Author response to reviewer comments and manuscript changes in "Interannual variability in Australia's terrestrial carbon cycle constrained by multiple observation types", by Trudinger et al

Reviewers Comments [RC], Authors' Response [AR] and Manuscript Changes [MC] are are follows:

Author response to comments by Reviewer #1

We thank reviewer 1 for their helpful comments, and provide the following responses:

[RC] (1) The models (CABLE and CASA-CNP) produce daily carbon and water fluxes, why do the authors used monthly mean flux measurements (P5 L16) as constraints? Also, please make a table for the other biometric data (e.g., leaf NPP, soil carbon and above-ground phytomass and litter etc.) if there are a few data points.

[AR] We use monthly mean flux measurements as constraints because the meteorological forcing is not as accurate at the daily timescale. There are too many data points for a table, instead we include map showing the location of the biometric and other data that are used for calibration.

[MC] Add the following to Section 2.3 on Observations "Although the model produces daily carbon and water fluxes, we used monthly rather than daily flux observations because the meteorological forcing is not as reliable at the daily timescale." Add a new figure (new Fig 2) that is a map of observation locations.

[RC] (2) Please provide more details on the calculation of the cost function. For example, how "Different types of observations were then weighted relative to each other"?

[AR] The weights for each observation group (e.g. ET, GPP etc) were scaled so that each group contributed approximately equally to the cost function calculated with the prior parameters. This is important because the different types of observations can have vastly different magnitudes, and the relative contribution of each group to the cost function should not depend on the units that are used.

[MC] Add the text "The weights ..." from [AR] to Section 2.4 Optimisation method.

[RC] (3) The approach of generating ensemble parameter sets is not clear. How does "the null space Monte Carlo method" work? The authors stated that the purpose of using this method is "to quantify uncertainty due to parameter equifinality in model predictions", so my understanding is that this method can generate the posterior parameter distributions. Although the authors have detailed introduction to this method (P6 L27-33), I feel it is still difficult for the readers to understand why the generated parameter sets are suffi- cient to represent the posterior distribution of the parameters. Are there some special features of this method? Otherwise I can't believe only 30 parameter sets are enough to represent possible combination of parameters that are consistent with the observations, given more than 10 parameters are involved in each model. Even if we assume there are only 10 parameters, and each parameter can only be two possible values, there will 210 possible parameter sets. I understand that it may not be feasible to run regional simulations for a huge number of sets of parameters, but the authors need to demonstrate the 30 parameter sets are a good sample of the posterior parameter space.

[AR] The null space Monte Carlo method does not calculate posterior parameter distribu- tions. Rather, it is an efficient method to generate multiple parameter sets that are constrained by the calibration observations (Tonkin and Doherty, 2009). By taking into account the calibration and null spaces, it allows for these parameter sets to have the most range in combinations of parameters that are least well constrained by the calibration observations, therefore allowing for a much better representation of parameter equifinality than random combinations of parameters. Clearly, the more parameter sets the better, but we need to balance that with the computational limitations. Figures 7 and S12 show the ensembles of parameter sets (colored symbols), and they have many different values for the parameters that are not well constrained (many more than two values). Our purpose of using this method would perhaps be better described as ”to see the effect of uncertainty due to parameter equifinality in model predictions”.

[MC] We change “quantify uncertainty due to parameter equifinality” to “see the effect of uncertainty due to parameter equifinality”. We add two additional references, Tonkin and Doherty (2009) and Sepulveda and Doherty (2015) that give further details about the method. We also add “The NSMC method does not specifically calculate the posterior parameter probability distributions, however it is an efficient way to generate multiple parameter sets that span a significant amount of the equifinality in both models.”

Specific points:

[RC] P1,L15, L16: ecosystem respiration − > heterotrophic respiration

[AR] Yes, thank you.

[MC] Change “ecosystem respiration” to “heterotrophic respiration” at the two locations on page 1

[RC] P1,L21: you can give a hint to the readers that the detailed description of BIOS-2 is in section 2.1

[AR] Yes

[MC] Add “described in detail in Section 2.1” to the Introduction.

[RC] P5, L12, L25: evaporation - evapotranspiration

[AR] Yes, thank you.

[MC] Change “evaporation” to “evapotranspiration” or “ET” in two places in Section 2.3 and Figure captions.

[RC] P8, L7: R2 is unitless

[AR] Yes

[MC] Remove units.

[RC] Fig 6, Fig S11: change the figure style to 2D. i.e., x axis is parameter, y axis is observation variable (ET, NPP etc), and use colors (light to dark) to represent the in- crease/decrease in variance

[AR] Presumably the reason for changing the figure is that some bars were obscured from view. We have changed the figure to 2D, but still use bars rather than colors, and all bars can be clearly seen now. New figures for observation worth for CABLE and CASA-CNP replace the previous Figures 6 and S11.

[MC] Change figures 6 (now 7) and S11 as described in [AR].

[RC] P9, L12: please give the sources of the six bioclimatic regions

[AR] The bioclimatic regions are an aggregation of the agro-climatic classification of Hutchinson et al. (2005) into six classes, as described and used by Haverd et al (2013a, 2013b).

[MC] Text in [AR] added to Section 3.2.

[RC] P9, L18: the anomaly from the best case

[AR] No, the anomaly is calculated for each ensemble member by subtracting the temporal average for that ensemble member.

[MC] Add "When anomalies are shown for ensemble members here and in subsequent figures, they are calculated for each ensemble member by subtracting the temporal average of the quantity for that ensemble member." to Section 3.1.1.

[RC] Consider citing these papers: Richardson et al., (2010), Chen et al., (2011), Keenan et al., (2012).

[AR] Thanks to the reviewer for drawing these to our attention.

[MC] We have added citations for all of these papers.

Author response to comments by Reviewer #2

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

[AR] 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 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?

[MC] Text added to Introduction as described in [AR].

[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).

[AR] 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).

[MC] Text added to Section 2.4 as described in [AR].

[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).

[AR]

- 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.

[MC] We have not added the suggested figure. The other points are addressed 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.

[AR]

- 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.

[MC] Clearly state the objectives, as described above. Add to the Conclusions "The timing of interannual variations in NEP at the flux sites is not particularly well captured by the model, as has been found in previous modelling studies, however most of the flux measurements are from locations that are not water limited, in contrast to the parts of the country that most influence Australian NEP.". Add text "Flux observations at more ..." from [AR] to the Discussion. Add to the Introduction "We include some improvements to the model structure and forcing data (specifically, ..."

[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.

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

[MC] We have merged a significant amount of the information from 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?

[AR] 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.

[MC] Add the following to Section 2.3 on Observations "Although the model produces daily carbon and water fluxes, we used monthly rather than daily flux observations because the meteorological

forcing is not as reliable at the daily timescale." and "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."

[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.

[AR] 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. Temporal correlations have been shown (Lasslop et al., 2008) to decrease significantly with increasing time lags, so we don't expect temporal correlations to be particularly important for monthly measurements.

[MC] No change made.

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

[AR] 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.

[MC] Add text from [AR] to Section 3.1.1 Comparison of model outputs with observations.

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

[AR] Yes, this will be clarified.

[MC] Add "(best case)" to figure caption (now Fig 5).

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

[AR] 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.

[MC] No change.

[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.

[AR] This should be evapotranspiration (ET) throughout.

[MC] "Evaporation" changed to "evapotranspiration" or "ET"

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

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

[MC] "... we used the PEST implementation (Parameter ESTimation, http://www.pesthomepage.org)"

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

[AR] 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).

[MC] Add the text from [AR] to Section 2.4.

[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?

[AR] 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.

[MC] Add to Section 2.4 "Haverd et al (2013b) used parameter sensitivity analysis to choose which parameters to optimise, avoiding parameters that were unlikely to be constrained by the available observations. Here we optimise a larger number of parameters, including some of the parameters that are not well constrained by the calibration dataset, to explore the effect of ..."

[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?

[AR] 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.

[MC] Change text at the end of Section 2.4 to 20 parameter sets for each model.

[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.

[AR] 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.

[MC] Add text to Section 3.1.1 "The degradation of the fit to some observations is a consequence

of trying to fit many different types of observations at once with a complex model, and it is not entirely surprising that there are some discrepancies. Richardson et al (2010) pointed out that this often occurs. Nonetheless, it is an indication of deficiencies in the model, including the forcing and specification of parameters, 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?

[AR] This is difficult to test, given all of the other changes since the Haverd et al study. We therefore replace the text with "possibly attributable".

[MC] Replace "likely attributable" 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)?

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

[MC] Add to each part of the figure caption, now figure 7, "(i.e. leaving out each group in turn)" and "(i.e. each observation group is used on its own)."

[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.

[AR] 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 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.

[MC] Add text in [AR] to Section 3.1.1.

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

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

[MC] Change figure caption (now Fig 8) to "Model-data mismatch for CABLE observations, Φ_{CABLE}*, ..."*

[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.

[AR] 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.

[MC] Add to Section 3.1.1 "... parameter identifiability (Doherty and Hunt, 2009), based on analysis of the posterior parameter covariance matrix using tools that are available with PEST (routine IDENTPAR)

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

[AR] Yes, thank you.

[MC] Figure numbers S11 and S12 changed to S12 and S13 in Section 3.1.1.

[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.

[AR] 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.

[MC] Add to Section 3.1.1 "For reference, Φ_{CABLE} *with prior parameters was 3962" and "*Φ_{CASA} *for prior parameters (but using inputs from CABLE calculated with optimised parameters) was 1449." Rather than talking about "little difference in* Φ*", we say "as we see quite different values of 2011 NEP for very similar values of total* Φ*."*

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

[AR] 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.

[MC] Text from [AR] added to Section 3.1.2.

[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.

[AR] 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. Prior to this work, it would not have been possible to assess the importance of parameter uncertainties on the soil respiration fuction because the function was fixed.

[MC] Remove "particularly those important for heterotrophic respiration". Add to Section 3.1.1 "Parameters w_2 *and* w_3 *are unconstrained, but due to the values of the other parameters in this function they are not used they so are irrelevant to the model."*

[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).

[AR] 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.

[MC] In the Abstract and Conclusions change text to "with similar contributions from equifinality in parameters associated with NPP and heterotropic respiration." Change text in Figure 3.2 to "The range in heterotrophic respiration is larger than the range in NPP in the tropics and temperate regions, but they are similar in other regions and for Australia as a whole (Figs. S14 and S15)."

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

[AR] 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.

[MC] We add to Section 2.4 "We use the parallel implementation of PEST called BEOPEST." and "(specifically, we used routines SUPCALC to calculate the dimension of the solution space, RANDPAR to generate random parameter sets, PNULPAR to retain the null-space components of

the random parameter sets and replace the solution space components with that of the calibrated model, and PARREP to replace parameters in the model's control file with the new parameter sets in preparation for recalibration)". In Section 3.1.1 we add that parameter identifiability was calculated with routine IDENTPAR. We had already mentioned that observation worth was calculated with routine GENLINPRED.

