Biogeosciences Discuss.,
doi:10.5194/bg-2016-186-AC2, 2016

Biogeosciences
Discussions
Open Access

Interactive comment on "Interannual variability in Australia's terrestrial carbon cycle constrained by multiple observation types" *by* Cathy M. Trudinger et al.

Cathy M. Trudinger et al.

cathy.trudinger@csiro.au

Received and published: 5 August 2016

We thank reviewer #2 for their helpful comments, and provide the following responses:

[RC] The objectives could be more clearly laid out in the introduction, especially given the number of different tests (forcing, model structure and optimisation set-up with different observations) that are then detailed, which I feel could lead to the reader feeling confused as to what the overall aims are. For example, stating (in the introduction), "...we use BIOS-2 to explore inter-annual variability in NEP..." lacks some detail. "We are also interested in the effect of uncertainty in model parameters on modelled IAV" is more clear, but it might be better to

Printer-friendly version

Discussion paper

formulate some key objectives as bullet points or questions and then directly link the results to these objectives.

We agree that the objectives should be more clearly stated, and suggest the following: The objectives of this study are to use multiple observation types to constrain the IAV of terrestrial carbon fluxes for Australia. Specifically, multiple observation types are used to optimise parameters in BIOS-2.1, by generating an ensemble of acceptable parameter sets that will allow us to see the effect of parameter equifinality. We then to use these parameter sets in the model to calculate IAV in Australian NEP over recent decades. We are interested in the following questions: What is our best estimate of IAV in Australian carbon fluxes? How does parameter equifinality affect modelled estimates of IAV and the 2011 anomaly for Australia? How does parameter equifinality effect estimates of the processes contributing to IAV in NEP, including NPP and heterotrophic respiration and the effect of soil moisture on heterotrophic respiration.

[RC] I like the description of the issues related to parameters that can and cannot be well constrained by the optimisation in Section 2.4 (p6 lines ∼10–20). And I think the aim of producing a range of estimates of NEP anomalies based on parameter uncertainty/equifinality is a good one. However I think it is important to detail what you mean by "parameters that are not informed by the calibration dataset". Equifinality can occur for two reasons (broadly speaking): i) parameters to which the model variables used in the optimisation are not sensitive (therefore will not be constrained by the optimisation), and ii) parameters to which the model variables are sensitive, but which are correlated with other parameters. In this latter case it is possible that the parameters will be well constrained, but to the "wrong" values (i.e. if you did a test with psuedo observations where you knew the right value for the parameters, the posterior value would be wellconstrained but not correct, possibly due to the particular noise realisations in the given set of observations. In my understanding, parameter

equifinality mostly Are you trying to account for both types in your ensembles? From p6 lines 28-29 (point i) it seems that perhaps just the first type? Also, it isn't clear to me that you are also accounting for uncertainty that still remains in parameters that have been well-constrained (but not perfectly) by the optimisation " i.e. calculated from the posterior parameter covariance matrix. Given these questions, I think the manuscript would benefit considerably from a more detailed description of exactly what uncertainty is represented in the parameter ensembles that are generated using the null point Monte Carlo method (p6 lines 21-34).

We will add the following: The ensemble from the Null Space Monte Carlo analysis will include the effect of uncertainty in parameters to which model outputs for comparison with observations are not sensitive, as well as parameters to which the model outputs are sensitive but which are correlated with other parameters, as both of these are part of the calibration null space. In addition, the recalibration process and the fact that solutions with a range of values of Phi are retained means that the ensemble also accounts for uncertainty in parameters that are well constrained by observations but affected by measurement noise or model structural error (Sepúlveda and Doherty, 2015).

[RC] Given your focus on the inter-annual anomalies, I think it would be good to have a plot showing the inter-annual variability of the observations, prior and posterior, perhaps as a bar graph of mean year to year anomalies, in order to asses how (if) the optimisation improves the modeling of the inter-annual anomalies. It is not very easy to evaluate this from the annual time series plots, and it is not shown on Figure 9. It is not clear that the simulations of the IAV are improved by all this assimilation work, which was a question I had when I started reading the paper, given previous studies have shown that parameter optimisation can improve the fit to seasonal fluxes but not necessarily to the IAV (Kuppel et al.

2012). Indeed on page 11 Line24 you state that the NEP anomalies (the key topic under investigation in this study) are not significantly improved by the optimisation, yet the majority of the paper is about the different optimisation tests, rather than an investigation of the IAV simulated by the model after the optimisation. Therefore at times I felt like I was reading a study about different optimisation configurations, and the link to the IAV was missing (until the discussion). There is a lot of information in Figures 9, S14 and 15 that has not been discussed and could contribute greatly to the discussion on IAV. The authors mention that it appears that NPP is the main driver but most of the uncertainty is in the heterotrophic respiration. This is a useful conclusion but I feel a greater discussion of this result would add value to this study (I have made further comments on this in the minor comments below).

