance NEE measurements alone. These
14 results may be specific to this site or type of ecosystem, but their study highlights the need for
15 further research in this area, and in particular, for synthetic data experiments that allow us to
16 understand which data will be the most useful for a given scientific question. This will also
17 enable researchers to plan more efficient measurement campaigns with experimentalists, as
18 also pointed out by Keenan et al. (2012).

19

20 **2.2 Issue of bias and inconsistencies between the observations and the**
21 **model**

22 Despite the theoretical benefit of adding data streams into an assimilation system as
23 additional constraints, several of the aforementioned studies at both site and global scale have
24 reported a bias or inconsistency either between the different observation data streams, or
25 between the observations and the model. This is easily detected when the optimisation of one
26 data stream results in a worse fit than the prior in one or more of the other data streams, as
27 seen in Section 2.2.2. Kato et al. (2013) assimilated SeaWiFS FAPAR (Gobron et al., 2006)
28 and eddy covariance LE measurements at the FLUXNET site in Maun, Botswana. They
29 showed that the individual assimilation of each the two data streams resulted in a perfect (i.e.
30 within the observational uncertainty) fit to the assimilated data set, but a considerable
31 degradation of the fit to the non-assimilated data set compared to the prior. A comparison
32 against eddy covariance measurements of gross carbon uptake (gross primary production –

8

Natasha MacBean 25/5/2016 15:18
Deleted: orthogonal

GPP) pointed to a bias in the FAPAR data because the fit to the independent GPP data was degraded after assimilating FAPAR data only, while the fit improved after assimilating the LE data only. Nevertheless, the simultaneous assimilation of both data streams achieved a compromise between the two suboptimal results achieved after assimilating only one data stream. The calibration further limited the number of parameters with correlated errors, and yielded a higher theoretical reduction in parameter uncertainty and a decrease in the RMS difference by 16% for the GPP data compared to the prior.

Bacour et al. (2015) also noted that when assimilating both in-situ and satellite-derived FAPAR data (from the SPOT and MERIS instruments) and in-situ NEE and LE flux data from two French FLUXNET sites into the ORCHIDEE LSM both separately and together, the posterior parameter values changed significantly for the photosynthesis and phenology-related parameters, depending on the bias between the model and the observations and the correlation between the parameter errors. If NEE data were assimilated alone there was an even stronger positive bias (model–observations) in the start of leaf onset in the FAPAR data than in the prior simulations, and no improvement in the maximum value. This was likely due to the fact that there were enough degrees of freedom to fit the NEE without changing the phenology-related parameters. Similarly, the fit to the NEE was degraded when the model was only optimized with FAPAR data. The model was able to fit the maximum FAPAR but this resulted in an adverse effect on the carbon assimilation capacity of the vegetation. The authors argued this was related to incompatibilities between the FAPAR and both the model and NEE measurements, possibly due to its larger spatial footprint of the satellite-derived FAPAR data and/or inaccuracies in the retrieval algorithm. However, given that assimilating in-situ FAPAR also degraded the fit to the NEE, another culprit may be an inconsistency between the model and the data. The authors suggested this could be due to the different characterisation of FAPAR or LAI in the model compared to what is described in the data. For example, satellite-derived greenness measures (FAPAR/NDVI) also contain information on the non-green elements of vegetation, but the model only simulates green LAI. Furthermore parameters and processes in models have been developed at certain temporal and spatial scales. Vegetation is often simply represented as a "big leaf" model in LSMs, taking no account of vertical canopy structure or the spatial heterogeneity in a scene, which is an additional source of inconsistency with what is measured. The joint (simultaneous) assimilation of all three data streams in Bacour et al. (2015) reconciled the different sources of information, with an improvement in the model-data fit for NEE, LE and FAPAR.

However, the compromise achieved in the joint assimilation was only possible when the FAPAR data were normalised to their maximum and minimum values, which thus partially accounted for any bias in the magnitude of the FAPAR or inconsistency with the model.

The story of biases and apparent inconsistencies in FAPAR data does not end there. A bias correction was also necessary in the study by Kaminski et al. (2012) with CCDAS-BETHY using the MERIS FAPAR product in addition to atmospheric CO_2 data (see above). They found that optimisation procedure failed when using the original FAPAR product because the FAPAR values were biased towards higher values. Only after applying a bias correction on the FAPAR data before assimilation was the optimisation successful. Schürmann et al. (2016) also reported the need to reduce a prior model bias in FAPAR. Even though the assimilation successfully corrected for this FAPAR bias, the impact of the prior bias was evident in the spatial patterns of the modelled heterotrophic respiration. Assimilating FAPAR data alone therefore resulted in a slight degradation in the net C flux and consequently led to incorrect simulations of the atmospheric CO_2 growth rate. The addition of CO_2 as a constraint prevented this degradation and resulted in a compromise in which FAPAR helped to disentangle these processes and find different parameter values compared to the CO_2-only case, thus improving the fit to both data streams. Forkel et al. (2014) discovered an apparent inconsistency between satellite-derived FAPAR and GPP data in tundra regions when using these data (plus satellite-derived albedo) to optimise the LPJmL LSM. They too speculated that the data might be positively biased, in this case due to issues with satellite measurements taken at high sun zenith angles. However, they gave alternative suggestions, one being that an inadequate model structure may be at fault – for example, LPJmL does not include vegetation classes corresponding to shrub, moss and lichen species that are dominant in these ecosystems. They also noted that the GPP product they used, which is based on a model tree ensemble up-scaling of FLUXNET data (Jung et al., 2011), might contain representation-related biases, given that there are very few FLUXNET stations in tundra regions. The issue of representation errors of sites has been touched upon before (e.g. Raupach et al., 2005). Alton (2013), who performed a global multi-site optimisation of the JULES LSM with a diverse range of data including satellite-derived LAI, FLUXNET, soil respiration and global river discharge, raised the point that FLUXNET sites are known to be large carbon sinks, which could potentially result in biased global NEE estimates. Resolving these apparent inconsistencies was beyond the scope of most of these studies, aside from applying a bias correction where one was evident. Aside from simple corrections, Quaife et

10

al. (2008) and Zobitz et al. (2014) suggested that LSMs should be coupled to radiative transfer models to provide a more realistic and mechanistic observation operator between the quantities simulated by the model and the raw radiance measured by satellite instruments. This proposition followed the experience gained in the case of atmospheric models for several decades (Morcrette, 1991).

2.3 Step-wise versus simultaneous assimilation

The paper by Alton (2013) documents the only previous study to have used a step-wise assimilation approach with more than two data streams, stating that the final parameter values were independent of the order of data streams assimilated. No studies in the LSM community to date have explicitly examined a step-wise versus simultaneous assimilation framework with the same optimisation system and model. The step-wise assimilation with the ORCHIDEE-CCDAS detailed in Peylin et al. (2016) has been compared to a simultaneous optimisation using the same three data streams as part of an on-going study. At each step, the resulting simulations (using the posterior parameters) were compared to the data stream from the previous steps. The fit to the MODIS NDVI (used in a similar manner to FAPAR as a proxy for vegetation greenness) was unchanged after further optimization of the phenology-related parameters in the second and third steps using in-situ flux and atmospheric CO_2 concentration data. In the simultaneous optimisation, the addition of NEE or atmospheric CO_2 concentration measurements resulted in a lower improvement to the fit to MODIS NDVI (Bacour et al. submitted?). As the NDVI data were normalised this was not a result of a simple bias in the magnitude of the data. Rather, it was likely due to inconsistencies between the model and data as discussed by Bacour et al. (2015). It is important to reiterate that there should be no difference between the step-wise and the simultaneous given an adequate description of the error covariance matrices and compliance with the assumptions associated with the inversion algorithm used. However, in practice it is very difficult to define a probability distribution that properly characterises the model structural uncertainty and observation errors accounting for biases and non-Gaussian distributions. This leads to issues within a simultaneous assimilation, and a greater risk of differences between a step-wise and simultaneous assimilation. However, as discussed in the introduction a step-wise assimilation may be useful on a provisional basis for dealing with possible inconsistencies. For example in the step-wise approach of Peylin et al. (2016) as the error covariance of the phenology-related

11

parameters was strongly constrained by the satellite data in the first step (and was propagated to the second step), the later optimisations with NEE and atmospheric CO_2 data in steps 2 and 3 found alternative solutions for the C flux-related parameters that provided a reasonable fit to all data streams. Wherever possible however, a simultaneous optimisation is favourable because the strong parameter linkages between different processes are maintained, and therefore biases and inconsistencies between the model and observations should be addressed prior to optimisation.

3   Demonstration with two simple models and synthetic data

The three sub-sections in Section 2 highlight examples within a carbon cycle modeling context of the three main challenges faced when performing a multiple data stream assimilation, namely, i) the possible negative influence of including additional data streams into an optimization on other model variables; ii) the impact of bias in the observations, missing model processes or incompatibility between the observations and with the model (as discussed in Section 2.2), and iii) the difference between a step-wise and simultaneous optimization (and the order of data stream assimilation) if the assumptions of the inversion algorithm are violated, which is more likely to be the case with non-linear models when using derivative-based algorithms and least-squares formulation of the cost function (as discussed in Section 2.3). The latter point is important because derivative methods (compared to global search) are the only viable option for large-scale, complex LSMs given the time taken to run a simulation. 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.

This section aims to demonstrate these challenges using simple toy models and synthetic experiments where the true values of the parameters are known. Thus the following sections include a description of the toy models together with the derivation of synthetic observations, the inversion algorithm used to optimise the model parameters and the experiments performed, followed by the results for each test case.

Natasha MacBean 26/6/2016 23:06
Deleted: assimilations

12

3.1 Methods

3.1.1 Simple carbon model

To demonstrate the challenges of multiple data stream assimilation in a carbon cycle context, we have chosen a test model that represents a simplified version of the carbon cycle dynamics typically implemented in most LSMs. The model has been well-documented in Raupach (2007) and has been used previously in the OptIC DA inter-comparison project (Trudinger et al., 2007). It is based on two equations that describe the temporal evolution (on a daily time step) of two living biomass (carbon) stores, s_1 and s_2, and the biomass fluxes between these two stores:

$$\frac{ds_1}{dt} = F(t)\left(\frac{s_1}{p_1 + s_1}\right)\left(\frac{s_2}{p_2 + s_2}\right) - k_1 s_1 + s_0 \tag{1}$$

$$\frac{ds_2}{dt} = k_1 s_1 - k_2 s_2 \tag{2}$$

In this model formulation, s_1 and s_2 are approximately equivalent to above- and belowground biomass stocks. The unknown parameters p_1, p_2, k_1 and k_2 will be optimised in the inversions. The first term on the right-hand side of Eq. (1) corresponds to the Net Primary Production (NPP) i.e. the carbon input to the system as a function of time, represented by $F(t)$, weighted by factors (the two fractions in parentheses) that account for the size of both pools, in order to introduce a limitation on NPP. The $F(t)$ forcing term is a random function of time ("log-Markovian" random process) representing the effect of fluctuating light and water availability due to climate on the NPP (Raupach, 2007 – Section 5.3). The litterfall is an output of s_1 (aboveground biomass store) and an input to s_2 (belowground biomass store) and is calculated as a constant fraction (k_1) of s_1 (defined by $k_1 s_1$). Heterotrophic respiration (Rh) is a constant fraction (k_2) of the belowground carbon reserve s_2 and is represented $k_2 s_2$. The constant s_0 is a "seed production" term set to 0.01 (i.e. not optimised) to ensure the model does not verge towards zero. A more detailed description of the properties of the model is given in Trudinger et al. (2007 – Section 2.1) and an in-depth analysis of the dynamical behaviour of the model is provided in Raupach (2007). Synthetic observations of both s_1 and s_2 variables were used to optimise all the unknown parameters in the model (see Section 2.1.5).