Other changes

Added Holgate et al reference at the end of Section 2.1.1.

We added some additional references from the OzFlux Special Issue (Isaac et al 2016, Beringer et al 2016b) to Section 2.3.

Added "exceptional" in the abstract for emphasis.

Added Tilo Ziehn acknowledgement.

Added an additional step to the NSMC description in Section 2.4 "(vi) Eliminate any parameter sets that are not considered plausible."

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.

Hutchinson, M. F., McIntyre, S., Hobbs, R. J., Stein, J. L., Garnett, S., and Kinloch, J.: Integrating a global agro-climatic classification with bioregional boundaries in Australia, Global Ecology and Biogeography, 14, 197-212, doi:10.1111/j.1466- 822X.2005.00154.x, 2005.

Lasslop, G., Reichstein, M., Kattge, J., and Papale, D.: Influences of observation errors in eddy flux data on inverse model parameter estimation, Biogeosciences, 5, 1311-1324, doi:10.5194/bg-5-1311-2008, 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.

Tonkin, M., and J. Doherty, Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, Water Resour. Res., 45, W00B10, doi:10.1029/2007WR006678, 2009.

Interannual variability in Australia’s terrestrial carbon cycle constrained by multiple observation types

Cathy M. Trudinger[1], Vanessa Haverd[2], Peter R. Briggs[2], and Josep G. Canadell[2]

[1]CSIRO Oceans and Atmosphere, Aspendale, Australia
[2]CSIRO Oceans and Atmosphere, Canberra, Australia

Correspondence to: Cathy Trudinger (cathy.trudinger@csiro.au)

Abstract. Recent studies have shown that semi-arid ecosystems in Australia may be responsible for a significant part of the interannual variability in the global concentration of atmospheric carbon dioxide. Here we use a multiple constraints approach to calibrate a land surface model of Australian terrestrial carbon and water cycles, with a focus on interannual variability. We use observations of carbon and water fluxes at 14 OzFlux sites, as well as data on carbon pools, litter-fall and streamflow. We include calibration of the function describing the response of heterotrophic respiration to soil moisture. We also explore the effect on modelled interannual variability of parameter equifinality, whereby multiple combinations of parameters can give an equally acceptable fit to the calibration data. We estimate interannual variability of Australian ~~net ecosystem production (NEP)~~ NEP of 0.12–0.21 PgC yr^{-1} (1σ) over 1982-2013, with a high anomaly of 0.43–0.67 PgC yr^{-1} in 2011 relative to this period associated with exceptional wet conditions following a prolonged drought. The ranges are due to the effect on calculated NEP anomaly of parameter equifinality, ~~which we find to be dominated by the effect of parameter equifinality in heterotrophic respiration rather than NPP .~~ with similar contributions from equifinality in parameters associated with NPP and heterotropic respiration. Our range of results due to parameter equifinality demonstrates how errors can be underestimated when a single parameter set is used.

1 Introduction

The growth rate of carbon dioxide (CO$_2$) in the atmosphere has significant interannual variability, mostly driven by variability in CO$_2$ uptake by terrestrial ecosystems (Rayner et al., 2008; Bastos et al., 2013; Le Quéré et al., 2015). Recent studies have shown that while mean terrestrial CO$_2$ uptake is dominated by tropical forests, the trend and interannual variability in terrestrial uptake are dominated by semi-arid ecosystems (Poulter et al., 2014; Ahlström et al., 2015; Liu et al., 2015). Uptake of CO$_2$ by land (net ecosystem production, NEP) is the balance between net primary production (NPP) and ~~ecosystem~~ heterotrophic respiration, and NPP rather than ~~ecosystem~~ heterotrophic respiration appears to be the main driving mechanism behind variability in the land sink (Poulter et al., 2014; Ahlström et al., 2015). Evidence from ecosystem models, atmospheric inversions and satellite observations (Poulter et al., 2014; Detmers et al., 2015) suggests that a strong land carbon sink in 2011 (Le Quéré et al., 2015; Bastos et al., 2013) was driven by growth of semi-arid vegetation in the southern hemisphere, with a large contribution from Australia associated with wet conditions of an extraordinary La Niña event following a prolonged drought.

Here we use the BIOS-2 implementation for Australia (Haverd et al., 2013a) of the Community Atmosphere Biosphere Land Exchange (CABLE) land surface model (Wang et al., 2010, 2011), described in detail in Sect. 2.1, to explore interannual variability (IAV) in Australian carbon fluxes between 1982 and 2013. This work builds on the study by Haverd et al. (2013b) that used BIOS-2 to estimate the mean carbon budget for Australia. We use a multiple constraints approach (Raupach et al., 2005) to optimise model parameters. We include some improvements to the model structure and forcing data (specifically, optimisation of the function describing the sensitivity of heterotrophic respiration to soil moisture, that has been shown by Exbrayat et al. (2013a) to be important for modelling NEP, and an improved product for vegetation cover). The implementation of the model described here is denoted BIOS-2.1.

We are also interested in the effect of uncertainty in model parameters on modelled IAV. It is now widely recognised in many areas of research that there is usually no single ‘true’ parameter set, but that there may be many parameter sets for a given model structure that are equally acceptable for simulating the data. There are different aspects to this. Firstly, two or more parameters may have a similar effect on model outputs, so can be difficult to distinguish. This is called equifinality (Aalderlink and Jovin, 1997; Beven, 2006; Tang and Zhuang, 2008) and leads to correlated errors in the estimates of model parameters. Whether parameters are uniquely identifiable from comparison of model outputs with observations will depend on how the model is formulated (over-parameterisation will increase the chances of equifinality), and what types of observations are used to calibrate the model. Different kinds of measurements, or measurements with information about processes acting on different timescales, will constrain different model parameters (Luo et al., 2009), so thought should be given to whether the available observations are likely to contain information about parameters of interest. Secondly, even without compensatory parameters, all models and the observations used to calibrate them are in error to some extent. Therefore, there is no reason to believe that the global optimum parameter set is more correct than other parameter sets that fit the observations to a lesser, but still acceptable, extent (Beven and Binley, 1992).

Although global search methods like Markov Chain Monte Carlo can usually do a more thorough search of parameter space than down-gradient methods, they are much more computationally expensive. Ziehn et al. (2012) showed that both can do a good job at obtaining the probability density function for parameters in a land surface model. Here we use a down-gradient search method to optimise a set of model parameters, with an efficient method (null-space Monte Carlo) to explore parameter equifinality.

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 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? The outline of the paper is as follows. In Sect. 2 we describe the model, the forcing data, the observations used for calibration and validation, and the optimisation method. In Sect. 3 we present results, followed by discussion in ~~Section~~Sect. 4 and conclusions in Sect. 5.

2 Methods

2.1 BIOS-2.1 model

Haverd et al. (2013a) described an implementation of the CABLE land surface model (Kowalczyk et al., 2006; Wang et al., 2011), CASA-CNP biogeochemical model (Wang et al., 2010) and Soil-Litter-Iso (SLI) soil model (Haverd and Cuntz, 2010) for Australia at fine spatial resolution ($0.05^\circ \times 0.05^\circ$ grid, which is roughly 5×5 km) using daily meteorology (downscaled for CABLE using a weather generator). This composite model and environment was referred to as BIOS-2, and makes use of the modelling environment built for the Australian Water Availability Project (AWAP) (King et al., 2009; Raupach et al., 2009). Modifications to CABLE, SLI and CASA-CNP for use in BIOS-2 are described in Haverd et al. (2013a). Some of the main features are that plant functional types are not used, instead each cell is partitioned into woody and grassy tiles. CABLE was run at an hourly timestep, and daily values of GPP, soil moisture and soil temperature were used to drive CASA-CNP at daily timesteps. Nitrogen and phosphorous cycles in CASA-CNP were disabled, and land management was not considered explicitly.

Here we extend this work as described in the following sections, with a new implementation for Australia denoted BIOS-2.1. We optimise model parameters for CABLE and CASA-CNP separately, using the method described in Sect. 2.4. The model parameters that are optimised are given in Tables 1 and 2. We run the model from 1900, but focus our analysis on the period between 1982 and 2013 for which we have consistent remotely sensed vegetation cover (Sect. 2.2).

2.1.1 Heterotrophic respiration function of soil moisture

Soil moisture has an important effect on heterotrophic respiration (Exbrayat et al., 2013a, b; Sierra et al., 2015). Soil moisture simulated by land surface models is known to be very model dependent (Koster et al., 2009), so a function of soil moisture that performs well in one model may not perform well in other models. The standard version of CASA-CNP uses a function from Kelly et al. (2000) for the dependence of heterotrophic respiration on soil moisture (Exbrayat et al., 2013b). The function is

$$f(\overline{s}) = \left(\frac{\overline{s} - 1.70}{0.55 - 1.70}\right)^{6.6481} \times \left(\frac{\overline{s} + 0.007}{0.55 + 0.007}\right)^{3.22} \quad (1)$$

where $\overline{s}$ is the root-mass weighted mean soil moisture content relative to saturated.

Here we optimise the heterotrophic respiration dependence on soil moisture, using a new function with six parameters. These parameters are optimised along with other model parameters in CASA-CNP. The form of the function was designed such that the function in Eq. 1 can be replicated with a particular choice of parameters, but the function also allows the type of behaviour seen in many of the functions compared in Fig. 4c of Sierra et al. (2015). The equations for the function, with parameters q, c, w_0, w_1, w_2 and w_3 to be optimised, are as follows:

$$f(\overline{s}) = \begin{cases} \max\begin{cases} \mathrm{q}\times\overline{s}^2 \\ \frac{0.5}{1-f(0)}\times\left[1-\cos\left(\pi\frac{\overline{s}-\mathrm{w}_0+(\mathrm{c}-1)\times\mathrm{w}_1}{\mathrm{w}_1\times\mathrm{c}}\right)-f(0)\right] \end{cases} & \text{if } \overline{s}<\mathrm{w}_0+\mathrm{w}_1 \\ 1.0 & \text{if } \mathrm{w}_0+\mathrm{w}_1\le\overline{s}\le\mathrm{w}_0+\mathrm{w}_1+\mathrm{w}_2 \\ 0.5\left[1+\cos\left(\pi\left(\overline{s}-\mathrm{w}_0-\mathrm{w}_1-\mathrm{w}_2\right)/\mathrm{w}_3\right)\right] & \text{if } \overline{s}>\mathrm{w}_0+\mathrm{w}_1+\mathrm{w}_2 \end{cases} \tag{2}$$

where

$$f(0) = 0.5\times\left[1-\cos\left(\pi\frac{0-\mathrm{w}_0+(\mathrm{c}-1)\times\mathrm{w}_1}{\mathrm{w}_1\times\mathrm{c}}\right)\right] \tag{3}$$

The function and parameters are shown in Fig. 1. The first part of the function (in red in Fig. 1a) is a quadratic with the rate of increase described by the parameter ‘q’. The second part of the curve (green) is an increasing part of a cosine curve. The curvature is controlled by the parameter ‘c’ as it determines whether the full range of the cosine curve from -1 to +1, or just part of it near the top, is used. The solid and dashed green lines in Fig. 1a demonstrate the effect of parameter c, with different values of c (1.0 and 5.0, respectively) but identical values of the other parameters. The width of this part of the curve is set by parameter ‘w_1’. As this function starts from zero at $\overline{s}=\mathrm{w}_0$, in order to match up with the first (quadratic) part of the function, the maximum of the two functions is used where they overlap (the solid green line in Fig. 1a shows the maximum of the quadratic and cosine functions where they overlap, the dotted lines show the parts of the functions that are not used). If w_0+w_1 hasn’t already exceeded the maximum $\overline{s}$ value of 1.0, the next part of the curve (yellow) is a constant value of 1.0. The width is set by parameter ‘w_2’ and can be zero if no flat top is required. Then if $\mathrm{w}_0+\mathrm{w}_1+\mathrm{w}_2 < 1.0$, the final part of the function (blue) is a decrease described by a cosine, with width ‘w_3’. If w_0+w_1 is greater than 1.0, then w_2 is not used. Similarly, if w_0+w_1+w_2 is greater than 1.0, w_3 is not used. The function is continuous throughout the full range, but there is a discontinuity in the gradient when the quadratic meets the first cosine part of the function.