- Suggested figure and assessing whether optimisation improves IAV - The question of whether optimisation improves IAV isn't so important, as it depends on what parameters you start with. Some of the parameters were already optimised in Haverd et al. (2013a). We were more interested in the uncertainty due to parameter equifinality, as well as how well we could model IAV at the flux sites (the answer to this question is, not particularly well). There are already timeseries plots, scatter plots and statistics showing how well we model IAV at the flux sites, so we don't believe an additional plot as suggested is warranted. In addition, the flux records at some of the sites only have a few years of data, so we wouldn't want to calculate mean year to year anomalies for all sites.
- By merging some of the Discussion into the Results section, as mentioned below, we hope to keep the focus on IAV and consolidate most of the discussion of Figs 9, S14 and S15.
- The conclusion of greater uncertainty in heterotrophic respiration - see comments below.

[RC] Given that the optimisation does not dramatically improve the IAV I was left feeling the paper was somewhat disjointed in its objectives. Also, the fact that the optimisation does not improve the simulation of anomalies is a key finding and could be discussed more. What could the modeling community do to better optimise the model/simulate the NEP anomalies? Some of the tests, for example the focus on the soil moisture affect on soil respiration, the contribution of different data streams, are not well linked to the initial objectives that are laid out. For example it was unclear to me why there is a focus on the soil respiration function despite the study of Exbrayat et al. (2013), particularly because in the introduction the authors discuss that previous studies have shown that NPP is the key driver. While obviously it would appear this function is important to investigate in terms of the development of the model in general, it was not clear that this was crucial for this study investigating the NEP anomalies. I would detail more clearly up front (in the introduction) why you have included different aspects (including the soil respiration function) that haven't been included in previous modelling studies with BIOS-2, i.e. explain (as you do later in the discussion p11 lines 30-35), that the purpose is to have best model set-up, otherwise I'm searching for a reason as to why these tests are particularly important for simulating the NEP anomalies.

- The objectives need to be more clearly stated, see above.
- We will put more emphasis on the point that although the model does not match IAV well at the flux sites, measurements at most of the Ozflux sites do not show a strong relationship between NEP and available soil water, and therefore are not particularly representative of IAV in NEP for Australia as a whole.
- What could the modelling and observation community do to improve simulation of IAV - We suggest the following: Flux observations at more representative sites might help. It is also not clear whether the meteorological drivers can explain the

IAV at the current flux sites, and a study similar to Abramowitz et al (2008) using statistical models but focussed on the interannual timescale at Australian sites may be useful to answer that question.

- Why focus on soil respiration function - yes, we did want to have the best model set-up. Soil moisture is important for soil respiration, and precipitation is important for IAV in NEP, so we wanted to ensure that we had the best estimate of the timing of heterotrophic respiration, and to have uncertainty in this function contribute to parameter uncertainty in NEP IAV. We're not aware of previous studies that have formally optimised this function. These points would be included in a revised manuscript.

[RC] This is maybe a style issue, but to deal with the issues raised in the previous comments, I think it may be better to merge some of the discussion into the results. I found the 'story' of the manuscript came together more in the discussion, but the description of results seemed incomplete at times or in need of interpretation. As already mentioned I was left wondering how each section (that may have contained a very nice test in itself) fitted in with the overall goals of better quantifying the NEP anomalies, instead of just being an interesting parameter optimisation test bed.

We agree that it would help the flow of the manuscript to merge some of the discussion into the results.

Minor comments

[RC] Why did you use monthly observations when I presume daily (or half hourly) fluxes are available? I guess this is because you are focusing on seasonal to annual (interannual) timescales but this needs a justification. Out of interest

have you done a test with daily observations to see if the optimisation of IAV is improved? Similarly, why were long-term means used for streamflow data?

We use monthly mean flux measurements as constraints because the meteorological forcing is not as accurate at the daily timescale, and not at all accurate at subdaily timescales (we use daily meteorological forcing downscaled to the CABLE timestep using a weather generator, as mentioned in Section 2.1). We have not tested the optimisation with daily observations, and would not expect an improvement. Long-term means are used for streamflow observations because BIOS-2 does not model streamflow dynamics well, something that we plan to address in future work. This information will be added.

[RC] How were the observation error correlations taken into account? This will be a particular issue for GPP observations that were derived from the NEP eddy covariance data via a flux partitioning method.

We did not take into account observation error correlations. We acknowledge that GPP and NEP would have correlated errors, but we don't have good information about their error statistics.