Natasha MacBean 27/6/2016 13:12
Deleted: pools

Natasha MacBean 27/6/2016 13:11
Deleted: assimilated

Natasha MacBean 27/6/2016 13:11
Deleted: into

Natasha MacBean 27/6/2016 13:19
Deleted: (the two fractions in parentheses)

Natasha MacBean 27/6/2016 13:30
Deleted: the aboveground carbon reserve

Natasha MacBean 24/6/2016 17:18
Deleted: as represented by

Natasha MacBean 27/6/2016 13:16
Deleted: an output of

3.1.2 Non-linear toy model

Although the simple carbon model contains a non-linear term it is essentially still a quasi-linear model. In order to illustrate the challenges associated with multiple data stream data assimilation for more complex non-linear models, especially when using derivative methods, we defined a simple non-linear toy model based on two equations with two unknown parameters:

$$s_1 = a \exp^b + at^2 \tag{3}$$

$$s_2 = \sin(10a + 10b) + 10t^2 \tag{4}$$

where s_1 and s_2 also correspond to two model state variables (as for the simple C model), a and b are the unknown parameters included in the optimisation, and t is the independent variable, which could represent time in a real-world scenario. Note that this model is not based on any particular physical process associated with land surface biogeochemical cycles, but it does contain typical mathematical functions that are observed in reality and implemented in LSMs. For example, the sinusoidal function (Eq. (4)) could represent diurnal variations of various processes such as photosynthesis and respiration. Exponential response functions (such as in Eq. (3)) are also observed for certain processes, including the temperature sensitivity of soil microbial decomposition. As for the simple carbon model, synthetic observations corresponding to the s_1 and s_2 variables were used to optimise both parameters (see Section 2.1.5).

3.1.3 Bayesian inversion algorithm

Most data assimilation approaches follow a Bayesian formalism which, simply put, allows prior knowledge of a system (in this case the model parameters) to be updated, or optimised, based on new information (from the observations). In order to achieve this we define a "cost function" that describes the misfit between the data and the model, taking into account their respective uncertainties, as well as the uncertainty on the prior information. If we follow a Bayesian formalism and least-squares minimisation approach, and assume Gaussian probability distributions for the model parameter and observation error variance/covariance, we derive the following cost-function (Tarantola, 1987):

14

$$J(\mathbf{x}) = \frac{1}{2}[(H(\mathbf{x}) - \mathbf{y})^T.\mathbf{R}^{-1}.(H(\mathbf{x}) - \mathbf{y}) + (\mathbf{x} - \mathbf{x}^b)^T.\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)] \tag{5}$$

where $\mathbf{y}$ is the observation vector, $H(\mathbf{x})$ the model outputs given parameter vector $\mathbf{x}$, $\mathbf{R}$ the observation error covariance matrix (including measurement and model errors), $\mathbf{x}^b$ the a priori parameter values, and $\mathbf{B}$ the prior parameter error covariance matrix. This framework leads to a Gaussian posterior parameter probability distribution function and requires that the model and its observation operator are linear.

The aim of the inversion algorithm is to find the minimum of this cost function, thereby achieving the best possible fit between the model simulations and the measurements, conditioned on their respective uncertainties and prior information. For cases where there is a strong linear dependence of the model to the parameters (at least for variations in $\mathbf{x}$ of the size of those expected in the data assimilation system), and where the dimensions of the problem are not too large, the solution can be derived analytically. If not, as is usually the case with LSMs, there are different numerical methods to find the most optimal parameter values. These include global search methods that randomly search the parameter space and test the likelihood of a particular parameter set at each iteration, and derivative methods, which calculate the gradient of the cost function at each iteration to find its minimum. In this study we use the latter class of methods. More specifically we use a quasi-Newton algorithm that uses both the gradient of the cost function and its derivative (Hessian) to evaluate if the minimum has been reached (i.e. where the gradient is zero). Thus we obtain the following algorithm for iteratively finding the minimum (Tarantola, 1987, p195):

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \varepsilon_i(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}(\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x})) + \mathbf{B}^{-1}(\mathbf{x}_i - \mathbf{x}^b)) \tag{6}$$

where i is the iteration number and $\mathbf{H}$ is the Jacobian, or first-order derivatives, of H, which in this study is determined using a finite difference method. Note that as we are potentially dealing with non-linear models, the quasi-Newton method has been slightly adapted to include the constant scaling factor ε_i (with a value <1.0) to ensure that the algorithm will converge.

Of course no inversion algorithm is perfect, and therefore it is possible that the true "global" minimum of the cost function has not been found. Derivative methods in particular can get stuck in so-called "local minima", preventing the algorithm from finding the true minimum. To address this issue we carry out a number of assimilations with different random

15

first guess points in the parameter space. If they all result in the same reduction in cost function value, we can have more confidence that the true minimum has been found.

Once the minimum of the cost function has been found, the posterior parameter error covariance can be approximated (using the linearity assumption) from the inverse Hessian of the cost function around its minimum, which is calculated using the Jacobian of the model at the minimum of $J(\mathbf{x})$ (for the set of optimized parameters), $\mathbf{H}_\infty$, following Tarantola (1987):

$$\mathbf{A} = [\mathbf{H}_\infty^T \mathbf{R}^{-1} \mathbf{H}_\infty + \mathbf{B}^{-1}]^{-1} \tag{7}$$

Note that the posterior error covariance matrix can be propagated into the model space to determine the posterior uncertainty on the simulated state variables as a result of the parametric uncertainty (as shown in the coloured error bands in the time series plots – Figures 1 and 5) using the following matrix product and the hypothesis of local linearity (Tarantola, 1987):

$$\mathbf{R}_{post} = \mathbf{H}_\infty . \mathbf{A} . \mathbf{H}_\infty^T \tag{8}$$

3.1.4 Step-wise versus simultaneous assimilation

Step-wise approach

In the step-wise approach each data stream (in our cases s_1 and s_2, see above) is assimilated sequentially, and the posterior error covariance matrix of Eq. (7) is propagated to the next step as the prior in Eq. (6). Note that the error covariance matrix can only be propagated if it is calculated within the inversion algorithm, which is the case here but may not be possible in other studies. The following details an example for two data streams.

Step 1: Assimilation of the first data stream, s_1. The prior parameters, including their values and error covariance ($\mathbf{x}^b$ and $\mathbf{B}$), are optimised to produce a first set of posterior optimised parameters $\mathbf{x}_1$ with error covariance $\mathbf{A}_1$.

Step 2: Assimilation the second data stream, s_2. The parameters, $\mathbf{x}_1$, and their error covariance, $\mathbf{A}_1$, are used as a prior to the optimisation system and further optimised to produce the second (and final) set of posterior optimised parameters, $\mathbf{x}_{post}$, and the associated error covariance $\mathbf{A}$.

16

Simultaneous approach

Both data streams s_1 and s_2 are included in the optimisation and all parameters are optimised at the same time. The prior parameters, including their values and error covariance ($\mathbf{x}^b$ and $\mathbf{B}$) are optimised to produce the posterior parameter vector ($\mathbf{x}_{post}$) and associated uncertainties $\mathbf{A}$.

3.1.5 Optimisation set-up: parameter values and uncertainty, and generation of synthetic observations

In this study we used synthetic observations that were generated by running the model with known (or 'true') parameter values and adding random Gaussian noise corresponding to the defined observation error for both s_1 and s_2 (see Table 1). We optimised a ten-year time window for the simple carbon model, in order to capture the dynamics of the s_1 and s_2 pools over a time period compatible with typically available observations. For the non-linear toy model, which did not correspond to physical processes in the terrestrial biosphere, we ran a simulation over a window of 100 integrations (steps) of the equations. The observation frequency was daily, corresponding to the time-step of the simple carbon model (a value of 1 for the non-linear toy model), and the observation error was set to 10% of the mean value for each set of pseudo-observations derived from multiple first guesses of the model.

The true values of all parameters for both models are given in Table 1, together with their upper and lower bounds (following Trudinger et al., 2007). We have not performed a prior sensitivity analysis to decide to which parameters are important to include in the optimisation, as the model variables are sensitive to all of the (small set of) parameters. However, in the case of a more complex, large-scale LSM it is advisable to carry out such an analysis, particularly given the computational burden of optimising many parameters. In this study the parameter uncertainty (1 sigma) was set to 40% of the parameter range following recent studies (e.g. Bacour et al., 2015). Prior values were chosen from a uniform random distribution bounded by the parameter bounds.

3.1.6 Experiments

The specific objective of the following experiments is to test the impact of a bias in the observations that is not accounted for in the R matrix, and the impact of using derivative

17

Natasha MacBean 24/8/2016 15:42
Deleted: We assumed independence (i.e. uncorrelated errors) for both the parameters and observation covariance matrices, thus the **R** and **B** matrices were diagonal.

Natasha MacBean 24/6/2016 20:42
Formatted: Normal, Indent: First line: 0,62 cm, No bullets or numbering

methods with non-linear models (as may be necessary with large-scale LSMs), particularly with reference the differences that may arise between step-wise and simultaneous optimisations.

Table 2 details the experiments that were carried out based on all possible combinations for assimilating the two data streams. Three approaches were compared: i) separate – where only one data stream was included in the optimisation; ii) step-wise – where each data stream was assimilated sequentially; and ii) simultaneous – where both data streams were included in the optimisation. All parameters for both models were optimised in all experiments, therefore in the step-wise cases the parameters were optimised twice. Tests for the step-wise were also carried out with and without the propagation of the full posterior parameter error covariance matrix, $\mathbf{A}_1$, in between steps 1 and 2 (test cases 2b and d – see Table 2) – i.e. for these tests only the posterior variance was propagated. An additional test was included for the simultaneous assimilation in order to test the impact of having a substantial difference in the number of observations for the data stream included in the optimisation, as may be the case for belowground (e.g. soil) biomass observations in reality. Therefore in test case 3b, only one observation was included for data stream s_2.

The differences in the parameter values and the theoretical reduction in their uncertainty $(1 - (\sigma_{post} / \sigma_{prior}))$ were examined for all eight test cases, as well as the fit (RMSE) to both data streams after the optimisation. For the step-wise approach we investigated if the fit to the first data stream is degraded in the second step by comparing the RMSE after each step. Note that the reduction in uncertainty is a theoretical, or approximate, estimate of the real uncertainty reduction because of the assumptions made in the inversion scheme.