The parameters required for our function to approximately match the equation from Kelly et al. (2000) are (q,c,w_0,w_1,w_2,w_3) = (0.0, 1.0, 0.025, 0.522, 0, 0.64). We use these as prior values in the parameter optimisation. The Kelly et al. (2000) function is shown by the dashed line in Fig. 1b. We chose not to optimise the temperature dependence of soil respiration because soil moisture (due to precipitation) has a much greater influence on interannual variability in soil respiration than temperature (see Supplementary Fig. S1). Exbrayat et al. (2013a) noted that the impact of the soil moisture response function on heterotrophic respiration is intimately connected with the skill of the land surface model to simulate soil moisture. It has been demonstrated elsewhere ~~(Frost et al., 2015)~~ (Frost et al., 2015; Holgate et al., 2016) that BIOS-2 performs well for soil moisture. Haverd et al. (2013a) manually adjusted the soil moisture dependence of soil respiration but did not include it in their formal parameter estimation.

2.2 Forcing data

The model is forced using gridded meteorological data, soil properties and vegetation cover at $0.05^\circ \times 0.05^\circ$ spatial resolution. The meteorological data and soil properties used here are as described in Haverd et al. (2013b). Briefly, they consist of daily meteorology from the Bureau of Meteorology's contribution to the Australian Water Availability Project (Grant et al., 2008; Jones et al., 2009), downscaled to hourly timesteps for CABLE using a weather generator, and soil properties taken from the McKenzie and Hook (1992) and McKenzie et al. (2000) interpretations of soil types mapped in the Digital Atlas of Australian Soils (Northcote et al., 1960, 1975).

In BIOS-2 (Haverd et al., 2013a, b), vegetation cover came from LAI derived from fPAR (fraction photosynthetic absorbed radiation) estimates obtained from the AVHRR and MODIS time series. These time series covered the periods 1990–2006 and 2000–2011, respectively. Here, vegetation cover is derived from the third generation (NDVI3g) of the GIMMS NDVI time series (Tucker et al., 2005; Zhu et al., 2013). This gives us a consistent vegetation cover timeseries over several decades (1982-2013). Total fPAR is partitioned into persistent (mainly woody) and recurrent (mainly grassy) vegetation components, following the methodology of Donohue et al. (2009) and Lu et al. (2003). This methodology takes advantage of low levels of seasonal change in LAI in woody vegetation, allowing seasonal variation in fPAR to be attributed principally to grassy vegetation. The remaining and relatively constant fPAR signal is attributed to woody vegetation. LAI for woody and grassy components are estimated by Beer's Law (e.g. Houldcroft et al., 2009):

$$LAI_W = -\frac{1}{k}\log_e\left(1 - fPAR_W\right) \tag{4}$$

$$LAI_G = -\frac{1}{k}\log_e\left(1 - \frac{fPAR_G}{1 - fPAR_W}\right) \tag{5}$$

where W denotes the persistent or mainly woody vegetation type, G denotes the recurrent or mainly grassy vegetation type and k is an extinction coefficient, set here to 0.5. In contrast to earlier BIOS-2 simulations (Haverd et al., 2013a, 2016b), Eq. 5 accounts for the effect of shading of grass by woody vegetation. We include one case without this correction for shading of grass, i.e. using

$$LAI_G = -\frac{1}{k}\log_e\left(1 - fPAR_G\right) \tag{6}$$

instead of Eq. 5, but all other results presented here use the shade correction.

2.3 Observations

The following observations were used for calibration of model parameters. We used monthly eddy flux data (~~evaporation~~evapotranspiration (ET), GPP and NEP) from 14 OzFlux sites ~~(Beringer et al., 2016a)~~ (Beringer et al., 2016a; Isaac et al., 2016) listed in Table 3. The eddy flux data were processed using the DINGO (Dynamic INtegrated Gap filling and partitioning for OzFlux) methodology for processing raw flux tower data, as described in Donohue et al. (2014)~~and Haverd et al. (2016a)~~ , Haverd et al. (2016a) and

Beringer et al. (2016b). The period for which we have observations at each site is shown in the 4th column of Table 3 (giving a total of 70 site-years of data). A subset of these monthly observations (40 site-years of data) was used for parameter estimation (column 5 of Table 3). We excluded eddy flux observations at Tumbarumba around 2003 because it is known that an insect attack combined with drought stress had a significant effect on growth at this site (Keith et al., 2012), but BIOS-2 does not currently simulate the impact of disturbance including insect attacks. Although the model produces daily carbon and water fluxes, we used monthly rather than daily flux observations because the meteorological forcing is not as reliable at the daily timescale.

We also used long-term averaged streamflow at 51 unimpaired catchments (Vaze et al., 2011; Zhang et al., 2011). Monthly mean streamflow was calculated from daily measurements for months with data available for at least 90% of days. The long-term means were then calculated by averaging the monthly means where they exist, for comparison with modelled long-term means (calculated by averaging modelled streamflow for the same months). 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. Long-term means of leaf NPP, above-ground phytomass, above-ground litter and soil carbon density in the top 15 cm (Barrett, 2001; Raison et al., 2003) were also used. Figure 2 shows the location of the observations used for calibration.

The observations used to optimise CABLE parameters were ~~evaporation~~ET, GPP, streamflow and leaf NPP. The observations used to optimise CASA-CNP parameters were NEP, soil carbon and above-ground phytomass and litter. Haverd et al. (2013b) used many of the same observations, except in that study NEP was not used for parameter optimisation and the eddy flux data were processed in a different way.

2.4 Optimisation method

To optimise parameters, we used the PEST implementation (Parameter ESTimation, http://www.pesthomepage.org) of the Levenberg-Marquardt method (Doherty, 1999). This is a down-gradient search method, meaning that it 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). The cost function that was minimised was the sum of weighted square residuals (Φ) plus the mismatch of some parameters from prior estimates (Φ_P) where we were confident in our prior estimates based on the literature. Flux observations were weighted in Φ so that each flux site contributed equally to the cost function, regardless of the length of the record used for calibration. ~~Different types of observations were then weighted relative to each other so that they~~ The weights for each observation group (e.g. ET, GPP etc) were then scaled so that each group contributed approximately equally to Φ ~~.~~ calculated with the prior parameters. This is important because the different types of observations can have vastly different magnitudes, and the relative contribution of each group to Φ should not depend on the units that are used. We first optimised CABLE parameters using PEST, then used the optimised CABLE parameters to generate GPP, soil moisture and soil temperature inputs for CASA-CNP. We then optimised CASA-CNP parameters with PEST. We use the parallel implementation of PEST called BEOPEST.

Due to the large number of processes and parameters, calibration of land surface models is generally an under-determined problem, where there is no unique, correct parameter set and multiple parameter combinations can give an adequate match to observations. Doherty et al. (2010) describes some of the issues involved in optimisation of highly parameterised models. Some ~~parameters, or~~ combinations of parameters ~~,~~ are informed by the calibration dataset, and these are described as comprising the ‘calibration solution space’. Errors in solution space parameter combinations are due to measurement noise. Orthogonal to the calibration solution space is the ‘calibration null space’, containing combinations of parameters that are not informed by the calibration dataset. Expert knowledge, where available, gives the best estimates of parameters ~~in~~ that are part of the calibration null space. Often, parameters that have little effect on model outputs for comparison with the calibration dataset are fixed at prior values, this is sometimes called regularisation. As discussed by Doherty (2015), "the purpose of regularisation is to attain uniqueness where none in fact exists". We should avoid simply fixing parameters to possibly incorrect values just because the model outputs that correspond to calibration observations are not sensitive to them. If models are used to make predictions of quantities unlike those used for calibration of parameters, we need to be aware that the predictions may be sensitive to parameters, or parameter combinations, that were not well constrained through the calibration process (Doherty and Johnston, 2003). If this occurs, the uncertainties in these predictions are likely to be under-estimated.

Haverd et al. (2013b) used parameter sensitivity analysis to choose which parameters to optimise, avoiding parameters that were unlikely to be constrained by the available observations. Here we optimise a larger number of parameters, ~~even though some of them~~ including some of the parameters that are not well constrained by the calibration dataset, to explore the effect of different values of the poorly-constrained parameters on model predictions. Instead of aiming to estimate a single parameter set, we generate an ensemble of parameter sets, to represent parameters that are not well constrained. We can then use this ensemble of parameter sets to ~~quantify~~ see the effect of uncertainty due to parameter equifinality in model predictions, similar to Chen et al. (2011). We use some of PEST’s linear analysis tools~~(http://www.pesthomepage.org)~~, including null space Monte Carlo (NSMC). Null space Monte Carlo (Tonkin and Doherty, 2009) is an efficient way to generate multiple parameter sets that are consistent with the observations. As described in Doherty et al. (2010) ~~, the null space Monte Carlo~~ and Sepúlveda and Doherty (2015), the NSMC method consists of the following steps: (i) Identify the null-space of the model’s parameter field from sensitivities calculated during optimisation. (ii) Generate many stochastic realizations of model parameters. (iii) Project these realizations onto the null- and calibration spaces. (iv) Retain the null-space component, but replace the solution space component with that of the calibrated model. (v) Recalibrate these stochastic parameter sets (only a few iterations). Recalibration is required due to non-linearities in the model and an indistinct boundary between the solution- and null-spaces. (vi) Eliminate any parameter sets that are not considered plausible. Utilities exist as part of the PEST package to perform these steps after initial optimisation of parameters (specifically, we used routines SUPCALC to calculate the dimension of the solution space, RANDPAR to generate random parameter sets, PNULPAR to retain the null-space components of the random parameter sets and replace the solution space components with that of the calibrated model, and PARREP to replace parameters in the model’s control file with the new parameter sets in preparation for recalibration by PEST).