[RC] Figure 2: it would be informative to see prior and posterior RMSE or R, some metric to show the improvement.

We add the following: Phi for the best case divided by Phi for prior parameters, split into observation groups, is as follows: ET 0.88, GPP 0.46, NPP 0.08 and streamflow 1.06 for CABLE observations and NEP 0.32, soil carbon 0.80, phytomass 0.95 and litter 0.78 for CASA-CNP observations. The best to prior ratio for total Φ was 0.36.

[RC] Does Figure 4 only show the posterior results?

Yes, this will be clarified.

[RC] It might be good to put which observations are used to optimise the CABLE and CASACNP parameters in Table 1 and 2.

This information is already given in the text at the end of Section 2.3, after the observations have been described. If the suggestion here is to add the types of observation to Tables 1 and 2, the reference to these tables comes in Section 2.2 before the observations are introduced, so we believe it is not beneficial to mention the observations in Tables 1 and 2.

[RC] Section 2.3: do you mean evaporation or evapotranspiration measurements were used? In the figures you say "ET", which is often used to denote the latter.

This should be evapotranspiration (ET) throughout.

[RC] P5 Line 30: define the PEST acronym (it might also be better to put the website here instead of later).

Parameter ESTimation. Yes, the website should be mentioned here instead of later.

[RC] Section 2.4: it might be good to describe briefly what the "down-gradient" method is for the non-specialist.

We would add the following: A down-gradient search method uses information about the gradient of the cost function with respect to the parameters to decide how to iteratively alter parameters to locate parameter values corresponding with the minimum in the cost function (Raupach et al., 2005).

[RC] P6 Line 21: How did you define the larger number of parameters from the initial SA used in the Haverd et al. (2013b) study?

Haverd et al excluded some parameters that were uncertain yet were identified in the sensitivity analysis as being unlikely to be constrained by the calibration observations. A number of these parameters were included here.

[RC] P6 Line 35 to next page: apologies if I've missed this, but why 21 parameter sets for CABLE and 19 for CASA-CNP?

The case generated by the original optimisation of CABLE and CASA-CNP could be included in either ensemble. The numbers 21 and 19 counted this case in the CABLE ensemble, when it is actually more appropriate (and aligned with our original thinking) to count it in the CASA-CNP ensemble, giving 20 members each. The CASA-CNP ensemble is therefore all cases with the same CABLE parameters. This also reflects the colors that were used in figures such as Figure S12.

[RC] P7 Line 35: What does this tell us? It shouldn't be the case that an optimisation with one data stream degrades the fit to another if the models are consistent with each other and with the observations, and the prior error covariance matrix are properly characterised.

Although the fit to most observations was improved, the fit to a few observations was degraded. We are using a range of different types of observations of carbon and water in a complex model, so it is not entirely surprising that there are some discrepancies. Richardson et al (2010) pointed out that this often occurs. Nonetheless, this must be an indication of deficiencies in the model and/or the observations and their uncertainty characterisation, but we have not yet been able to identify the specific causes of these deficiencies in our model.

[RC] P8 Line 5: "likely attributable" ! can you do a test to quantify this further, given the focus on this new function?

This is difficult to test, given all of the other changes since the Haverd et al study. We

Printer-friendly version

Discussion paper

therefore replace the text with “possibly attributable”.

[RC] Figure 6b: Is this for each observation type when it is used on its own, or when it is left out (as I understand for Figure 6a)?

This is for each observation group when used on its own. This will be clarified.

[RC] As mentioned above, Section 3.2 is one example of a section that lacks some more detailed description, rather than just a summary and explanation of the figures and how to interpret them (which is also useful). Aside from NPP, which observations are useful for which processes (even qualititatively, or using a summary metric), particularly in reference to constraining the IAV? There is also no description of Figure S11.

We will include the following: Many of the CABLE parameters are constrained by more than one observation group. The eddy flux data (ET and GPP) provide the tightest constraints on the biophysical parameters, as also found by Haverd et al (2013a), presumably because they contain temporal information. Streamflow seems to contain mostly redundant information that is available from the other observations, but is still worth including to mitigate against the effect of biases in any single observation type. In future work, we plan to improve streamflow dynamics in the model, and would then hope to take advantage of temporal information in the streamflow measurements.