In a second stage the impact of an unknown, un-accounted for bias in the model was examined. This bias could be a systematic bias in the observations due to the algorithm used for their derivation, the result of missing or incomplete processes in the model, or an incompatibility between the observations and the model, for example due to differences in spatial resolution or an inconsistent characterisation of a variable between the model and the observations. To test the impact of such an occurrence, we introduced a constant scalar bias into the modelled s_2 variable with a value of 10 (i.e. twice the magnitude of the defined observation uncertainty). All eight experiments were repeated, but a bias was introduced into the model calculation of s_2 that was not accounted for in the cost function (i.e. the error distributions retained a mean of zero). This was treated as an unknown bias, and therefore not

18

corrected or accounted for in the inversion scheme and the defined observation uncertainty (Table 1) was not changed for this set of experiments.

In all experiments for both models twenty assimilations were performed starting from different random "first guess" points in the parameter space. As discussed in Section 2.1.3 this was done to test the ability of the algorithm to converge to the true global minimum of the cost function. Note that the global minimum and possible reduction in $J(x)$ will be different for each experiment, as each is based on a different cost function.

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 s_1 for each time step was correlated to the noise added to s_2. 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.

3.2 Results

The twenty random first guess assimilations were examined for each set of experiments for both models (before the results for each test were examined in more detail), in order to check that the algorithm converged to a global minimum. As shown in the supplementary information (Fig. S1), a high proportion of the twenty first guess assimilations across all test cases for both models resulted in a similar reduction in $J(x)$, even though the overall

19

1   magnitude of the reduction was sometimes different between tests. This indicates that the

2   algorithm does not easily get stuck in any local minima (if they exist). The examples shown in

3   the results below were taken from one first guess parameter set for each model that belonged

4   to the cluster that had the highest cost function reduction. Any differences seen in the

5   parameter values, their posterior uncertainty or the resultant RMSE reduction described below

6   therefore are due to the specific details of each test and not the inability of the algorithm to

7   find the minimum.

8

9   3.2.1 Typical performance with a quasi-linear model and no bias

10   Figures 1a and b show the simple carbon model simulations for test case 3a (in which

11   both data streams are assimilated simultaneously) for the s_1 and s_2 variables. A large reduction

12   in RMSE is achieved after optimisation (blue curve) with respect to the observations (black

13   curve). Overall, there is a good reduction in RMSE for all test cases (including the individual

14   assimilations 1a and 1b) with a reduction of ~80% for s_1 and s_2. In addition, the optimisation

15   of the s_1 and s_2 variables resulted in a good or moderate reduction in RMSE for variables not

16   included in any assimilation: ~60% for the litterfall (Eqn. 1) and ~16% for the heterotrophic

17   respiration (Rh – Eqn. 2) across all test cases (not shown), although there was already a good

18   prior fit to the data. As would be expected from these results, the parameter values and the

19   theoretical reduction in parameter uncertainty do not vary between the tests (Figures 2a and b

20   blue symbols), except for a slight difference in the value of the k_2 parameter in test cases 1a

21   and 3b, for which there is also a lower reduction in uncertainty (~82% compared to >95%).

22   Note that Fig. 2a shows the normalised parameter values to account for differences in the

23   magnitude of the different parameters and their range (the zero line represents the "true"

24   parameter value – see caption). In this situation therefore, where we have a relatively simple

25   linear model and two data streams to which the model parameters are highly sensitive, we see

26   that the differences between the step-wise and simultaneous approaches are minimal. This is

27   even the case when the error covariance is not propagated between the two steps (test cases 2b

28   and d), suggesting that under this assimilation set-up with this model both s_1 and s_2

29   individually contain enough spatio-temporal information to retrieve the true values of all

30   parameters, as we can see from the separate test cases 1a and b. However, we cannot

31   definitively say whether this is due to the simplicity or relative linearity of the model – it is

possible that observations of variables in more complex linear model would not be able to retrieve the true values of all parameters.

3.2.2 Impact of unknown bias in one data stream – example with a simple carbon model

In Section 2.2.1 we saw that there is little difference between a step-wise and simultaneous optimisation if there is no bias in the model or observations, and if the model is quasi-linear and therefore the critical assumptions behind the inversion approach were not violated. However, it is not uncommon to have a bias between your observations and model that is not obvious and therefore not accounted for in the optimisation, as the cost function used in most inversion algorithms (and in this study) assume Gaussian error distributions with zero mean. Note that this is also the case when defining a likelihood function for accepting or rejecting parameter values in a global search method. To test the impact of a bias, we added a constant value to the simulated s_2 variable in a second test (see Section 3.1.6) that was treated as an unknown bias, and therefore not corrected or accounted for in the inversion scheme. The impact of this bias on s_1 and s_2 is shown in Figures 1c-d, and the reduction in RMSE between the model and observations is seen in Fig. 3 for all variables (including Rh and litterfall). The red symbols in Fig. 2 show the resultant parameter values and theoretical reduction in uncertainty as a result of the bias. The inversion cannot accurately find the correct values for all parameters in any test case and there are now considerable differences between the simultaneous and step-wise approach. Furthermore the order in which the data streams are assimilated in the step-wise cases also results in different posterior parameter values (test cases 2a and b versus 2c and d in Fig. 2a and Fig. 3). Nevertheless the optimisation results in a similar reduction in uncertainty on the parameters, except in test case 1b where only s_2 data are assimilated (Fig. 2b).

The main impact of the bias in the modelled s_2 variable is on the value of k_2 parameter (Fig. 2a), which is consistently offset from the true value (dashed line in Fig. 2a) in all test cases. This was expected given that it is the parameter most directly related to the calculation of s_2. However, in test cases 2a and 3a, the values of p_1 and p_2 are also incorrect (and p_1 for test case 2b). Note that these parameters only indirectly influence the s_2 pool in the model, and therefore we might have expected that they would be less affected by the bias. This nicely

21

Natasha MacBean 27/6/2016 13:43
Deleted: uniform

Natasha MacBean 27/6/2016 13:43
Deleted: 2

demonstrates one issue that could arise in all DA studies, where the bias in a particular variable (in the observations or the model) is aliased onto another process in the model (Wutzler and Carvalhais, 2014). Such an aliasing of bias onto indirectly related parameters is even more evident when only s_2 is included in the assimilation and s_1 does not provide any constraint (test case 1b) – in this case all parameters are incorrect but the p_2 parameter in particular shows a strong deviation from the true value (Fig. 2a). As a result we see a deterioration in the RMSE for the s_1, litterfall and Rh variables in test case 1b and in the step-wise cases where s_2 is assimilated in the second step (Figures 3a, c and d – test case 1b, 2a and 2b). However, the RMSE reduction remains high for the s_2 variable for these test cases (Fig. 3b), as the inversion has found a solution that accounts for the bias even though all inferred parameter values are incorrect. The assimilation of s_1 in the second step lowers the reduction in RMSE for s_2 gained in the first step to ~70%, but it is not a considerable degradation.

Even though the posterior parameter values are incorrect, and despite the fact that the first step results in a degradation, the final reductions in RMSE are largely the same as the situation with no bias for all variables when s_1 is included in a simultaneous assimilation or optimised in the second step (test cases 2c, d and 3a in Fig. 3). This shows that the inclusion of s_1 observations can find a solution to counter the bias in s_2 and prevents a degradation in the fit to the data. If s_2 is assimilated in the second step there is a negative impact on all other variables as discussed above, demonstrating again that the order of data stream assimilation can matter when there are biases or inconsistencies between the data and the model.

The analysis of the impact of the bias presented here is specific to this model and the type and magnitude of the bias that was added, but the broader findings can be generalised to any situation in which there is a bias or inconsistency between a model and data that is not accounted for in the assigned error distributions. Exactly what might constitute a bias or inconsistency is discussed more in Section 3.2. Also note that it is important to examine the impact on the other variables. For the separate test case 1b in which only s_2 data are used to optimise the model, the negative impact on the other variables (Fig. 3) would have been concealed if we had only examined the posterior reduction in RMSE for the s_2 variable. Again this is a concern that is inherent to all DA experiments, whether single- or multi-data stream, but we can see from these results (i.e. by comparing the separate test cases 1b with 2a and b)

Natasha MacBean 27/6/2016 13:46
Deleted: than

that adding another data stream in a multi-constraint approach does not always reduce the problem.

3.2.3 Difference between the step-wise and simultaneous approaches in the presence of a non-linear model

As discussed in Section 2.2.1, there is little difference between the step-wise and the simultaneous assimilation approaches for simple, relatively linear models, unless the observation error (including measurement and model errors) distribution deviates strongly from the Gaussian assumption. However in reality, large-scale, complex LSMs may contain highly non-linear responses to certain model parameters. To demonstrate the impact of non-linearity in a multiple data stream assimilation context, we used a non-physically based toy model chosen for its non-linear characteristics (see Section 2.1.2).

Fig. 4a shows the posterior parameter values for both the a and b parameters of the non-linear toy model for all test cases. The values were not normalised as both parameters have the same range. The horizontal dashed line shows the "true" known values of the parameters (both equal to 1.0) that were used to generate the synthetic observations. Note that no bias has been introduced into the model in the results described here. The prior and posterior model s_1 and s_2 simulations for the non-linear toy model are compared to the synthetic observations in Fig. 5 for both step-wise cases in which the posterior error covariance matrix from step 1 ($\mathbf{A}_1$ – see section 2.1.4) was propagated to step 2 (experiments 2a and c – Fig. 5a-d) and both simultaneous cases 3a and b (Fig. 5 e-h). Finally Fig. 6 summarises the reduction in RMSE between the simulated and observed s_1 and s_2 variables for the non-linear toy model for all test cases and, in the step-wise cases, the reduction in RMSE after both the first and second steps (light versus dark green bars).

Assimilating each data stream individually (test cases 1a and b) does not result in an accurate retrieval of the posterior parameters (Fig. 4a), nor in a strong constraint on either parameter, as shown by the lack of theoretical reduction in the parameter uncertainty after the optimisation (Fig. 4b). Despite this, there is a 91-92% reduction in RMSE for the data stream that was included in the optimisation (i.e. for s_1 in test case 1a – Fig. 6a, and s_2 in test case 1b – Fig. 6b). However, the improvement on the other data stream is much less (28% reduction in RMSE for s_1 when s_2 is assimilated) or even results in a degradation compared to the prior

23

1 fit (e.g. in the case of s_2 when s_1 is assimilated – Fig. 6b). Lack of improvement, or even

2 degradation, in the RMSE of other variables in the model is a common issue for data

3 assimilation in general – one that is not often evaluated in model-data fusion studies. It is also

4 is not necessarily the result of a bias or incompatibility between the observations and the

5 model.

6 Only the simultaneous case, in which all s_1 observations have been included in the cost

7 function (test case 3a), manages to retrieve the correct parameter values after the optimisation.

8 All other posterior parameter values are incorrect, and are considerably different between

9 each case, unlike for the simple carbon model (without a model bias). Most step-wise test

10 cases (particularly 2b-d) do not result in the same parameter values as the simultaneous test

11 case 3a in which all the observations are included (Fig. 4a). This highlights that strong non-

12 linearity in the model sensitivity to parameters, together with the use of an algorithm that is

13 only adapted to weakly non-linear problems, can result in differences between a step-wise and

14 simultaneous approach in multiple – data stream assimilation (see Section 1).