The ensemble from the NSMC 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 Φ are retained means that the ensemble also accounts for the uncertainty in parameters that are well constrained by observations but affected by measurement noise or model structural error (Sepúlveda and Doherty, 2015).

We generated ensembles of parameter sets in the following way. After optimising CABLE parameters with PEST, we used the ~~null space Monte Carlo~~ NSMC method to generate 30 additional parameter sets for CABLE. ~~We~~ Based on their values of Φ, we chose the best ~~21~~ 20 of these CABLE parameter sets from the NSMC and used them to generate GPP, soil moisture and soil temperature inputs for CASA-CNP, then we optimised the CASA-CNP parameters with PEST for each of these inputs. This gave us ~~21~~ 20 corresponding parameter sets for both models, and we call this ensemble of parameters for both models the 'CABLE parameter ensemble' because they originated from null space Monte Carlo applied to CABLE. We then used the null space Monte Carlo method to generate 30 parameter sets for CASA-CNP, all using a single set of CASA-CNP inputs ~~from~~ calculated with parameters from the original optimisation of CABLE. We retained the best ~~19~~ 20 of these parameter sets; we call this the 'CASA-CNP parameter ensemble' because it originated from null space Monte Carlo applied to CASA-CNP. Overall, this gave us 40 combinations of CABLE and CASA-CNP parameters~~that span at least some of the~~. The NSMC method does not specifically calculate the posterior parameter probability distributions, however it is an efficient way to generate multiple parameter sets that span a significant amount of the uncertainty due to equifinality in both models.

3 Results

In this section, we will focus first on the results of the parameter optimisation (e.g. how well the model matches the observations, which observations inform which parameters, and parameter equifinality), then look at modelled IAV in carbon fluxes (what the model predicts for Australia, and how parameter equifinality affects our estimates) and our modelled 2011 NEP anomaly.

3.1 ~~Comparison of model outputs with observations~~Optimisation results

3.1.1 Comparison of model outputs with observations

Figures 3 and 4 show monthly and annual timeseries of GPP, ET, ecosystem respiration and the anomaly in NEP at three contrasting OzFlux sites: Howard Springs (tropical savanna), Tumbarumba (cool temperate) and Alice Springs Mulga (sparsely vegetated). Timeseries for all 14 OzFlux sites considered here are shown in Supplementary Material (Figs. S2-S9). Our best case corresponds to the case that has the lowest total Φ (~~i.e.~~ where total Φ is the sum of Φ for both CABLE and CASA-CNP calculated separately, i.e. $\Phi_{\mathrm{CABLE}} + \Phi_{\mathrm{CASA}}$). This also happens to be the case with the lowest ~~Φ for the CASA-CNP observations~~Φ_{CASA}. The annual timeseries plots also show model results for the ensembles of parameter sets. In addition, we include one case of ecosystem respiration and NEP calculated with the Kelly et al. (2000) soil respiration function (parameters other than those used in Eq. 1 have been re-optimised for this case). When anomalies are shown for ensemble members here and in subsequent figures, they are calculated for each ensemble member by subtracting the temporal average of the quantity for that ensemble member.

Figure 5 shows results for monthly and annual ET, GPP and ecosystem respiration for the best case plotted as scatter plots of model versus observations, with different colors used for each site. Figure 6 shows the modelled versus measured soil carbon density in the top 15 cm for the best case. Modelled versus measured streamflow, leaf NPP, above-ground litter and above-ground phytomass for the best case are shown in Supplementary Fig. S10. Supplementary Table S1 compares our flux and pool estimates averaged over 1990-2010 for the best case and ensemble mean with values from BIOS-2 in Haverd et al. (2013b). We also show our uncertainty range due to parameter equifinality (1σ) calculated from the ensemble and the total uncertainty from Haverd et al. (2013b) due to parameter and forcing uncertainty.

There was no apparent relationship between ~~Φ for CABLE and CASA-CNP calculated separately~~ Φ_{CABLE} and Φ_{CASA}, indicating that a better fit to the observations in CABLE did not lead to a better fit to CASA-CNP observations. Monthly ET, GPP and ecosystem respiration at Howard Springs and Tumbarumba are both dominated by a strong seasonal cycle, whereas the variability at Alice Springs Mulga is very episodic. NEP is the difference between two large fluxes of opposite sign, and it is difficult to model the seasonal cycle well, particularly at Howard Springs where GPP and ecosystem respiration are highly correlated at monthly time-scales.

Φ for the best case divided by Φ 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. Note that the prior CASA-CNP case used prior CASA-CNP parameters but inputs from CABLE calculated with optimised CABLE parameters, and therefore shows the change in the agreement with observations due to optimisation of only CASA-CNP parameters. Optimisation of parameters has improved the agreement with many of the observations~~(compared to the simulation with prior parameters)~~, but has degraded the fit to a few observations. ~~This is a consequence of trying to fit many different types of observations at once.~~ For example, run-averaged NPP is significantly improved by parameter optimisation~~. Conversely, the~~ , but run-averaged streamflow is slightly worse. Figure 4d shows that mean GPP at Howard Springs has moved away from the observations (the parameter responsible for this change is vcmax_g, the maximum RuBP carboxylation rate to leaf for grass, that has moved away from its prior value). The degradation of the fit to some observations is a consequence of trying to fit many different types of observations at once with a complex model, and it is not entirely surprising that there are some discrepancies. Richardson et al. (2010) pointed out that this often occurs. Nonetheless, it is an indication of deficiencies in the model, including the forcing and specification of parameters, 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.

Overall we capture the observed level of NEP variability well. The agreement with observed annual NEP flux anomalies at the OzFlux sites has improved relative to the original BIOS-2 calculations: RMSE 0.39 $\mathrm{gCm^{-2}d^{-1}}$ (this work) compared with 0.58 $\mathrm{gCm^{-2}d^{-1}}$ (Haverd et al., 2013a), an improvement ~~likely~~ that is possibly attributable to optimization of the heterotrophic respiration response to soil moisture. However, the correlation between modelled and observed annual NEP values is poor ($\mathrm{R}^2 = 0.1$ ~~$\mathrm{gCm^{-2}d^{-1}}$~~, Fig. 5f). IAV in NEP at flux sites is difficult to capture well in the model for a number of reasons: 1) NEP is the difference between two very large fluxes, and these fluxes are temporally highly correlated in the non-temperate regions, with both fluxes being highly sensitive to soil moisture. 2) Flux measurements are quite local, whereas the model ~~is~~

has a resolution of 0.05^ox0.05^o. 3) We are missing some processes from the model, such as disturbance (e.g. the insect attack at Tumbarumba) and fire that may be important at the local scale of the flux measurements. 4) Flux measurements are also subject to errors, particularly due to the partitioning algorithm.

3.2 ~~Observation worth~~

3.1.1 Observation worth

Fig. 7 shows how each observation group ~~(i.e. type of observation)~~ contributes to the reduction of uncertainty in CABLE parameters. In Fig. 7a we show how the post-calibration uncertainty variance *increases* as each observation group is left out one at a time. A rise in uncertainty variance occurs for observation groups that contain unique information about a parameter that is not contained in the other groups. Figure 7b shows the *decrease* in pre-calibration uncertainty variance for each observation group used on its own. A decrease will occur when any observation group contains information about a parameter, even if this information is redundant. For example, the parameter lgamma_g (that controls the drought response of grass stomatal conductance) is constrained by all observation groups, and leaving out any individual observation group makes little difference to the post-calibration uncertainty, indicating redundancy in the information provided by the observation groups. ~~Parameters~~ Many of the CABLE parameters are constrained by more than one observation group. In contrast, parameters alloclg and alloclw (describing allocation of carbon to leaves) are mainly constrained by NPP observations, and leaving NPP observations out of the optimisation significantly increases the uncertainty in these parameters. 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.

The value of observation groups to estimation of CASA-CNP parameters is shown in Supplementary Fig. S11. ~~The observation~~ Many of the CASA-CNP parameters are not well constrained by the calibration observations. Most of those that are constrained to some extent are influenced by only one observation group, demonstrating the benefit of including all four observation groups. Specifically, parameters age_leaf_w and age_clitt2 describing turnover times of woody leaves and fine structural litter are influenced by observations of above-ground litter; age_wood describing turnover of wood and falloc_w describing the fraction of carbon allocated to wood are influenced by measurements of above-ground phytomass; and parameter soilc0_frac for the fraction of soil carbon in the top 15 cm and soil carbon pool turnover times age_csoil1, age_csoil2 and age_csoil3 are influenced by observations of soil carbon. 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 constrain which parameters for both CABLE and CASA-CNP gives results that are as we would have expected. The observation worth is calculated using PEST's linear analysis tools (routine GENLINPRED).

3.2 ~~Parameter equifinality~~

3.1.1 Parameter equifinality

Figure 8 shows scatter plots of the model-data mismatch for CABLE observations, Φ_{CABLE}, ~~Φ,~~ against each CABLE parameter, where the range of the x-axis corresponds to the prior range for the parameter. For reference, Φ_{CABLE} with prior parameters was 3962. Relative to their prior range, some parameters cover a small range for low values of ~~Φ~~Φ_{CABLE}, such as alloclw, alloclg (leaf carbon allocation coefficients in leaves and grass) and lgamma_g (controls drought response of grass stomatal conductance), indicating that they are relatively well constrained by the optimisation. Other parameters, such as f10_w (fraction of woody roots in the top 10cm), dleaf_g (grassy leaf length) and zr_g (maximum grassy rooting depth) cover a wide range of parameter values (relative to their prior range) for very little variation in ~~Φ~~Φ_{CABLE}, implying that their value is not so well constrained by the optimisation. In some cases, combinations of parameters might be well constrained by the optimisation but the individual parameters are not. We can see this by looking at the parameter identifiability ~~, which can be estimated using the analysis~~ (Doherty and Hunt, 2009), based on analysis of the posterior parameter covariance matrix using tools that are available with PEST (routine IDENTPAR). Figure 9 shows the identifiability of combinations of parameters in CABLE. Early eigenvectors (dark colours) are most identifiable (comprise the calibration solution space), later eigenvectors (pastel colours) are least identifiable (comprise the calibration null space). Eigenvectors split across parameters indicate whether combinations of parameters, rather than individual parameters, are identifiable. In general, parameters with the smallest ranges for low ~~Φ~~ Φ_{CABLE} in Fig. 8 are part of the most identifiable eigenvectors, as expected. The parameter f10_w (fraction of grass roots in the top 10cm) that had a wide range for low values of ~~Φ~~ Φ_{CABLE} in the scatter plots in Fig. 8, has identifiability that is comprised of a number of the most identifiable eigenvectors, as indicated by the dark colors. Whether it is a problem that parameters are not individually identified depends on what the model is being used to predict or calculate.