Figure S11 shows that many parameters are not well constrained by the calibration observations, and most of those that are constrained to some extent are influenced by only one observation group (e.g. age_leaf_w and age_clitt2 by litter, age_wood and falloc_w by phytomass and soilc0_frac, age_csoil1, age_csoil2 and age_csoil3 by soil carbon), demonstrating the benefit of including all of these observation types. The function describing the effect of soil moisture on soil respiration is constrained by observations of both NEP and soil carbon. This analysis of which observation groups

Printer-friendly version

Discussion paper

constrain which parameters gives results that are mostly as we would have expected. However, we would have expected the soil respiration function to be constrained by litter observations, but this appears not to be the case, perhaps because the litter observations are quite sparse.

[RC] Figure 7: is this the total model-data mismatch using all observations?

No, just the observations for CABLE. This will be clarified.

[RC] P9 Line 1: parameter identifiaility ! I like this analysis, but one could also simply look at the parameter posterior covariance matrix which may be easier for some readers to follow as an initial example of parameter correlation.

Parameter identifiability is based on analysis of the posterior parameter covariance matrix, including eigenvector analysis. While a simple look at the covariance matrix will tell you about parameter correlations, identifiability involves more sophisticated analysis too.

[RC] P9 Line 10: Figures 12 and 13, not 11 and 12.

Yes, thank you.

[RC] P10 Line 1: We don"t know that there is only a little difference in the misfit function (when looking at fig. 11a) if we don't know what the prior misfit is.

The total cost function calculated with the prior parameters was 5411. However, the point we were trying to make here was that there seems to be no relationship between NEP IAV and Phi, as we see quite different values of 2011 NEP for very similar values of Phi.

Printer-friendly version

Discussion paper

[RC] P10 Lines 13-15 (Section 3.5): would be good to have a reference to a figure or table here.

This information is not shown clearly in any of the figures or tables, so we have added some numbers to the text as follows: Without the shade correction (Eq. 5), the agreement with calibration observations is a bit worse than our best case for some observation types (e.g. the ratio of optimised to prior Phi for the noshade case for NPP and soil carbon are 0.27 and 0.91, compared to 0.08 and 0.80 for our best case) and a bit better for others (GPP and NEP Phi ratio for the noshade case are 0.32 and 0.29, compared to 0.46 and 0.32 for our best case), but overall the total Phi is not significantly different.

[RC] P11 Line 16: Is it really clear or fair to say that the parameters related to heterotrophic respiration are particularly important when you have introduced a new function for the impact of soil moisture on respiration which has quite a number of parameters? Is it surprising that you have more equifinality for heterotrophic respiration given this new function? It would be good to describe this more in the results because at this point I was wondering if this was a clear conclusion of the analysis (i.e. that the heterotrophic respiration-related parameters are particularly important). A link to a figure at this point in the discussion may also help.

We agree that it is not clear that the parameters related to heterotrophic respiration are more important for Australian NEP than other parameters, and that the text "particularly those important for heterotrophic respiration" should be removed. However, we note that the number of parameters in this function is not an issue, as of the six parameters in this function, two parameters are not used at all (w2 and w3) so will not effect model outputs. And the function is reasonably well constrained by observations as discussed elsewhere.

[RC] P11 Lines 21-22: Similar point to above. It is not immediately obvious to me,

looking at Figure 9, S14 and S15 that the NEP anomaly is dominated by equifinality in the heterotrophic respiration. It depends which ecosystem contributes to the anomaly, which was only briefly discussed. Looking at Australia overall for 2011, it looks like the uncertainty due to parameter equifinality is the roughly the same for NPP and heterotrophic respiration. I think the same can be said for the ecosystems the authors say below contribute most to the NEP anomalies (savanna and sparse vegetation).

Yes, the reviewer is right. The uncertainty due to equifinality is larger for heterotrophic respiration than NPP in the tropics and temperate regions, but they are similar for Australia as a whole, or for the savanna and sparsely vegetation regions that contribute most to Australian NEP.

[RC] In general all the routines used from PEST should be described in more detail somewhere (supplementary?).

There is documentation of the PEST routines in the PEST manual, so we do not believe it is worth describing them in detail, however we can add a short summary of what each routine does. The routines that we used were RANDPAR, PNULPAR, PARREP for null space Monte Carlo, and IDENTPAR for identifiability. We have already mentioned that GENLINPRED was used to calculate observation worth.

References:

Abramowitz, G., Leuning, 5 R., Clark, M., and Pitman, A.: Evaluating the Performance of Land Surface Models, Journal of Climate, 21, 5468–5481, doi:10.1175/2008JCLI2378.1, 2008.

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.

Change Biol., 11, 378-25 397, 2005.

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, 25-40, doi:10.1007/s00442-010-1628-y, 2010.

Sepúlveda, N. and Doherty, J.: Uncertainty analysis of a groundwater flow model in East-central Florida, Groundwater, 53, 764-474, doi:10.1111/gwat.12232, 2015.

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2016-186, 2016.