Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-25-AC1, 2016
© Author(s) 2016. CC-BY 3.0 License.

Geoscientific Model Development Discussions

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

Natasha MacBean et al.

natasha.macbean@lsce.ipsl.fr

Received and published: 27 June 2016

Response to 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 frame- work 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 alleviate 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 âĂÍ fraction of the aboveground carbon reserve as represented by k1s1." → "The litterfall is an output of s1 (aboveground biomass) and an input to s2 (belowground biomass) and is calculated as a constant fraction of the aboveground carbon reserve, defined by k1s1".

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 âĂÍtwo carbon pools, s1 and s2:" to: "It is based on two equations that describe the temporal evolution of two living biomass (carbon) pools, s1 and s2, 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, s1 and s2, 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 s1 and s2 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 explicitely 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). âĂÍ 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.

» RESPONSE Yes, we have added this point to the end of that sentence, thank you. »

- 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 s2.” »

- 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 assumptions 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: âĂć 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). âĂć 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. âĂć 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. âĂć 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$0.5x0.5°) 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 s2 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 Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-25, 2016.