~~Parameter identifiability and scatter plots~~ Scatter plots and parameter identifiability for CASA-CNP are shown in the Supplementary Figs. ~~S11 and~~ S12 ~~.~~ and S13. Φ_{CASA} for prior parameters (but using inputs from CABLE calculated with optimised parameters) was 1449. Many of the CASA-CNP parameters are not well constrained relative to their prior ranges by the observations. The soil respiration function seems to be fairly well constrained compared to the range of curves shown in Fig. 4c of Sierra et al. (2015), and parameters s and w_1 in this function are among some of the best constrained parameters. Parameters w_2 and w_3 are unconstrained, but due to the values of the other parameters in this function they are not used they so are irrelevant to the model.

3.1.2 Model structural choices

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 Φ 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 Φ 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 Φ is not significantly different. We have used the shade correction here because it is more physically realistic, although the comparison with observations does not favour either parameterisation.

The function describing the effect of soil moisture on soil respiration for our ensemble of parameter sets is shown in Fig. 1b, with lines colored by the total Φ from both models. The optimised functions are all quite different to the function from Kelly et al. (2000), with the optimised functions increasing throughout the range of $\overline{s}$ from 0 to 1.0, rather than having a peak followed by a decrease. A case using the Kelly et al. (2000) function with the other CASA-CNP parameters re-optimised has a higher value of Φ, particularly for NEP observations, but this is to be expected when we optimise fewer parameters.

3.2 Interannual variation in NEP for Australia

Figure 10 shows modelled annual values of NEP anomaly for six bioclimatic regions and the continent. Results are shown for both the CABLE and CASA-CNP ensembles of parameters in grey. The red lines show the case re-optimised without the shade correction for deriving vegetation cover from fPAR, and the blue lines show the case re-optimised with the Kelly et al. soil respiration dependence on soil moisture. The bioclimatic regions~~are~~, shown in Fig. 10h, are an aggregation of the agro-climatic classification of Hutchinson et al. (2005) into six classes, as described and used by Haverd et al. (2013a, b). Annual values of GPP and heterotrophic respiration anomaly for the bioclimatic regions and Australia are shown in Supplementary Figs. S14 and S15. Supplementary Fig. S14 shows total NPP as well as the contributions from grassy and woody vegetation. IAV in NPP for grassy vegetation types is larger than the IAV in NPP for woody types in the tropics, savanna, Mediterranean and Australia. IAV in NPP for woody vegetation is similar to or larger than IAV in NPP for grassy vegetation in warm and cool temperate regions.

Figure 11 shows the modelled annual NEP anomaly together with annual precipitation for each region and the continent, and the Southern Oscillation Index. There is a strong relationship between NEP and precipitation in all regions, as has been shown in many previous studies. Precipitation is clearly the most important factor influencing interannual variations in NEP, predominantly precipitation in the current year but also to some extent the precipitation in the years leading up to the current year due to ‘memory’ effects (Schimel et al., 2005; Poulter et al., 2014). Our results show high NEP anomalies for Australia in 1983-84 (after a strong El Nino), 1989 (after a strong El Nino), 2000 (in the middle of a prolonged La Nina but Australian precipitation was very high) and very high values in 2010 and particularly 2011 (precipitation was very high, at the end of a decade-long drought). We see low NEP anomalies in 1982 (during a strong El Nino), 1994 (during the third of three consecutive El Ninos, with very low precipitation), 2002 (at the beginning of an El Nino, with very low precipitation) and 2014 (at the beginning of an El Nino). Using an earlier version of the BIOS-2.1 configuration (but without the correction to the vegetation cover for shaded grass), Haverd et al. (2016b) found no significant change in the sensitivity of Australian NEP to rainfall, contrary to the suggestion by Poulter et al. (2014) of a shift during recent decades in the sensitivity of vegetation activity to moisture availability. Haverd et al. (2016c) also used the BIOS-2.1 model and showed that at continental scale, annual variations in production are dampened by annual variations in decomposition, with both fluxes responding positively to precipitation anomalies, in contrast to previous global modelling results (Poulter et al., 2014) suggesting that IAV in Australian net carbon uptake is amplified by lags between production and decomposition.

Some years have significant ~~spread~~ range in NEP anomaly due to parameter equifinality, usually when the anomaly is furthest from zero (either positive or negative), while at other times the ~~spread~~ range is quite small. The ~~spread~~ range is particularly large compared to the calculated IAV in the tropics, medium in the warm and cool temperate and fairly small in the other regions and Australia as a whole. The ~~spread~~ range in heterotrophic respiration is larger than the ~~spread in NPP~~ range in NPP in the tropics and temperate regions, but they are similar in other regions and for Australia as a whole (Figs. S14 and S15)~~, indicating that the spread in NEP anomaly is mostly due to the effect of parameter equifinality on heterotrophic respiration rather than NPP.~~ .

~~IAV in NPP for grassy vegetation types is larger than the IAV in NPP for woody types in the tropics, savanna, Mediterranean and Australia. IAV in NPP for woody vegetation is similar to or larger than IAV in NPP for grassy vegetation in warm and cool temperate regions.~~

~~Figure 11 shows the modelled annual NEP anomaly together with annual precipitation for each region and the continent, and the Southern Oscillation Index.There is a strong relationship between NEP and precipitation in all regions, as has been shown in many previous studies~~Annual NEP anomalies for a case optimised without the shade correction are shown by the red lines in Fig. 10 and are mostly within the range given by the ensemble of cases with the shade correction. We generated an ensemble of results for the case without the shade correction (not shown here but used in Haverd et al. (2016b)) and it had a similar spread of results to the cases shown here. Therefore we place no importance in the difference between the red lines and our other cases. Annual NEP anomalies for a case calculated using the Kelly et al. (2000) soil respiration function are shown by the blue lines in Fig. 10 and fall at the high IAV end of the range given by the other cases.

We estimate that IAV in Australian NEP is 0.12–0.21 PgC yr^{-1} for the period 1982-2013. This quantity is the standard deviation of annual NEP anomalies calculated separately for each ensemble member, with the range given by the ensemble. NEP IAV relative to mean NPP is 6–10%. In the earlier BIOS-2 implementation, Haverd et al. (2013b) gave a continental value of 8% for NEP variability (1σ) relative to mean NPP for the period 1990–2011. Over this shorter period, our range is 7–10%.

3.3 2011 NEP anomaly

Our estimate for the 2011 NEP anomaly (relative to the 1982-2013 mean) is 0.43–0.67 PgC yr^{-1}. In Figure 12a we show the ensemble of estimates of the 2011 anomaly plotted against the corresponding model-data mismatch (total Φ). ~~This shows a large range in 2011 anomaly for little difference in Φ, indicating~~ There seems to be no relationship between the modelled magnitude of the 2011 peak and how well the model fits the ~~measurements used for calibration of both CABLE and CASA-CNP~~calibration observations, as we see quite different values of 2011 NEP for very similar values of total Φ. Figure 12b shows the size of the 2011 anomaly against Φ for just the monthly NEP flux measurements (Φ_{NEP}). While there is a suggestion of a relationship here (lower Φ_{NEP} corresponds to lower 2011 anomaly values), we note that the flux measurements are not without error, NEP at the flux sites is difficult to model well and the model does not include all of the processes that may be important at the local scale. We therefore do not have particularly higher confidence in the estimates that have better agreement with flux observations, but prefer to take account of the agreement of the model to all of the different types of measurements. ~~Fox et al. (2009) and Luo et al. (2015)~~ Previous studies (e.g. Fox et al., 2009; Richardson et al., 2010; Keenan et al., 2012b; Luo et al., 2015; Du et al., 2015) ha emphasised the importance of using both pool- and flux-based datasets to constrain land surface models, and a strength of our work is that we have used observations of both types in this study. In Fig. 12c we show the ensemble of estimates of the 2011 anomaly plotted against IAV (1σ), indicating a strong relationship between the size of the 2011 anomaly and overall IAV of each ensemble member.

3.4 ~~Shade correction for woody/grassy partition~~

~~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. soil carbon and NPP) and a bit better for others (NEP and GPP), but overall the total Φ is not significantly different. We have used the shade correction here because it is more physically realistic, however the comparison with observations has not shown either case to be more correct, so this difference in forcing data should be considered as part of the total uncertainty. Annual NEP anomalies for a case optimised without the shade correction are shown by the red lines in Fig. 10 and are mostly within the range given by the ensemble of cases with the shade correction. We generated an ensemble of results for the case without the shade correction (not shown here but used in Haverd et al. (2016b)) and it had a similar spread of results to the cases shown here. Therefore we place no importance in the difference between the red lines and our other cases.~~

3.4 ~~Soil respiration function~~

~~The function describing the effect of soil moisture on soil respiration for our ensemble of parameter sets is shown in Fig. 1b, with lines colored by the total Φ from both models. The optimised functions are all quite different to the function from Kelly et al. (2000), with the optimised functions increasing throughout the range of $\bar{s}$ from 0 to 1.0, rather than having a peak followed by a decrease. A case using the Kelly et al. (2000) function with the other CASA-CNP parameters re-optimised has a higher value of Φ, particularly for NEP observations, but this is to be expected when we optimise fewer parameters. Annual NEP anomalies for this case are shown by the blue lines in Fig. 10 and fall at the high IAV end of the range given by the other cases. The observations that have most influence on the parameters in the soil respiration function are NEP and soil carbon (Fig. S11).~~

4 ~~Discussion~~

~~Precipitation is clearly the most important factor influencing interannual variations in NEP, predominantly precipitation in the current year but also to some extent the precipitation in the years leading up to the current year due to 'memory' effects (Schimel et al., 2005; Poulter et al., 2014). Our results show high NEP anomalies for Australia in 1983-84 (after a strong El Nino), 1989 (after a strong El Nino), 2000 (in the middle of a prolonged La Nina but Australian precipitation was very high) and very high values in 2010 and particularly 2011 (precipitation was very high, at the end of a decade-long drought). We see low NEP anomalies in 1982 (during a strong El Nino), 1994 (during the third of three consecutive El Ninos, with very low precipitation), 2002 (at the beginning of an El Nino, with very low precipitation) and 2014 (at the beginning of an El Nino).~~

~~Using an earlier version of the BIOS-2.1 configuration (but without the correction to the vegetation cover for shaded grass), Haverd et al. (2016b) found no significant change in the sensitivity of Australian NEP to rainfall, contrary to the suggestion by Poulter et al. (2014) of a shift during recent decades in the sensitivity of vegetation activity to moisture availability.~~

The 2011 ~~NEP anomaly relative to the 1982-2013 mean is 0.43-0.67 PgC yr^{-1}, and~~ NEP anomaly stands out as extreme compared to all other years. The best case (with lowest ~~overall~~ total Φ) has 2011 NEP anomaly near the lower end of the range (0.47 PgC yr^{-1}). However during the development of this work we generated a few different ensembles of parameter sets with only small differences to the model and inputs, and found that the range stayed quite constant but that the 2011 anomaly for the best case could be anywhere within the range. ~~In Fig. 12a there is a wide range of values for the 2011 NEP anomaly for the half a dozen cases with the lowest Φ.~~ Parameter equifinality has an important effect on our estimate of the 2011 NEP anomaly, and we are currently not able to identify where within the range 0.43–0.67 PgC yr^{-1} the true value is most likely to sit. Further constraints on the model parameters ~~, particularly those important for heterotrophic respiration,~~ are needed to reduce the uncertainty in this estimate.