15 In the simultaneous optimisation in which all observations are included (test case 3a)

16 the posterior fit to the data dramatically improves for both the s_1 and s_2 data streams after the

17 assimilation (blue dashed line in Fig. 5e and f). This was expected given that the correct

18 values of the parameters were found. For the step-wise cases (test case 2a in Figures 5a and b,

19 and test case 2c in Fig. 5c and d), the black dashed line shows the prior, and the posterior after

20 step 1 is shown by green dashed line. In the step-wise assimilation we see two different

21 scenarios depending on which data stream was assimilated first. In the first step the results are

22 the same as the case where each individual data stream is assimilated separately. In both cases

23 the first step results in a good fit to the data that was included in the optimisation in that step.

24 When the s_1 data was assimilated in the first step (Fig. 5 first row), the fit to s_2 deteriorated

25 after the optimisation (Fig. 5b green dashed line and Fig. 6b – test case 2a_s1), but when the

26 s_2 data were assimilated first (Fig. 5 second row) the optimisation step did manage to achieve

27 an improvement in the s_1 data stream (Fig. 5c green dashed line and Fig. 6a – test case 2c_s1).

28 In the second step the optimisation of s_2 in test cases 2a and b does not degrade the fit

29 to s_1 when the full parameter error covariance matrix ($\mathbf{A}_1$) is propagated between step 1 and 2

30 (Figures 5a blue curve and 6a 2a_s2). Furthermore optimising s_2 in the second step reverses

31 the deterioration in s_2 caused by assimilating s_1 in the first step (Figures 5b blue curve and 6b

32 2a and b dark green bars). However, when s_1 data were assimilated in the second step (test

24

Natasha MacBean 27/6/2016 13:49
Deleted: ,

Natasha MacBean 27/6/2016 13:49
Deleted: highlighting

Natasha MacBean 27/6/2016 13:49
Deleted: , as well as the assumption of linearity in calculating the posterior error covariance matrix at the minimum of the cost function

cases 2c and d), we found that the good fit achieved with s_2 observations in the first step was effectively reversed (Fig. 5d blue curve). Therefore assimilating s_1 in the second step degraded the fit to the s_2 observations, even compared to the prior case (Fig. 6b, dark green bars for test cases 2c and d). This nicely highlights one of the main possible issues with a step-wise assimilation framework.

The fact that the final reduction in RMSE values after both steps was ~90% for most cases, even though the values were not correct for all but case 3a (Fig. 4), indicates that the error correlation between the two parameters (~ -1.0 – calculated from the posterior error covariance matrix but not shown) led to alternative sets of values that resulted in a similar improvement to the data – a phenomenon known as model equifinality.

3.2.4 Order of assimilation of data streams and propagation of parameter error covariance matrices in a step-wise approach

Comparing the step-wise cases 2a and b with 2c and d for the non-linear toy model reveals that neither order in the assimilation, s_1 then s_2, or s_2 then s_1, results in the correct posterior parameter values that match the simultaneous test case (Fig. 4a). This is not a result that can be generalised to all step-wise assimilations as it will depend on the data stream involved and whether they contain enough spatio-temporal information to accurately constrain all the parameters included in the optimisation, as well as any biases in the model or observations (as discussed in Section 3.2.2) or model non-linearity (section 3.2.3). In the case of the non-linear toy model, neither s_1 nor s_2 find the right parameter values when assimilated individually, therefore it is not surprising that neither order manages to achieve the right posterior parameter values. Nevertheless, the theoretical uncertainty of both parameters is reduced by >95% for the step-wise cases in which $\mathbf{A}_1$ from step 1 is propagated between step 1 and 2 (test cases 2a and c – Fig. 4b), even though the posterior values for the step-wise cases are incorrect. This demonstrates that a good theoretical reduction in uncertainty is not always indicative that the right parameters have been found by the optimisation. The lower theoretical reduction in parametric uncertainty for cases 2b and d (Fig. 4b) demonstrates that information is lost between the steps if the posterior error covariance terms of $\mathbf{A}_1$ after step 1 are not propagated to step 2, and therefore cannot be used to further constrain the optimisation.

25

From a mathematical standpoint the most rigorous approach is to propagate the full parameter error covariance matrices between each step. Without that constraint not only is information lost in the second step, but the information contained in the second data stream may have a stronger influence compared to a simultaneous or step-wise case with a propagated error covariance matrix. The inversion may therefore be more vulnerable to any strong biases or incompatibilities between the model and the observations of the second data stream, or indeed the particular sensitivity of its corresponding model state variable to the parameters. This is one possible explanation for the degradation seen in s_1 in the non-linear toy model when s_2 is optimised in the second step and $\mathbf{A}_1$ is not propagated between the steps (Fig. 6a test case 2b_s2). The same was also true for the simple carbon model for test case 2b when a bias was introduced into the s_2 simulation (see Section 3.2.2 and Fig. 3a).

However, the reverse is also true – if the first data stream contains strong biases then the associated error correlations will be also propagated with $\mathbf{A}_1$. If autocorrelation in the observation errors, or indeed correlation between the errors of the data streams, is not accounted for, it is likely that the posterior simulations are over-tuned, i.e. we will overestimate the reduction in parameter uncertainty. If this is the case and the first step results in incorrect parameter values, the propagation of $\mathbf{A}_1$ could restrict the parameter values to the wrong location in the parameter space and thus inhibit the ability of the inversion to find the correct global minimum. These issues are likely to be more considerable for non-linear models, as seen by the lack of difference between test cases 2a-d in the simple carbon model example (Fig. 2).

3.2.5 Impact of accounting for correlated observation errors in the prior observation error covariance matrix

In a final test we introduced time invariant correlated noise between the two data streams (see Section 3.1.6). We investigated the impact of ignoring cross-correlation between two data streams by comparing the results of i) an optimisation in which the correlated errors were included in the off-diagonal elements of the prior observation error covariance matrix, $\mathbf{R}$, to ii) an optimisation in which the correlated observation errors were excluded (i.e. R was kept diagonal). Note that this experiment is only relevant to simultaneous multiple data stream assimilation, as it is not possible to account for cross-correlation between data streams when

26

one is assimilated after the other in a step-wise approach. The presence of error correlations increases redundancy in the inversion, which would therefore reduce the expected theoretical error reduction compared to uncorrelated observations (experiments not shown). We would expect a further limitation on the expected error reduction with a sub-optimal system, as represented by optimisation ii) in which there was cross-correlation between the data streams, but the correlated observation errors were ignored (as seen in Chevallier, 2007).

Figure 7 shows the difference between the two optimisation, (i.e. including off-diagonal minus diagonal elements in the **R** matrix), for the reduction in the cost function value (Figures 7a and d) and posterior s_1 and s_2 observation errors (1 sigma – Figures 7b, c, e and f), for both the simple C model (top row) and the non-linear toy model (bottom row). The plot shows the median difference across all twenty random first guess parameters, and the reduction is calculated as 1 – (posterior/prior).

At low observation error there is no discernible difference between accounting for the correlated observation errors in the **R** matrix or not. This is likely because there is enough information in the observations to find the global minimum of the cost function. Trudinger et al. (2007) also found that similar posterior values were obtained when comparing observations with correlated and uncorrelated Gaussian errors. However, at a certain point with increasing observation error increases along the x-axis (i.e. decreasing information content) there is a difference in the cost function and parameter error reduction between the two optimisations for both models (Figure 7). The sub-optimal set-up in optimisation ii) is clearly be seen as the optimisation with off-diagonal correlated errors in R results in a higher reduction in the cost function and posterior observation error (blue values). This pattern fits our expectation, as detailed above. Furthermore we see a pattern emerging suggesting that the difference between the two optimisations increases with higher observation correlation for the same error magnitude. However, for some combinations of observation error and correlation, the pattern is opposite to what we expect (red values), particularly for the s_1 data stream in the simple C model (Figure 7b). This is likely because the accuracy of the solution becomes limited by observation uncertainty at higher observation errors, and also due to presence of model non-linearity, which prevents a fully accurate characterisation of the posterior error covariance matrix.

The key finding of this preliminary investigation into correlated observation errors is that it becomes increasingly important to properly characterise and account for correlations

between data streams if the observations do not contain enough information (i.e. too few observations or high observation uncertainty). However, this is a wide topic that has received little-to-no attention in the carbon cycle data assimilation literature to date, aside from the 2 out of 21 experiments in the wider-ranging study of Trudinger et al. (2007). We therefore suggest that an investigation such as this should be extended in order to fully understand the impact of cross-correlation between data streams; however, this is beyond the scope of this paper.

4   Perspectives and advice for Land Surface Modellers

Although it is clear that in many cases, increasing the number of different observations in a model optimisation provides additional constraints, challenges remain that need to be addressed. Many of the issues that we have investigated are relevant to any data assimilation study, including those only using one data stream. However, most are more pertinent when considering more than one source of data. Based on the simple toy model results presented here, in addition to lessons learned from existing studies, we recommend the following points when carrying out multiple data stream carbon cycle data assimilation experiments:

- If technical constraints require that a step-wise approach be used, it is preferable (from a mathematical standpoint) to propagate the full parameter error covariance matrix between each step. Furthermore, it is important to check that the order of assimilation of observations does not affect the final posterior parameter values, and that the fit to the observations included in the previous steps is not degraded after the final step (e.g. Peylin et al., 2016).

- 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.

Natasha MacBean 27/6/2016 00:30
Deleted: Lessons to be learned when dealing with non-linearity . ... [2]

Natasha MacBean 26/6/2016 16:44
Deleted: , orthogonal

Natasha MacBean 26/6/2016 22:28
Deleted: should not be ignored

Natasha MacBean 26/6/2016 23:20
Moved (insertion) [2]

Natasha MacBean 28/6/2016 14:30
Deleted: se

Natasha MacBean 28/6/2016 14:25
Deleted: in this study

- The presence of a bias in a data stream, or an incompatibility between the observations and the model, will hinder the use of multiple observation types in an assimilation framework. Therefore it is imperative to analyse and correct for biases in the observationsand to determine if there is an incompatibility between the model and data. Alternatively, it may be possible account for any possible bias/inconsistency in the observation error covariance matrix, **R**, using the off-diagonal terms or inflated errors (Chevallier, 2007), or by using the prior model-data RMSE to define the observation uncertainty.

- Most optimisation studies with a large-scale LSM require the use of derivative-based algorithms based on a least-squares formulation of the cost function, and therefore rely on assumptions of Gaussian error distributions and quasi model linearity. However, if the these assumptions are not met it may not be possible to find the true global minimum of the cost function and the characterisation of the posterior probability distribution will be incorrect. This is a particular problem if the posterior parameter error covariance matrix is then propagated in a step-wise approach, although these issues are relevant to both step-wise and simultaneous assimilation. Therefore it is important to assess the non-linearity of your model, and if the model is strongly non-linear, use a global search algorithms for the optimisation – although at the resolution of typical LSM simulations ($\geq 0.5 \times 0.5°$) this will likely only be computationally feasible at site or multi-site scale.