Using the Lund-Potsdam-Jena (LPJ) dynamic global vegetation model, Poulter et al. (2014) estimated a 2011 NEP anomaly relative to the 2003-2012 mean for Australia of 0.66 PgC yr^{-1}. Our estimate for the 2011 NEP anomaly relative to the 2003-2012 mean is 0.40-0.61 PgC yr^{-1}, just under the estimate from Poulter et al. Like Poulter et al., we see that the IAV in NEP (including the 2011 anomaly) is dominated by IAV in NPP rather than respiration. ~~However, our uncertainty in NEP anomaly~~ Our uncertainty in the NEP anomaly has roughly equal contributions from equifinality in parameters important for NPP and heterotrophic respiration for Australia and the savanna, sparsely vegetated and Mediterranean regions. Uncertainty in NEP in the tropical and temperate regions is dominated by the uncertainty due to parameter equifinality in ecosystem respiration rather than NPP.

4 Discussion

Our correlation of modelled and observed annual NEP at the flux sites is quite poor and, despite a reduction in Φ_{NEP} by optimisation of parameters, IAV in annual NEP at the flux sites appears not to be significantly improved by optimisation. ~~IAV in~~ Previous studies have found that land surface models have difficulty similating the correct timing of IAV in carbon fluxes (e.g. Urbanski et al., 2007; Keenan et al., 2012a, b), although see Desai (2010). A key point is that IAV in NEP at the flux sites is not particularly representative of IAV in NEP for the country as a whole. Other than Alice Springs, measurements at the flux sites do not show a strong relationship between NEP and available soil water. This is in contrast to the parts of the country that most influence the continental NEP, where vegetation growth is mostly water limited. Using BIOS-2.1, Haverd et al. (2016c) found that 90% of Australian IAV in NEP is due to the savanna and sparsely vegetated regions, and the majority of the flux sites are outside these regions.

~~Differences~~ Without many flux observations at sites that are water limited (the record from Alice Springs is currently only a few years long), it is difficult for us to assess how well the model simulates regional and continental carbon fluxes for Australia. Additional flux observations at sites that are water limited, and therefore more representative of Australian carbon fluxes, might help us to assess how well the model matches observations, and would also be valuable for parameter estimation and model development. It is 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.

Differences in the present study from Haverd et al. (2013a) include the use of an improved product for vegetation cover (GIMMS NDVI3g) that extends over several decades, the correction for shaded grass used in calculating the vegetation cover, the use of OzFlux NEP observations for parameter optimisation, inclusion of the function describing the effect of soil moisture on soil respiration in the optimisation (Haverd et al. (2013a) manually tuned this function), optimisation of a greater number of parameters, and more rigorous analysis of parameter uncertainty by the generation of multiple parameter sets that are used to explore parameter equifinality. Our estimates for the carbon pools and fluxes generally agree with Haverd et al. (2013a) within the uncertainty ranges (Supplementary Table S1). An exception is the fraction of continental NPP attributable to recurrent (assumed grassy) vegetation, which is 0.40 ± 0.04 (1σ), compared with 0.67 ± 0.14 in the 2013 analysis. Litter pools are also higher: continental average of 8.4 ± 2.3 tCha^{-1}, compared with 2.5 ± 1.3 tCha^{-1} in the BIOS-2 analysis. The increase in litter in the current work is attributable to a correction to the Haverd et al. (2013a) analysis in which litter observations were incorrectly assumed to be comparable with total (above and below-ground) fine-structural litter, when in fact they should be compared with only the above-ground component. We now have increased confidence in our estimates for IAV, principally due to the use of the improved product for vegetation cover, optimisation of the soil respiration function and more rigorous parameter uncertainty analysis.

We have focused here on the range of model results that come from parameter equifinality when many other choices are fixed, such as model structure, choice of the cost function and weights for observations and prior estimates of parameters, observations and forcing data. Many of these choices can also lead to uncertainty in the results. We have not calculated the total uncertainty here. Our range of results due to equifinality highlights the dangers of taking a single parameter set. In particular, comparison of model results for the single best parameter set for two different model configurations could easily lead to incorrect conclusions if the effect of parameter equifinality was ignored. Different types of observations can be tested to see whether they would reduce the uncertainty in parameters that are not well constrained. Future work already underway will include the effect of nutrients, land-use change, fire and tree demography on Australian carbon fluxes, with a more comprehensive assessment of the uncertainties.

5 Conclusions

We have used a multiple constraints approach to optimise model parameters in BIOS-2.1, an updated fine resolution implementation of the CABLE, CASA-CNP and SLI models for Australia, with a particular focus on interannual variability. In addition to other parameters, we optimised a function for the dependence of soil respiration on soil moisture. We have explored the effect of parameter equifinality on calculated interannual variation in NEP anomalies. The timing of interannual variations in NEP at the flux sites is not particularly well captured by the model, as has been found in previous modelling studies, however most of the flux measurements are from locations that are not water limited, in contrast to the parts of the country that most influence Australian NEP. We estimate that the 1σ variation in IAV in Australian NEP is 0.12–0.21 PgC yr^{-1}. The value of the IAV in NEP is dominated by NPP, but the range of estimates due to parameter equifinality ~~is dominated by the effect of parameter equifinality on heterotrophic respiration rather than~~ has roughly equal contributions from parameters associated with both heterotrophic respiration and NPP. The 2011 Australian NEP anomaly relative to the 1982-2013 mean is 0.43-0.67 PgC yr^{-1}. We find a strong relationship between the size of the 2011 anomaly and the overall IAV. Our range of results due to parameter equifinality demonstrates how errors can be underestimated when a single parameter set is used.

Acknowledgements. This work has been undertaken as part of the Australian Climate Change Science Program, funded jointly by the Department of the Environment, the Bureau of Meteorology and CSIRO. We thank researchers in the Australian Ozflux network for making the OzFlux data available. We thank Ranga Myneni and Jian Bi for supplying GIMMS NDVI3g updated to 2013, and Randall Donohue for processing these data. We thank John Doherty of Watermark Numerical Computing for making PEST freely available, and for advice on the use of PEST and its utilities. Tilo Ziehn provided helpful comments on the manuscript.

References

Aalderlink, R. H. and Jovin, J.: Identification of the parameters describing primary production from continuous oxygen signals, Wat. Sci. Tech., 36, 43–51, 1997.

Abramowitz, G., Leuning, 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.

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., Kato, E., Poulter, B., Sitch, S., Stocker, B. D., Viovy, N., Wang, Y. P., Wiltshire, A., Zaehle, S., and Zeng, N.: The dominant role of semi-arid ecosystems in the trend and variability of the land CO_2 sink, Science, 348, 895–899, doi:10.1126/science.aaa1668, 2015.

Barrett, D. J.: NPP Multi-Biome: VAST Calibration Data, Oak Ridge National Laboratory Distributed Active Archive Centre, Oak Ridge, TN, 2001.

Bastos, A., Running, S. W., Gouveia, C., and Trigo, R. M.: The global NPP dependence on ENSO: La Niña and the extraordinary year of 2011, Journal of Geophysical Research: Biogeosciences, 118, 1247–1255, doi:10.1002/jgrg.20100, 2013.

Beringer, J.: Whroo OzFlux tower site, OzFlux: Australian and New Zealand Flux Research and Monitoring. hdl:102.100.100/14232, 2013a.

Beringer, J.: Wallaby Creek OzFlux tower site, OzFlux: Australian and New Zealand Flux Research and Monitoring. hdl:102.100.100/14231, 2013b.

Beringer, J., Hutley, L. B., Tapper, N. J., and Cernusak, L. A.: Savanna fires and their impact on net ecosystem productivity in North Australia, Global Change Biology, 13, 990–1004, doi:10.1111/j.1365-2486.2007.01334.x, 2007.

Beringer, J., Hutley, L. B., Hacker, J. M., Neininger, B., and U, K. T. P.: Patterns and processes of carbon, water and energy cycles across northern Australian landscapes: From point to region, Agricultural and Forest Meteorology, 151, 1409–1416, doi:10.1016/j.agrformet.2011.05.003, 2011.

Beringer, J., Hutley, L. B., McHugh, I., Arndt, S. K., Campbell, D., Cleugh, H. A., Cleverly, J., Resco de Dios, V., Eamus, D., Evans, B., Ewenz, C., Grace, P., Griebel, A., Haverd, V., Hinko-Najera, N., Huete, A., Isaac, P., Kanniah, K., Leuning, R., Liddell, M. J., Macfarlane, C., Meyer, W., Moore, C., Pendall, E., Phillips, A., Phillips, R. L., Prober, S., Restrepo-Coupe, N., Rutledge, S., Schroder, I., Silberstein, R., Southall, P., Sun, M., Tapper, N. J., van Gorsel, E., Vote, C., Walker, J., and Wardlaw, T.: An introduction to the Australian and New Zealand flux tower network - OzFlux, Biogeosciences Discussions, 2016, 1–52, doi:10.5194/bg-2016-152, 2016a.

Beringer, J., McHugh, I., Hutley, L. B., Isaac, P., and Kljun, N.: DynamicINtegrated Gap-filling and partitioning for OzFlux (DINGO), Biogeosciences Discussions, 2016, 1–36, doi:10.5194/bg-2016-188, 2016b.

Beven, K.: A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18–36, 2006.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes, 6, 279–298, 1992.

Chen, M., Zhuang, Q., Cook, D. R., Coulter, R., Pekour, M., Scott, R. L., Munger, J. W., and Bible, K.: Quantification of terrestrial ecosystem carbon dynamics in the conterminous United States combining a process-based biogeochemical model and MODIS and AmeriFlux data, Biogeosciences, 8, 2665–2688, doi:10.5194/bg-8-2665-2011, 2011.

Cleverly, J., Boulain, N., Villalobos-Vega, R., Grant, N., Faux, R., Wood, C., Cook, P. G., Yu, Q., Leigh, A., and Eamus, D.: Dynamics of component carbon fluxes in a semi-arid Acacia woodland, central Australia, Journal of Geophysical Research: Biogeosciences, 118, 1168–1185, doi:10.1002/jgrg.20101, 2013.

Desai, A. R.: Climatic and phenological controls on coherent regional interannual variability of carbon dioxide flux in a heterogeneous landscape, Journal of Geophysical Research: Biogeosciences, 115, doi:10.1029/2010JG001423, 2010.

Detmers, R. G., Hasekamp, O., Aben, I., Houweling, S., van Leeuwen, T. T., Butz, A., Landgraf, J., Köhler, P., Guanter, L., and Poulter, B.: Anomalous carbon uptake in Australia as seen by GOSAT, Geophysical Research Letters, doi:10.1002/2015GL065161, 2015GL065161, 2015.