In addition to the above points, we have investigated the impact of a difference in the number of observations in each data stream in this study. Test case 3b, in which only one observation was included for the s_2 data stream instead of the complete time-series, shows that a substantial difference in number of observations between the data streams can influence the resulting parameter values and posterior uncertainty (compare test cases 3a and b in Fig. 2 for the simple C model and Fig. 4 for the non-linear toy model) as each data stream will have a different overall "weight" in the cost function. Xu et al. (2006), among others, have mentioned the possible need to weight the cost function for different data sets. Different arguments abound on this issue. Some contend that the cost function should not be weighted by the number of observations because the error covariance matrices (**B** and **R**) already define this weight in an objective way (e.g. Keenan et al., 2013), and we would agree with this

29

Natasha MacBean 26/6/2016 17:01
Deleted: Devote time to characterising the error structure for the observations and parameter error distributions, including their correlations (Raupach et al., 2005). For the observations this should include the model structural errors (Kuppel et al. 2013), the temporal or spatial autocorrelation and correlation between different data streams.

Natasha MacBean 26/6/2016 16:58
Deleted: ¶
Prior information is key in a Bayesian Bayesian framework. Effort should be put into better constraining the prior parameter bounds of all parameters based on literature wherever possible. ¶ ... [3]

Natasha MacBean 26/6/2016 17:52
Deleted: strongly so

Natasha MacBean 26/6/2016 23:20
Deleted: avoid a least squares formulation of the cost function or

Natasha MacBean 26/6/2016 16:57
Deleted: Prior information is key in a Bayesian framework. Effort should be put into better constraining the prior parameter bounds of all parameters based on literature wherever possible. ¶ ... [4]

assertion. It should not be necessary to weight by the number of observations in the cost function if there is sufficient information to properly build the prior error covariance matrices (**B** and **R**).

It is always useful to investigate the issues such as those discussed here by performing synthetic experiments with pseudo observations, as in this study, to understand the possible constraint brought by different data streams, and the impact of a possible bias and observation or observation–model inconsistency. Note also that performing a number of tests starting from different random "first guess" points in parameter space can help to diagnose if the global minimum has been reached, as outlined in Section 3.1.6 and discussed at the beginning of the results (Section 3.2). Furthermore, several diagnostic tests exist to help infer the relative level of constraint brought about by different data streams, including the observation influence and degrees of freedom of signal metrics (Cardinali et al., 2004). Performing these tests was beyond the scope of this study, particularly given that the simple toy models contained so few parameters, but such tests may be instructive when optimising many hundreds of parameters in a large-scale LSM with a number of different data streams.

Aside from multiple data stream assimilation, other promising directions could also be considered to help constrain the problem of lack of information in resolving the parameter space within a data assimilation framework, including the use of other ecological and dynamical "rules" that limit the optimisation (see for example Bloom and Williams, 2015), or the addition of different timescales of information extracted from the data such as annual sums (e.g. Keenan et al., 2012). Finally we should also seek to develop collaborations with researchers in other fields who may have advanced further in a particular direction. Members of the atmospheric and hydrological modelling communities, for example, have implemented techniques for inferring the properties of the prior error covariance matrices, including the mean and variance, but also potential biases, autocorrelation and heteroscedasticity, by including these terms as "hyper-parameters" within the inversion (e.g. Michalak et al. 2005; Evin et al., 2014; Renard et al., 2010; Wu et al. 2013). Of course this extends the parameter space – making the problem harder to solve unless sufficient prior information is available (Renard et al., 2010), but such avenues are worth exploring.

Natasha MacBean 26/6/2016 17:06

Deleted: If technical constraints require a step-wise approach is used it is preferable (from a mathematical standpoint) to propagate the full parameter error covariance matrix between each step, if it can be calculated, and carefully consider the order of the assimilation of data streams (a synthetic experiment will aid in this regard).
Many of these issues are relevant to any data relevant to any data assimilation study, including those only using one data stream. However, most are more pertinent when considering more than one source of data.

Natasha MacBean 26/6/2016 23:20

Moved up [2]: Many of these issues are relevant to any data assimilation study, including those only using one data stream. However, most are more pertinent when considering more than one source of data.

Natasha MacBean 27/6/2016 00:56

Deleted: The impact of bias in the magnitude of satellite-derived FAPAR data has featured highly in past multiple data stream assimilation studies. Aside from simple corrections, Quaife et al. (2008) and Zobitz et al. (2014) suggested that LSMs should be coupled to radiative transfer models to provide a more realistic and mechanistic observation operator between the quantities simulated by the model and the raw radiance measured by satellite instruments. This proposition followed the experience gained in the case of atmospheric models for several decades (Morcrette, 1991).

Natasha MacBean 26/6/2016 17:16

Deleted: Of course, optimising the parameters of the model will not account for all the uncertainty in a model. Inaccurate or incomplete process representation is likely a key factor that may also bias the posterior values retrieved in any optimisation. Keenan et al. (2012) reflected that despite using multiple different constraints and different time increments in the cost function, the inter-annual variability and long-term trend of carbon uptake at Harvard forest FLUXNET site in the USA could not be reproduced without a temporal variation of the parameters, suggesting a missing process in the model. However, as this paper shows, the complexities of model-data fusion require that we continue to develop DA techniques alongside development of LSMs, with the hope of converging upon more reliable and accurate predictions of the global C budget in the near future.

5 Conclusions

In this study we have attempted to highlight and discuss some of the challenges associated with using multiple data streams to constrain the parameters of LSMs, with a particular focus on the carbon cycle. We demonstrated some of the issues using two simple models constrained with synthetic observations for which the 'true' parameters are known. We performed a variety of tests in Section 2 to demonstrate the differences between assimilating each data stream separately, sequentially (in a step-wise approach) and together in the same assimilation (simultaneous approach). In particular we focused on difficulties that may arise in the presence of biases or inconsistencies between the data and the model, as well as non-linearity in the model equations. In Section 3 we discussed the experimental results with reference to similar difficulties that have been documented in recent C cycle assimilation studies.

Many of the issues faced are inherent to all optimisation experiments, including those in which only one data stream is used. It is of upmost importance to determine if the observations contain biases, and/or if inconsistencies or incompatibilities exist between the model and the observations, and to correct for this or properly account for this in the error covariance matrices. 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. Secondly it is crucial to understand the assumptions and limitations related to the inversion algorithm used. Without these two points being met, there is a greater risk of obtaining incorrect parameter values, which may not be obvious by examining the posterior uncertainty and model-data RMSE reduction. Furthermore it is more likely that the implementation of a step-wise versus simultaneous approach will lead to different results.

This study was not able to examine an exhaustive list of all possible challenges that may be faced when assimilating multiple data streams, but we hope that this tutorial style paper will serve as a guide for those wishing to optimise the parameters of LSMs using the variety of C cycle related observations that are available today. Furthermore we hope that by increasing awareness about the possible difficulties of model-data integration, we can further bring the modelling and experimental communities together to work more closely on these issues.

31

Code availability

The model and inversion code will be made available via the ORCHIDAS website (upon registration): https://orchidas.lsce.ipsl.fr/multi_data_stream.php.

32

References

Alton, P. B.: From site-level to global simulation: Reconciling carbon, water and energy fluxes over different spatial scales using a process-based ecophysiological land-surface model, Agric. For. Meteorol., 176, 111–124, doi:10.1016/j.agrformet.2013.03.010, 2013.

Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R. and Zhu, Z.: Evaluating the land and ocean components of the global carbon cycle in the CMIP5 earth system models, J. Clim., 26(18), 6801–6843, doi:10.1175/JCLI-D-12-00417.1, 2013.

Bacour, C., Peylin, P., MacBean, N., Rayner, P. J., Delage, F., Chevallier, F., Weiss, M., Demarty, J., Santaren, D., Baret, F., Berveiller, D., Dufrêne, E. and Prunet, P.: Joint assimilation of eddy covariance flux measurements and FAPAR products over temperate forests within a process-oriented biosphere model, J. Geophys. Res. Biogeosciences, 120, 1839–1857, doi:10.1002/2015JG002966.Received, 2015.

Barrett, D. J., Michael J Hill, I., Hutley, L. B., Beringer, J., Xu, J. H., Cook, G. D., Carter, J. O. and Williams, R. J.: Prospects for improving savanna biophysical models by using multiple-constraints model-data assimilation methods, Aust. J. Bot., 53(7), 689–714, doi:10.1071/BT04139, 2005.

Bloom, A. A. and Williams, M.: Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological 'common sense' in a model–data fusion framework, Biogeosciences, 12(5), 1299–1315, doi:10.5194/bg-12-1299-2015, 2015.

Cardinali, C., S. Pezzulli, E. Andersson (2004), Influence-matrix diagnostic of a data assimilation system, Q. J. R. Meteorol. Soc.,130: 2767–2786, doi: 10.1256/qj.03.205

Chevallier, F., 2007: Impact of correlated observation errors on inverted CO2 surface fluxes from OCO measurements, Geophys. Res. Lett., 34, L24804, doi:10.1029/2007GL030463.

Dufresne, J. L., Foujols, M. a., Denvil, S., Caubel, a., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, a., Cugnet, D., de Noblet, N., Duvel, J. P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J. Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, a., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, a., Lefebvre, M. P., Lefevre, F., Levy,

33

C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N. and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5., 2013.

Evin, G., Thyer, M., Kavetski, D., McInerney, D. and Kuczera, G.: Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity, Water Resour. Res., 50(3), 2350–2375, doi:10.1002/2013WR014185, 2014.

Forkel, M., Carvalhais, N., Schaphoff, S., v. Bloh, W., Migliavacca, M., Thurner, M. and Thonicke, K.: Identifying environmental controls on vegetation greenness phenology through model-data integration, Biogeosciences, 11, 7025–7050, doi:10.5194/bg-11-7025-2014, 2014.

Gobron, N., Pinty, B., Aussedat, O., Chen, J. M., Cohen, W. B., Fensholt, R., Gond, V., Huemmrich, K. F., Lavergne, T., Mélin, F., Privette, J. L., Sandholt, I., Taberner, M., Turner, D. P., Verstraete, M. M. and Widlowski, J. L.: Evaluation of fraction of absorbed photosynthetically active radiation products for different canopy radiation transfer regimes: Methodology and results using Joint Research Center products derived from SeaWiFS against ground-based estimations, J. Geophys. Res., 111, D13110, doi:10.1029/2005JD006511, 2006.

Gobron, N., Pinty, B., Aussedat, O., Taberner, M., Faber, O., Mélin, F., Lavergne, T., Robustelli, M. and Snoeij, P.: Uncertainty estimates for the FAPAR operational products derived from MERIS - Impact of top-of-atmosphere radiance uncertainties and validation with field data, Remote Sens. Environ., 112(4), 1871–1883, doi:10.1016/j.rse.2007.09.011, 2008.

Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R. and Mathieu, P. P.: Consistent assimilation of MERIS FAPAR and atmospheric CO2 into a terrestrial vegetation model and interactive mission benefit analysis, Biogeosciences, 9(8), 3173–3184, doi:10.5194/bg-9-3173-2012, 2012.

Kato, T., Knorr, W., Scholze, M., Veenendaal, E., Kaminski, T., Kattge, J. and Gobron, N.: Simultaneous assimilation of satellite and eddy covariance data for improving terrestrial water and carbon simulations at a semi-arid woodland site in Botswana, Biogeosciences, 10(2), 789–802, doi:10.5194/bg-10-789-2013, 2013.

Keenan, T. F., Davidson, E., Moffat, A. M., Munger, W. and Richardson, A. D.: Using

model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling, Glob. Chang. Biol., 18(8), 2555–2569, doi:10.1111/j.1365-2486.2012.02684.x, 2012.

Keenan, T. F., Davidson, E. a., Munger, J. W. and Richardson, A. D.: Rate my data: Quantifying the value of ecological data for the development of models of the terrestrial carbon cycle, Ecol. Appl., 23(1), 273–286, doi:10.1890/12-0747.1, 2013.

Knorr, W.: Annual and interannual CO_2 exchanges of the terrestrial biosphere: process-based simulations and uncertainties, Glob. Ecol. Biogeogr., 9(3), 225–252, doi:10.1046/j.1365-2699.2000.00159.x, 2000.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S. and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cycles, 19(1), 1–33, doi:10.1029/2003GB002199, 2005.

Kuppel, S., F. Chevallier and P. Peylin,: Quantifying the model structural error in Carbon Cycle Data Assimilation Systems. Geosci. Model Dev., 6, 45-55, doi:10.5194/gmd-6-45-2013, 2013.

Michalak, A. M., Hirsch, A., Bruhwiler, L., Gurney, K. R., Peters, W. and co-authors: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions. J. Geophys. Res. 110, D24107. DOI: 10.1029/2005JD005970, 2005.

Morcrette, J.-J.: Evaluation of Model-generated Cloudiness: Satellite-observed and Model-generated Diurnal Variability of Brightness Temperature. *Mon. Wea. Rev.*, **119**, 1205–1224, 1991.

van Oijen, M., Rougier, J. and Smith, R.: Bayesian calibration of process-based forest models: bridging the gap between models and data, Tree Physiol., 25(7), 915–927, doi:10.1093/treephys/25.7.915, 2005.

Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P. J., Kuppel, S., Koffi, E. N., Kane, A., Maignan, F., Chevallier, F., Ciais, P., and Prunet, P.: A new step-wise Carbon Cycle Data Assimilation System using multiple data streams to constrain the simulated land surface carbon cycle, Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-13, in review, 2016.

35

Quaife, T., Lewis, P., De Kauwe, M., Williams, M., Law, B. E., Disney, M. and Bowyer, P.: Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter, Remote Sens. Environ., 112(4), 1347–1364, doi:10.1016/j.rse.2007.05.020, 2008.

Raupach, M. R.: Dynamics of resource production and utilisation in two-component biosphere-human and terrestrial carbon systems, Hydrol. Earth Syst. Sci., 11, 875–889, doi:10.5194/hess-11-875-2007, 2007.

Raupach, M. R., Rayner, P. J., Barrett, D. J., Defries, R. S., Heimann, M., Ojima, D. S., Quegan, S. and Schmullius, C. C.: Model-data synthesis in terrestrial carbon observation: Methods, data requirements and data uncertainty specifications, Glob. Chang. Biol., 11(3), 378–397, doi:10.1111/j.1365-2486.2005.00917.x, 2005.

Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R. and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), , 19, doi:10.1029/2004GB002254, 2005.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46(5), 1–22, doi:10.1029/2009WR008328, 2010.

Richardson, A. D., Williams, M., Hollinger, D. Y., Moore, D. J. P., Dail, D. B., Davidson, E. a., Scott, N. a., Evans, R. S., Hughes, H., Lee, J. T., Rodrigues, C. and Savage, K.: Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints, Oecologia, 164(1), 25–40, doi:10.1007/s00442-010-1628-y, 2010.

Schürmann, G. J., Kaminski, T., Köstler, C., Carvalhais, N., Voßbeck, M., Kattge, J., Giering, R., Rödenbeck, C., Heimann, M., and Zaehle, S.: Constraining a land surface model with multiple observations by application of the MPI-Carbon Cycle Data Assimilation System, Geosci. Model Dev. Discuss., doi:10.5194/gmd-2015-263, in review, 2016.

Scholze, M., T. Kaminski, W. Knorr, S. Blessing, M. Vossbeck, J.p. Grant, and K. Scipal. "Simultaneous Assimilation of SMOS Soil Moisture and Atmospheric CO2 In-situ Observations to Constrain the Global Terrestrial Carbon Cycle." *Remote Sensing of Environment* 180, 334-45, 2016.

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., Levis, S., Levy, P. E., Lomas, M., Poulter, B., Viovy, N., Zaehle, S., Zeng, N., Arneth, A., Bonan, G., Bopp, L., Canadell, J. G.,

36

1   Chevallier, F., Ciais, P., Ellis, R., Gloor, M., Peylin, P., Piao, S., Le Quéré, C., Smith, B.,
2   Zhu, Z. and Myneni, R.: Recent trends and drivers of regional sources and sinks of carbon
3   dioxide, Biogeosciences, 12, 653–679, doi:10.5194/bgd-12-653-2015, 2015.

4   Thum, T., N. MacBean, P. Peylin, C. Bacour, D. Santaren, B. Longdoz, D. Loustau and P.
5   Ciais, The potential benefit of using forest biomass data in addition to carbon and water flux
6   measurements to constrain ecosystem model parameters: case studies at two temperate forest
7   sites. In revision for Agric. For. Meteorol.

8   Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Park, B., Reichstein, M.,
9   Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett,
10 D. and Nikolova, S.: OptIC project: An intercomparison of optimization techniques for
11 parameter estimation in terrestrial biogeochemical models, J. Geophys. Res. Biogeosciences,
12 112(2), doi:10.1029/2006JG000367, 2007.

13 Williams, M., Schwarz, P. a, Law, B. E., Irvine, J. and Kurpius, M. R.: An improved analysis
14 of forest carbon dynamics using data assimilation, Glob. Chang. Biol., 11(1), 89–105,
15 doi:10.1111/j.1365-2486.2004.00891.x, 2005.

16 Wu, L., M. Bocquet, F. Chevallier, T. Lauvaux, and K. Davis, 2013: Hyperparameter
17 estimation for uncertainty quantification in mesoscale carbon dioxide inversions. Tellus B, 65,
18 doi:10.3402/tellusb.v65i0.20894.

19 Xu, T., White, L., Hui, D. and Luo, Y.: Probabilistic inversion of a terrestrial ecosystem
20 model: Analysis of uncertainty in parameter estimation and model prediction, Global
21 Biogeochem. Cycles, 20(2), 1–15, doi:10.1029/2005GB002468, 2006.

22 Zobitz, J. M., Moore, D. J. P., Quaife, T., Braswell, B. H., Bergeson, A., Anthony, J. a. and
23 Monson, R. K.: Joint data assimilation of satellite reflectance and net ecosystem exchange
24 data constrains ecosystem carbon fluxes at a high-elevation subalpine forest, Agric. For.
25 Meteorol., 195-196, 73–88, doi:10.1016/j.agrformet.2014.04.011, 2014.

26
27
28
29
30

Table 1: The optimisation set-up for both models, including the true parameter values, their range and the observation uncertainty (1 sigma), which was set to 10% of the mean value for each set of pseudo-observations derived from multiple first guesses of the model. The parameter uncertainty (1 sigma) was set to 40% of the range for each parameter.

Model	Parameter value (range)				Observation uncertainty	
Simple carbon	p_1	p_2	k_1	k_2	s_1	s_2
model	1 (0.5,5)	1 (0.5,5)	0.2 (0.03,0.9)	0.1 (0.01,0.12)	0.5	5
Non-linear	a		b		s_1	s_2
toy model	1 (0,2)		1 (0,2)		0.5	0.5

38

1   Table 2: List of experiments performed for both models with synthetic data. All parameters

2   are optimised in all cases (therefore in both steps for the step-wise approach).

3

Test case	Step 1	Step 2	Parameter error covariance terms propagated in step 2?
Separate			
1a	s_1	-	-
1b	s_2	-	-
Step-wise			
2a	s_1	s_2	yes
2b	s_1	s_2	no
2c	s_2	s_1	yes
2d	s_2	s_1	no
Simultaneous			
3a	s_1 and s_2	-	-
3b	s_1 and only 1 obs for s_2	-	-

4

5

6

7

8

9

10

11

12

13

14
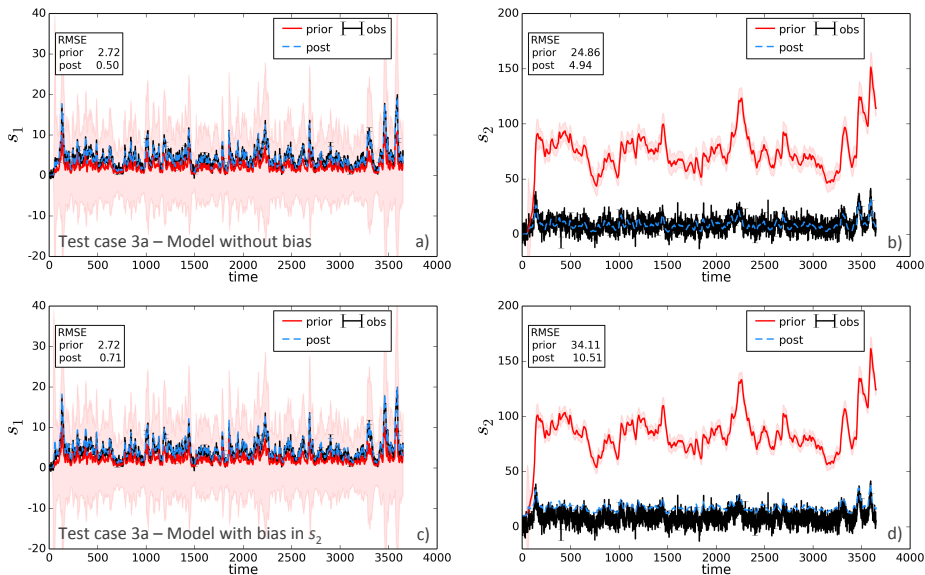
Figure 1: Prior and posterior model simulations compared to the synthetic observations for the simple carbon model for test case 3a for a) s_1 and b) s_2 simulations without any model bias, and c and d) with bias in the simulated s_2 variable. The coloured error band on the prior and posterior represents the propagated parameter uncertainty (1 sigma) on the model state variables (in the equivalent colour as the mean curve). This is mostly visible for the prior model simulation (pink band) as there is a high reduction in model uncertainty reduction as a result of the assimilation.
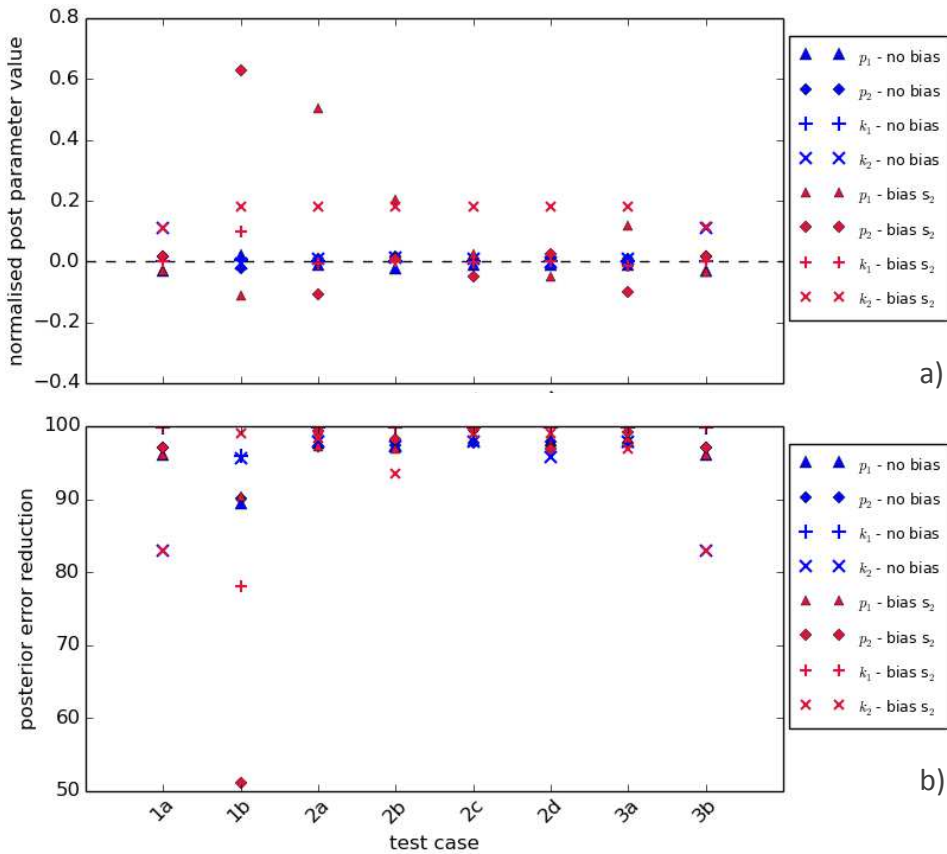
1