Doherty, J.: PEST: Model-Independent Parameter Estimation, Watermark Computing, Townsville, 1999.

Doherty, J.: Calibration and Uncertainty Analysis for Complex Environmental Models, Watermark Numerical Computing, Brisbane, Australia, 2015.

Doherty, J. and Hunt, R. J.: Two statistics for evaluating parameter identifiability and error reduction, Journal of Hydrology, 366, 119–127, doi:http://dx.doi.org/10.1016/j.jhydrol.2008.12.018, 2009.

Doherty, J. and Johnston, J. M.: Methodologies for calibration and predictive analysis of a watershed model, Journal of the American Water Resources Association, 39, 251–265, doi:10.1111/j.1752-1688.2003.tb04381.x, 2003.

Doherty, J. E., Hunt, R. J., and Tonkin, M. J.: Approaches to highly parameterized inversion: A guide to using PEST for model-parameter and predictive-uncertainty analysis, U.S. Geological Survey Scientific Investigations Report 2010-5211, 71 p., http://pubs.usgs.gov/sir/2010/5211/pdf/uncpest_sir2010-5211.pdf, 2010.

Donohue, R., Hume, I., Roderick, M., McVicar, T., Beringer, J., Hutley, L., Gallant, J., Austin, J., van Gorsel, E., Cleverly, J., Meyer, W., and Arndt, S.: Evaluation of the remote-sensing-based DIFFUSE model for estimating photosynthesis of vegetation, Remote Sensing of Environment, 155, 349–365, doi:10.1016/j.rse.2014.09.007, 2014.

Donohue, R. J., McVicar, T. R., and Roderick, M. L.: Climate-related trends in Australian vegetation cover as inferred from satellite observations, 1981-2006, Global Change Biology, 15, 1025–1039, doi:10.1111/j.1365-2486.2008.01746.x, 2009.

Du, Z., Nie, Y., He, Y., Yu, G., Wang, H., and Zhou, X.: Complementarity of flux- and biometric-based data to constrain parameters in a terrestrial carbon model, Tellus B, 67, doi:10.3402/tellusb.v67.24102, 2015.

Exbrayat, J., Pitman, A. J., Abramowitz, G., and Wang, Y.: Sensitivity of net ecosystem exchange and heterotrophic respiration to parameterization uncertainty, Journal of Geophysical Research: Atmospheres, 118, 1640–1651, doi:10.1029/2012JD018122, 2013a.

Exbrayat, J., Pitman, A. J., Zhang, Q., Abramowitz, G., and Wang, Y.: Examining soil carbon uncertainty in a global model: response of microbial decomposition to temperature, moisture and nutrient limitation, Biogeosciences, 10, 7095–7108, doi:10.5194/bg-10-7095-2013, 2013b.

Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C. M., and Van Wijk, M. T.: 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, 1597–1615, doi:10.1016/j.agrformet.2009.05.002, 2009.

Frost, A., Ramchurn, A., Hafeez, M., Zhao, F., Haverd, V., Beringer, J., and Briggs, P.: Evaluation of AWRA-L: the Australian Water Resource Assessment model, in: 21st International Congress on Modelling and Simulation, Gold Coast, Australia, 29 Nov to 4 Dec 2015, www.mssanz.org.au/modsim2015, 2015.

Grant, I., Jones, D., Wang, W., Fawcett, R., and Barratt, D.: Meteorological and remotely sensed datasets for hydrological modelling: A contribution to the Australian Water Availability Project, Catchment-scale Hydrological Modelling and Data Assimilation (CAHMDA-3) International Workshop on Hydrological Prediction: Modelling, Observation and Data Assimilation, Melbourne, 2008.

Haverd, V. and Cuntz, M.: Soil-Litter-Iso: A one-dimensional model for coupled transport of heat, water and stable isotopes in soil with a litter layer and root extraction, Journal of Hydrology, 388, 438–455, doi:10.1016/j.jhydrol.2010.05.029, 2010.

Haverd, V., Raupach, M. R., Briggs, P. R., Canadell, J. G., Isaac, P., Roxburgh, C. P. S. H., van Gorsel, E., Viscarra Rossel, R. A., and Wang, Z.: Multiple observation types reduce uncertainty in Australia's terrestrial carbon and water cycles, Biogeosciences, 10, 2011–2040, doi:10.5194/bg-10-2011-2013, 2013a.

Haverd, V., Raupach, M. R., Briggs, P. R., Canadell, J. G., M., S. J. D. R., Law, Meyer, C. P., Peters, G. P., Pickett-Heaps, C., and Sherman, B.: The Australian terrestrial carbon budget, Biogeosciences, 10, 851–869, doi:10.5194/bg-10-851-2013, 2013b.

Haverd, V., Smith, B., Raupach, M., Briggs, P., Nieradzik, L., Beringer, J., Hutley, L., Trudinger, C. M., and Cleverly, J.: Coupling carbon allocation with leaf and root phenology predicts tree-grass partitioning along a savanna rainfall gradient, Biogeosciences, 13, 761–779, doi:10.5194/bg-13-761-2016, 2016a.

Haverd, V., Smith, B., and Trudinger, C.: Dryland vegetation response to wet episode, not inherent shift in sensitivity to rainfall, behind Australia's role in 2011 global carbon sink anomaly, Global Change Biology, doi:10.1111/gcb.13202, 2016b.

Haverd, V., Smith, B., and Trudinger, C.: Process contributions of Australian ecosystems to interannual variations in the carbon cycle, Environmental Research Letters, doi:10.1088/1748-9326/11/5/054013, 2016c.

Holgate, C. M., De Jeu, R. A. M., van Dijk, A. I. J. M., Liu, Y. Y., Renzullo, L. J., Kumar, V., Dharssi, I., Parinussa, R. M., Van Der Schalie, R., Gevaert, A., Walker, J., McJannet, D., Cleverly, J., Haverd, V., Trudinger, C. M., and Briggs, P. R.: Comparison of remotely sensed and modelled soil moisture data sets across Australia, Remote Sensing of Environment, in press, 2016.

Houldcroft, C. J., Grey, W. M. F., Barnsley, M., Taylor, C. M., Los, S. O., and North, P. R. J.: New vegetation albedo parameters and global fields of soil background albedo derived from MODIS for use in a climate model, Journal of Hydrometeorology, 10, 183–198, doi:10.1175/2008JHM1021.1, 2009.

Hutchinson, M. F., McIntyre, S., Hobbs, R. J., Stein, J. L., Garnett, S., and Kinloch, J.: Integrating a global agro-climatic classification with bioregional boundaries in Australia, Global Ecology and Biogeography, 14, 197–212, doi:10.1111/j.1466-822X.2005.00154.x, 2005.

Hutley, L. B., Leuning, R., Beringer, J., and Cleugh, H. A.: The utility of the eddy covariance techniques as a tool in carbon accounting: tropical savanna as a case study, Australian Journal of Botany, 53, 663–675, doi:10.1071/BT04147, 2005.

Isaac, P., Cleverly, J., McHugh, I., van Gorsel, E., Ewenz, C., and Beringer, J.: OzFlux Data: Network integration from collection to curation, Biogeosciences Discussions, 2016, 1–41, doi:10.5194/bg-2016-189, 2016.

Jones, D. A., Wang, W., and Fawcett, R.: High-quality spatial climate data-sets for Australia, Aust. Meteorol. Oceanogr. J., 58, 233–248, 2009.

Keenan, T., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., Dragoni, D., Gough, C. M., Grant, R., Hollinger, D., Hufkens, K., Poulter, B., McCaughey, H., Raczka, B., Ryu, Y., Schaefer, K., Tian, H., Verbeeck, H., Zhao, M., and Richardson, A. D.: Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO_2 exchange, Global Change Biology, 18, 1971–1987, doi:10.1111/j.1365-2486.2012.02678.x, 2012a.

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, Global Change Biology, 18, 2555–2569, doi:10.1111/j.1365-2486.2012.02684.x, 2012b.

Keith, H., van Gorsel, E., Jacobsen, K., and Cleugh, H.: Dynamics of carbon exchange in a Eucalyptus forest in response to interacting disturbance factors, Agricultural and Forest Meteorology, 153, 67–81, doi:10.1016/j.agrformet.2011.07.019, 2012.

Kelly, R. H., Parton, W. J., Hartman, M. D., Stretch, L. K., Ojima, D. S., and Schimel, D. S.: Intra-annual and interannual variability of ecosystem processes in shortgrass steppe, Journal of Geophysical Research: Atmospheres, 105, 20 093–20 100, doi:10.1029/2000JD900259, 2000.

King, E. A., Paget, M. J., Briggs, P. R., Trudinger, C. M., and Raupach, M. R.: Operational Delivery of Hydro-Meteorological Monitoring and Modelling over the Australian Continent, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2, 241–249, doi:10.1109/JSTARS.2009.2031331, 2009.

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., , and Puma, M. J.: On the nature of soil moisture in land surface model, J. Climate, 22, 4322–4335, doi:10.1175/2009JCLI2832.1, 2009.

Kowalczyk, E. A., Wang, Y. P., and Law, R. M.: The CSIRO Atmospheric Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model, Tech. Paper. No. 13, (CSIRO Marine and Atmospheric Research, Aspendale, Australia), 2006.

Le Quéré, C., Moriarty, R., Andrew, R. M., Peters, G. P., Ciais, P., Friedlingstein, P., Jones, S. D., Sitch, S., Tans, P., Arneth, A., Boden, T. A., Bopp, L., Bozec, Y., Canadell, J. G., Chini, L. P., Chevallier, F., Cosca, C. E., Harris, I., Hoppema, M., Houghton, R. A., House, J. I., Jain, A., Johannessen, T., Kato, E., Keeling, R. F., Kitidis, V., Klein Goldewijk, K., Koven, C., Landa, C. S., Landschützer, P., Lenton, A., Lima, I. D., Marland, G., Mathis, J. T., Metzl, N., Nojiri, Y., Olsen, A., Ono, S., Peng, T., Peters, W., Pfeil, B., Poulter, B., Raupach, M. R., Regnier, P., Rödenbeck, C., Saito, S., Salisbury, J. E., Schuster, U., Schwinger, J., Séférian, R., Segschneider, J., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Werf, G. R., Viovy, N., Wang, Y., Wanninkhof, R., Wiltshire, A., and Zeng, N.: Global carbon budget 2014, Earth Syst. Sci. Data, 7, 47–85, doi:10.5194/essd–7–47–2015, 2015.

Leuning, R., Cleugh, H. A., Zegelin, S. J., and Hughes, D.: Carbon and water fluxes over a temperate Eucalyptus forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates, Agricultural and Forest Meteorology, 129, 151–173, doi:10.1016/j.agrformet.2004.12.004, 2005.

Liu, Y. Y., van Dijk, A. I. J. M., de Jeu, R. A. M., Canadell, J. G., McCabe, M. F., Evans, J. P., and Wang, G.: Recent reversal in loss of global terrestrial biomass, Nature Climate Change, 5, 470–474, doi:10.1038/nclimate2581, 2015.