2   Figure 2: a) Normalised posterior parameter values and b) posterior parameter error reduction

3   for all parameters of the simple carbon model for each test case, and for both the simulations

4   with no bias (blue) and simulations with a bias in the s_2 variable that was not accounted for in

5   the inversion (red). In a) parameters values were normalised to account for differences in the

6   magnitude of the different parameters and their range, thus it is a measure of the distance

7   from the true value as a fraction of the range and is calculated as: (posterior value – true value

8   / max parameter value – minimum parameter value). The closer the value to the zero dashed

9   line represents a better match to the "true" parameter value. To give an indication of the

10   optimisation performance, the following are the normalised first guess parameter values for

11   this particular example test (compare with posterior values in Fig. 2a): p_1 0.09, p_2 0.29, k_1 0.1,

12   k_2 0.15.

13

41
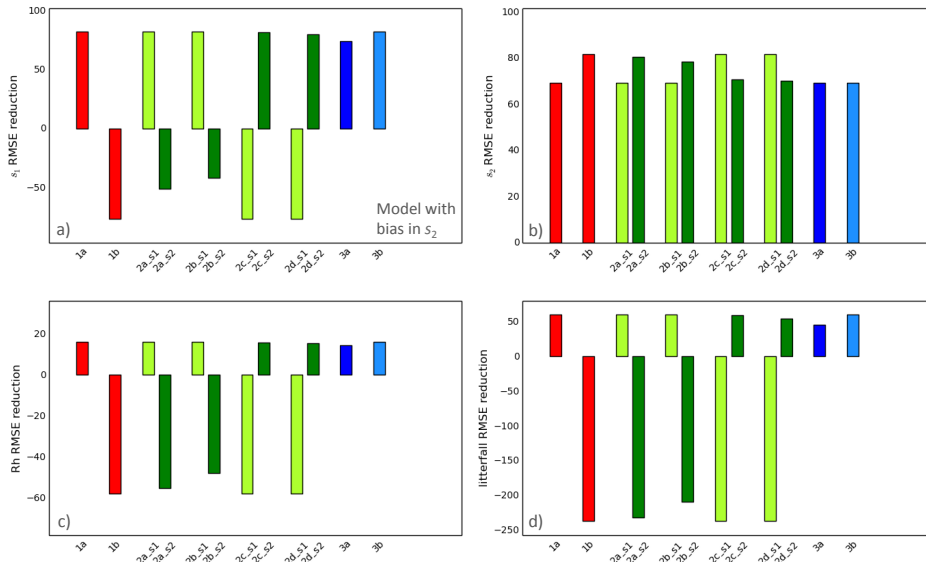
Figure 3: Reduction in RMSE for all test cases for simulations with a bias in the s_2 variable: a) s_1, b) s_2, c) litterfall and d) heterotrophic respiration (Rh). For the step-wise cases (2a, b, c and d) the reduction after both step 1 and step 2 are shown in light and dark green respectively, and are denoted in the x-axis labels with '_s1' for step 1 and '_s2' for step 2. The reduction (in %) is calculated as $1 - (RMSE_{post} / RMSE_{prior})$.
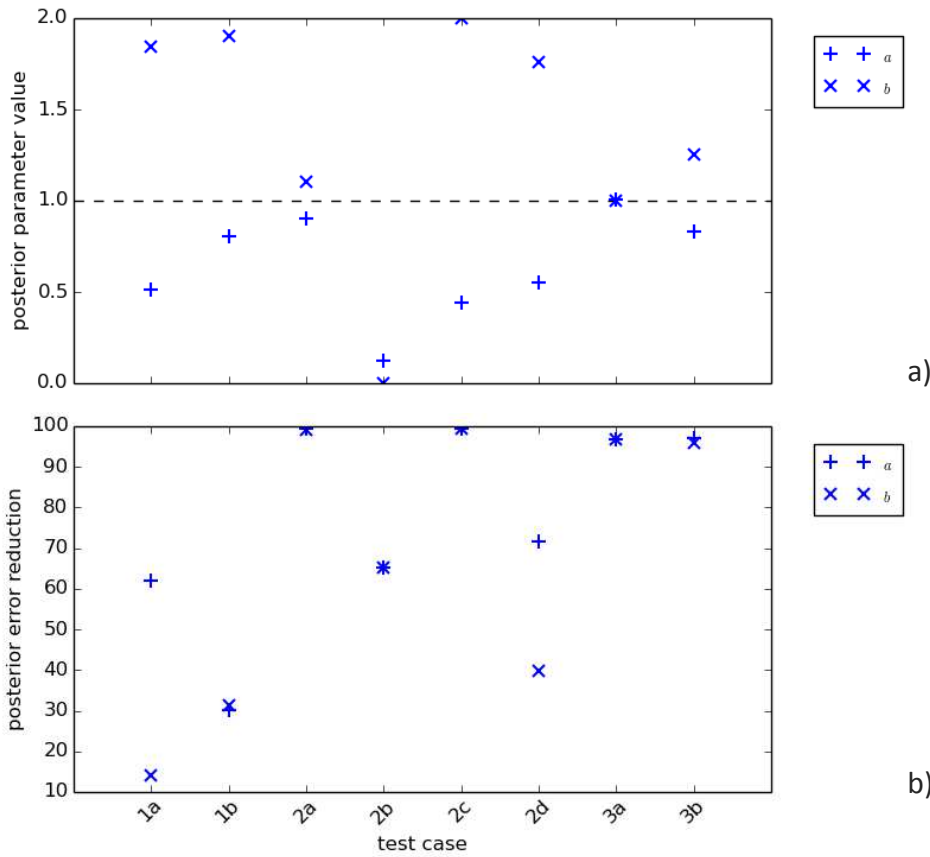
42

Figure 4: Posterior parameter values of both the non-linear toy model *a* and *b* parameters for each test case for the simulations with no model bias. The y-axis range corresponds to the parameter bounds and the dashed horizontal line represents the "true" known value of both parameters. To give an indication of the optimisation performance, the following are the first guess parameter values for this particular example test (compare with posterior values in Fig. 4a): *a* 0.87, *b* 1.98. b) Posterior uncertainty reduction for both parameters for all test cases.
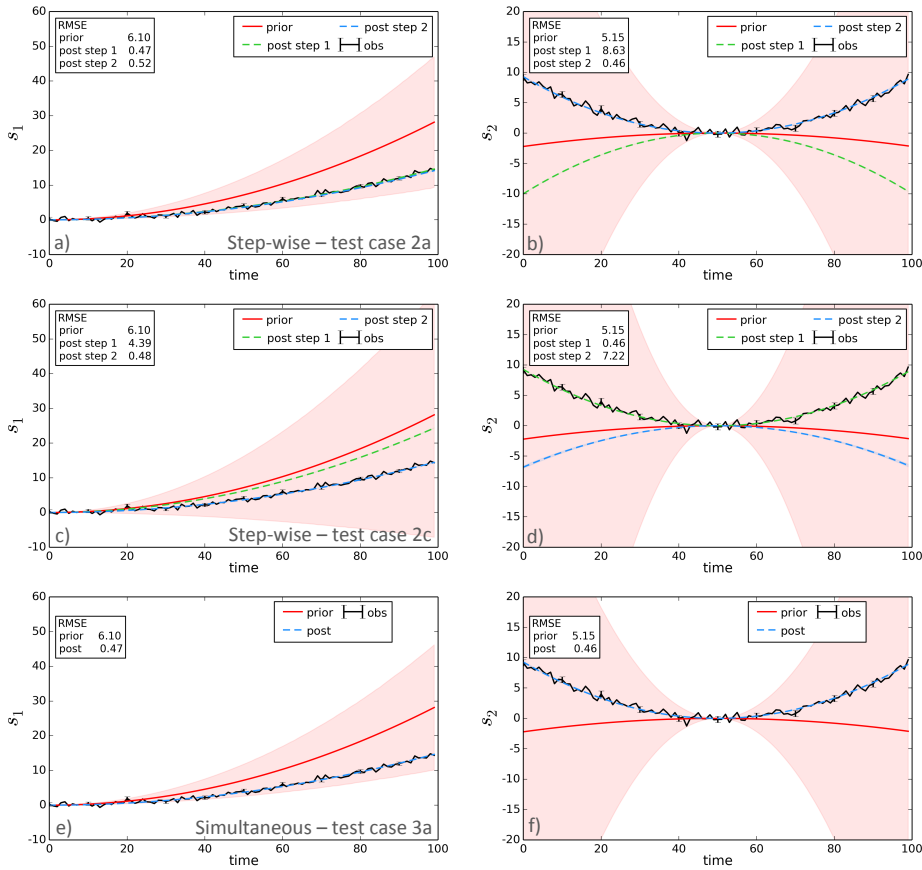
43

Figure 5: Prior and posterior model simulations compared to the synthetic observations for the non-linear toy model (with no bias) for both the s_1 (left column) and s_2 (right column) variables for a) and b) test case 2a (1st row) – step-wise approach with s_1 observations assimilated in the first step, followed by the s_2 observations in the second step; c) and d) test case 2c (2nd row) – step-wise approach with s_2 observations assimilated in the first step, followed by s_1 observations in the second step; and e) and f) test case 3a (3rd row) – the simultaneous case in which both data streams were included. For both step-wise examples $\mathbf{A}_1$ was propagated between the 1st and 2nd steps. The coloured error band on the prior and posterior represents the propagated parameter uncertainty (1 sigma) on the model state variables (in the equivalent colour as the mean curve). This is mostly visible for the prior

44

model simulation (pink band) as there is a high reduction in model uncertainty reduction as a result of the assimilation.

45
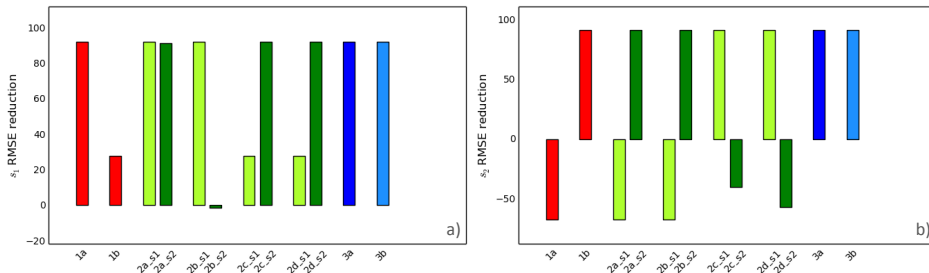
Figure 6: Reduction in RMSE for all test cases for both a) s_1 and b) s_2 variables for the non-linear toy model simulations with no model bias. For the step-wise cases (2a, b, c and d) the reduction after both step 1 and step 2 are shown in light and dark green respectively, and are denoted in the x-axis labels with '_s1' for step 1 and '_s2' for step 2. The reduction (in %) is calculated as $1 - (RMSE_{prior} / RMSE_{post})$.

46
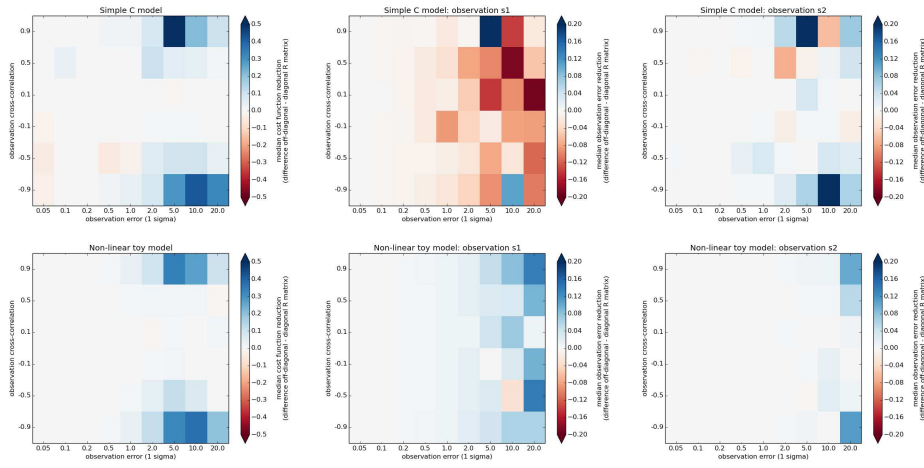
Figure 7: Median difference (across 20 first guess parameters) between including correlated observation errors in the R matrix (off-diagonal elements) minus ignoring the correlated observation errors (keeping R diagonal) for the reduction cost function (a and d: left column) and the reduction in s_1 and s_2 observation errors (b, c, e and f: middle and right columns), for both the simple C model (a, b and c: top row) and the non-linear toy model (d, e and f: bottom row). The reduction is calculated as 1 – (posterior/prior).

1 **Supplementary material**
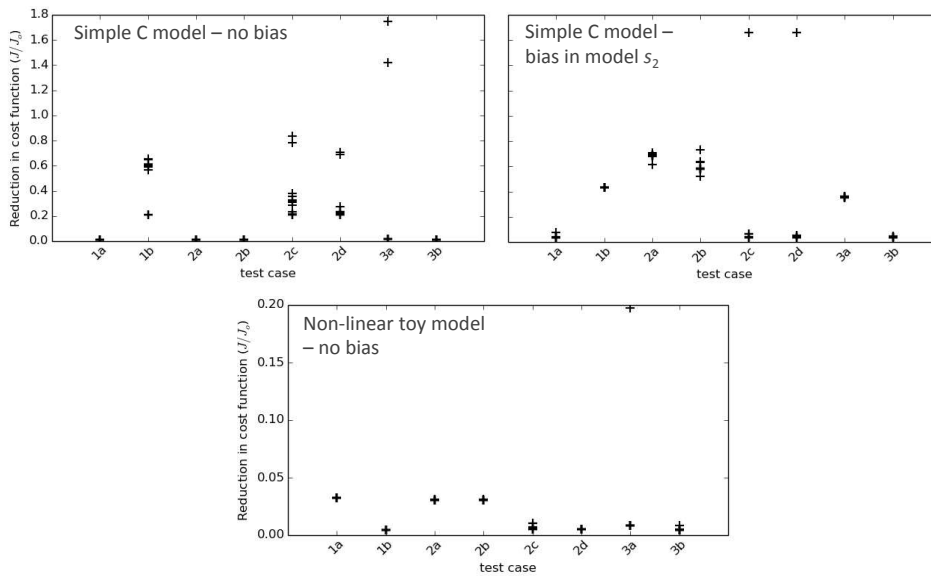
2

3 Figure S1: Reduction in the cost function (J/J_o) for each model and each test for all 20
4 assimilations with different random "first guess" points in the parameter space (i.e. each cross
5 represents the 20 random "first guess" tests). Top panel – simple C model without bias (left)
6 and with bias added to the simulated s_2 variable (right). Bottom panel – non-linear toy model
7 with no added bias. Note that the majority of the random "first guess" assimilations achieve
8 the same reduction in the cost function even though the final value is different for each test,
9 which is to be expected as each test (for each model) has a different cost function.

Test case 3b, in which only one observation was included for the s_2 data stream instead of the complete time-series, shows that a substantial difference in number of observations between the data streams can influence the resulting parameter values and posterior uncertainty (compare test cases 3a and b in Fig. 2 for the simple C model and Fig. 4 for the non-linear toy model) as each data stream will have a different overall "weight" in the cost function. However, the impact of having a different number of observations for each data stream in the cost function also depends strongly on the prescribed observation error and relative sensitivity of each corresponding model variable to the model parameters. If one variable has a greater sensitivity than the other, it will matter less if fewer observations of that variable are included in the cost function.

Xu et al. (2006), among others, have mentioned the possible need to weight the cost function for different data sets. Different arguments abound on this issue. Some contend that the cost function should not be weighted by the number of observations because the error covariance matrices (**B** and **R**) already define this weight in an objective way (e.g. Keenan et al., 2013). Certainly it should not be necessary to weight by the number of observations in the cost function if there is sufficient information to properly build the prior error covariance matrices (**B** and **R**). On the other hand, it is a difficult task to characterise the model structural uncertainty and the observation error correlations (see Kuppel et al., 2013 for practical solutions). Given this, our expert knowledge on the workings of the model processes and the sensitivity of the model to the parameters may permit us to specify a stronger weight to a data stream that could help to constrain a particular section of the model, but for which there are only a few data points. Clearly the definition of the prior error model, including for the covariance between errors of the data streams, is of the upmost importance (Trudinger et al., 2007) and merits close attention in future multiple data stream assimilation studies.

Lessons to be learned when dealing with non-linearity

Most optimisation studies with a large-scale LSM use derivative methods based on a least-squares approach, and therefore rely on assumptions of Gaussian probability

and linear model sensitivity. However, if the model is weakly non-linear within the probability distribution around the point in parameter space that is being analysed (see Tarantola, 1987, p72), it is possible to use an iterative algorithm, such as the one described in Eq. (6), to find the minimum of the cost function (i.e. the maximum likelihood of the posterior parameter distribution). Furthermore a linearization of the model around the maximum likelihood estimation (minimum of $J(\mathbf{x})$) of the parameters can be used to calculate the posterior error covariance (see Eq. (6)). If the model is too strongly non-linear and therefore these assumptions are not met, it may not be possible to find the true global minimum of the cost function and the characterisation of the posterior probability distribution will be incorrect. This is a particular problem if the posterior parameter error covariance matrix is then propagated in a step-wise approach, although these issues are relevant to both step-wise and simultaneous assimilation. Note that performing a number of tests starting from different random "first guess" points in parameter space can help to diagnose if the global minimum has been reached, as outlined in Section 2.1.6 and discussed at the beginning of the results (Section 2.2).

It is possible to avoid dealing with issues related to non-linearity in the model sensitivity and non-Gaussian error distributions by using a global search method (e.g. Markov Chain Monte Carlo or a genetic algorithm) that randomly, but effectively, searches the entire parameter space. However in large dimensional problems, as is likely the case when optimising a LSM at large scales with multiple data streams, using a global search method is likely not feasible due to computational time constraints. In these cases, a derivative method is likely the only option.

An important finding of the results presented for the non-linear toy model in Section 2.2.3 is that degradation in another data stream is not necessarily the result of a bias or incompatibility between the observations and the model. Rather if the model sensitivity to the parameters is very non-linear, multiple combinations of parameter values may exist that result in a similar reduction of the cost function (multiple minima), but provide a different fit to each data stream. Even though all data streams may be sensitive to all the parameters, the information content of each will not be the same. Finding the true global minimum in this instance may require a bit more careful thought in planning the

assimilation set-up, and may depend on having a reasonable idea of the level of information each data stream can bring to constrain each parameter. It may be the case that one data stream has a higher non-linear sensitivity to the parameters and therefore may act as an "troublemaker" and pull the parameters in a direction that results in a degradation to the other data streams, as seen in Section 2.2.3. If a simultaneous optimisation is not possible, it may be useful under such circumstances to identify any "troublemaker" data streams, and assimilate them in the first step of the optimisation. In the second step "peacemaker" data streams, with a lower non-linear sensitivity to the parameters, will then find a compromise set of parameter values that can fit both data streams well, provided the full posterior parameter error covariance matrix is propagated between the steps in order to retain all the information brought by the first data stream. As discussed this could be an explanation for the results seen for the non-linear toy model test case 2a where s_1 was assimilated prior to s_2 (Figures 6a and b) as discussed in Section 2.2.3.

Page 29: [3] Deleted	Natasha MacBean	26/06/16 16:58

In the case of non-Gaussian error distributions consider performing a transformation to make the distributions more Gaussian, or avoid a least squares formulation and instead use a method that avoids outliers (e.g. absolute deviations – Trudinger et al., 2007).

Analyse and correct for biases in the observations, or approximately account for it in the observation error covariance matrix, **R**, using the off-diagonal terms or inflated errors (Chevallier, 2007), or by using the prior model-data RMSE to define the observation uncertainty.

Investigate potential incompatibilities between your model and data. Take time to understand which physical quantities your data correspond to and whether that is consistent with the description of the equivalent variable in the model. As for the previous point, one way of attempting to account for unknown inconsistencies between the model and data is to set the observation uncertainty, **R**, the prior RMSE between the model and the data.

Evaluate the impact on other model variables with independent observations, and if the optimisation degrades the fit compared to the prior, investigate the reasons behind the inconsistency and address them as above.

Assess the non-linearity of your model (multiple first guess tests can help in this regard), and if

.

Prior information is key in a Bayesian framework. Effort should be put into better constraining the prior parameter bounds of all parameters based on literature wherever possible.

Conduct preliminary sensitivity analyses to determine which parameters should be constrained by each data stream.

Set up experiments with synthetic data, as in this study, to understand the constraints posed by the different data streams you will include in the experiment.