Lu, H., Raupach, M. R., McVicar, T. R., and Barrett, D. J.: Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series, Remote Sensing of Environment, 86, 1–18, doi:10.1016/S0034-4257(03)00054-3, 2003.

Luo, Y., Weng, E., Wu, X., Gao, C., Zhou, X., and Zhang, L.: Parameter identifiability, constraint and equifinality in data assimilation with ecosystem models, Ecological Applications, 19, 571–574, 2009.

Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J. W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R. B., Jones, C. D., Lara, M. J., Liang, J., McGuire, A. D., Parton, W., Peng, C., Randerson, J. T., Salazar, A., Sierra, C. A., Smith, M. J., Tian, H., Todd-Brown, K. E. O., Torn, M., van Groenigen, K. J., Wang, Y. P., West, T. O., Wei, Y., Wieder, W. R., Xia, J., Xu, X., Xu, X., and Zhou, T.: Towards More Realistic Projections of Soil Carbon Dynamics by Earth System Models, Global Biogeochemical Cycles, doi:10.1002/2015GB005239, 2015.

Martin, D., Beringer, J., Hutley, L. B., and McHugh, I.: Carbon cycling in a mountain ash forest: Analysis of below ground respiration, Agricultural and Forest Meteorology, 147, 58–70, doi:10.1016/j.agrformet.2007.06.008, 2007.

McKenzie, N. J. and Hook, J.: Interpretation of the Atlas of Australian Soils, Technical Report 94/1992, CSIRO Division of Soils, Canberra, 1992.

McKenzie, N. J., Jacquier, D. W., Ashton, L. J., and Creswell, H. P.: Estimation of Soil Properties Using the Atlas of Australian Soils, Canberra, 2000.

Meyer, W. S., Kondrlovà, E., and Koerber, G. R.: Evaporation of perennial semi-arid woodland in southeastern Australia is adapted for irregular but common dry periods, Hydrological Processes, 29, 3714–3726, doi:10.1002/hyp.10467, 2015.

Northcote, K. H., Beckmann, G. G., Bettenay, E., Churchward, H. M., Van Dijk, D. C., Dimmock, G. M., Hubble, G. D., Isbell, R. F., McArthur, W. M., Murtha, G. G., Nicolls, K. D., Paton, T. R., Thompson, C. H., Webb, A. A., and Wright, M. J.: Atlas of Australian Soils, Sheets 1 to 10, with Explanatory Data, CSIRO Australia and Melbourne University Press, Melbourne, 1960.

Northcote, K. H., Hubble, G. D., Isbell, R. F., Thompson, C. H., and Bettenay, E.: A Description of Australian Soils, CSIRO Australia, 1975.

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y., Running, S. W., Sitch, S., and van der Werf, G. R.: Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle, Nature, 509, 600–603, doi:10.1038/nature13 376, 2014.

Raison, J., Keith, H., Barrett, D. J., Burrows, B., and Grierson, P.: Spatial Estimates of Biomass in “Mature" Native Vegetation, Australian Greenhouse Office, 2003.

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, 378–397, 2005.

Raupach, M. R., Briggs, P. R., Haverd, V., King, E. A., Paget, M., and Trudinger, C. M.: Australian Water Availability project: Phase 3 report, CAWCR Technical Report No. 13, 2009.

Rayner, P. J., Law, R. M., Allison, C. E., Francey, R. J., Trudinger, C. M., and Pickett-Heaps, C.: Interannual variability of the global carbon cycle (1992-2005) inferred by inversion of atmospheric CO_2 and $\delta^{13}CO_2$ measurements, Global Biogeochemical Cycles, 22, doi:10.1029/2007GB003068, 2008.

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.

Schimel, D., Churkina, G., and Braswell, B.: Remembrance of weather past: ecosystem response to climate variability, in: A history of atmospheric CO_2 and its effects on plants, animals, and ecosystems, edited by Ehleringer, J. R., Cerling, T. E., and Dearing, M. D., pp. 350–368, Springer-Verlag, Berlin, 2005.

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.

Sierra, C. A., Trumbore, S. E., Davidson, E. A., Vicca, S., and Janssens, I.: Sensitivity of decomposition rates of soil organic matter with respect to simultaneous changes in temperature and moisture, Journal of Advances in Modeling Earth Systems, 7, 335–356, doi:10.1002/2014MS000358, 2015.

Tang, J. and Zhuang, Q.: Equifinality in parameterisation of process-based biogeochemistry models: A significant uncertainty source to the estimation of regional carbon dynamics, J. Geophysical Research, 113, G04 010, doi:10.1029/2008JG000 757, 2008.

Tonkin, M. and Doherty, J.: Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, Water Resources Research, 45, doi:10.1029/2007WR006678, 2009.

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., Vermote, E. F., and El Saleous, N.: An extended AVHRR 8 km NDVI dataset compatible with 25 MODIS and SPOT vegetation NDVI data, Int. J. Remote Sens., 26, 4485–4498, 2005.

Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M., and Munger, J. W.: Factors controlling CO_2 exchange on timescales from hourly to decadal at Harvard Forest, Journal of Geophysical Research: Biogeosciences, 112, doi:10.1029/2006JG000293, g02020, 2007.

van Gorsel, E., Leuning, R., Cleugh, H. A., Keith, H., and Suni, T.: Nocturnal carbon efflux: reconciliation of eddy covariance and chamber measurements using an alternative to the u*-threshold filtering technique, Tellus B, 59, 397–403, doi:10.1111/j.1600-0889.2007.00252.x, 2007.

Vaze, J., Chiew, F. H. S., Perraud, J. M., Viney, N., Post, D., Wang, J. T. B., Lerat, J., and Goswami, M.: Rainfall-runoff modelling across southeast Australia: Datasets, models and results, Aust. J. Water Resour., 14, 101–113, 2011.

Wang, Y. P., Law, R. M., and Pak, B.: A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere, Biogeosciences, 7, 2261–2282, doi:10.5194/bg-7-2261-2010, 2010.

Wang, Y. P., Kowalczyk, E., Leuning, R., Abramowitz, G., Raupach, M. R., Pak, B., van Gorsel, E., and Luhar, A.: Diagnosing errors in a land surface model (CABLE) in the time and frequency domains, Journal of Geophysical Research: Biogeosciences, 116, 10.1029/2010JG001 385, doi:10.1029/2010JG001385, g01034, 2011.

Zhang, Y. Q., Viney, N., Chen, Y., and Li, H. Y.: Collation of streamflow dataset for 719 unregulated Australian catchments, CSIRO: Water for a Healthy Country National Research Flagship, 2011.

Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R. R., and Myneni, R. B.: Global data sets of vegetation leaf area index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the period 1981 to 2011, Remote Sens., 5, doi:10.3390/rs5020927, 2013.

Ziehn, T., Scholze, M., and Knorr, W.: On the capability of Monte Carlo and adjoint inversion techniques to derive posterior parameter uncertainties in terrestrial ecosystem models, Global Biogeochemical Cycles, 26, doi:10.1029/2011GB004185, 2012.

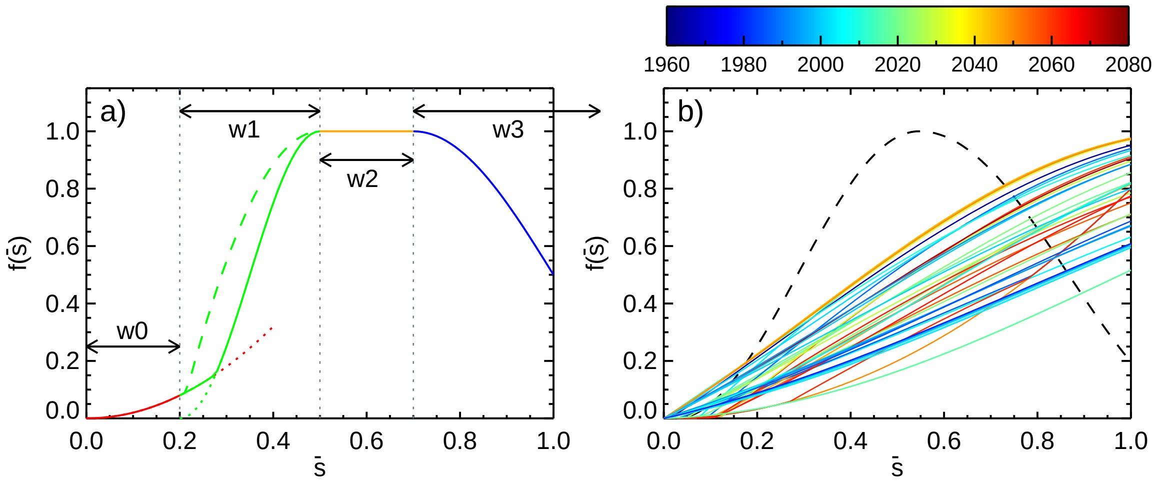

Figure 1. Function describing the heterotrophic respiration dependence on soil moisture, $\overline{s}$. a) Schematic figure to explain the form and parameters in the function in Eq. 2. Parameters used for the example shown by the solid line were $(q, c, w_0, w_1, w_2, w_3) = (0.3, 0.2, 0.6, 1.0, 0.2, 2.0)$. b) The dashed black line shows the function given in Kelly et al. (2000) for heterotrophic respiration dependence on soil moisture. The solid lines show our estimated function for the ensemble of parameters, colored by the corresponding value of Φ for both CABLE and CASA-CNP combined.

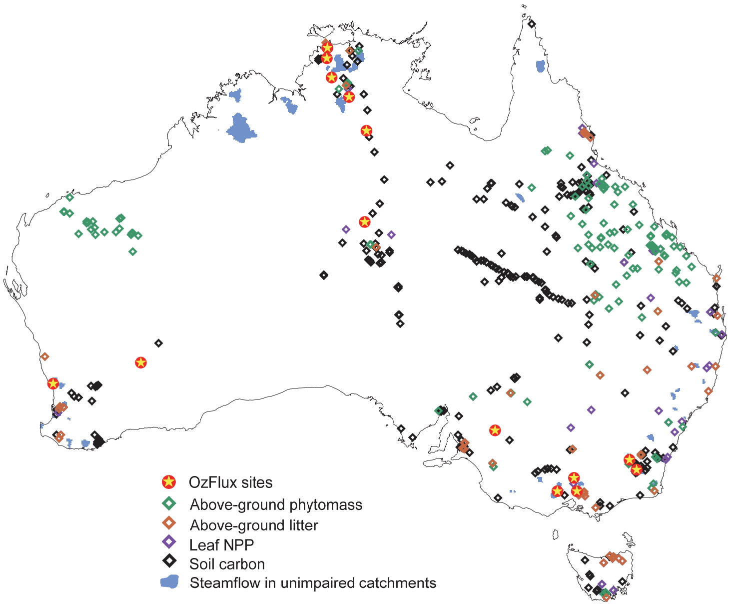

Figure 2. Location of observations used for calibration.

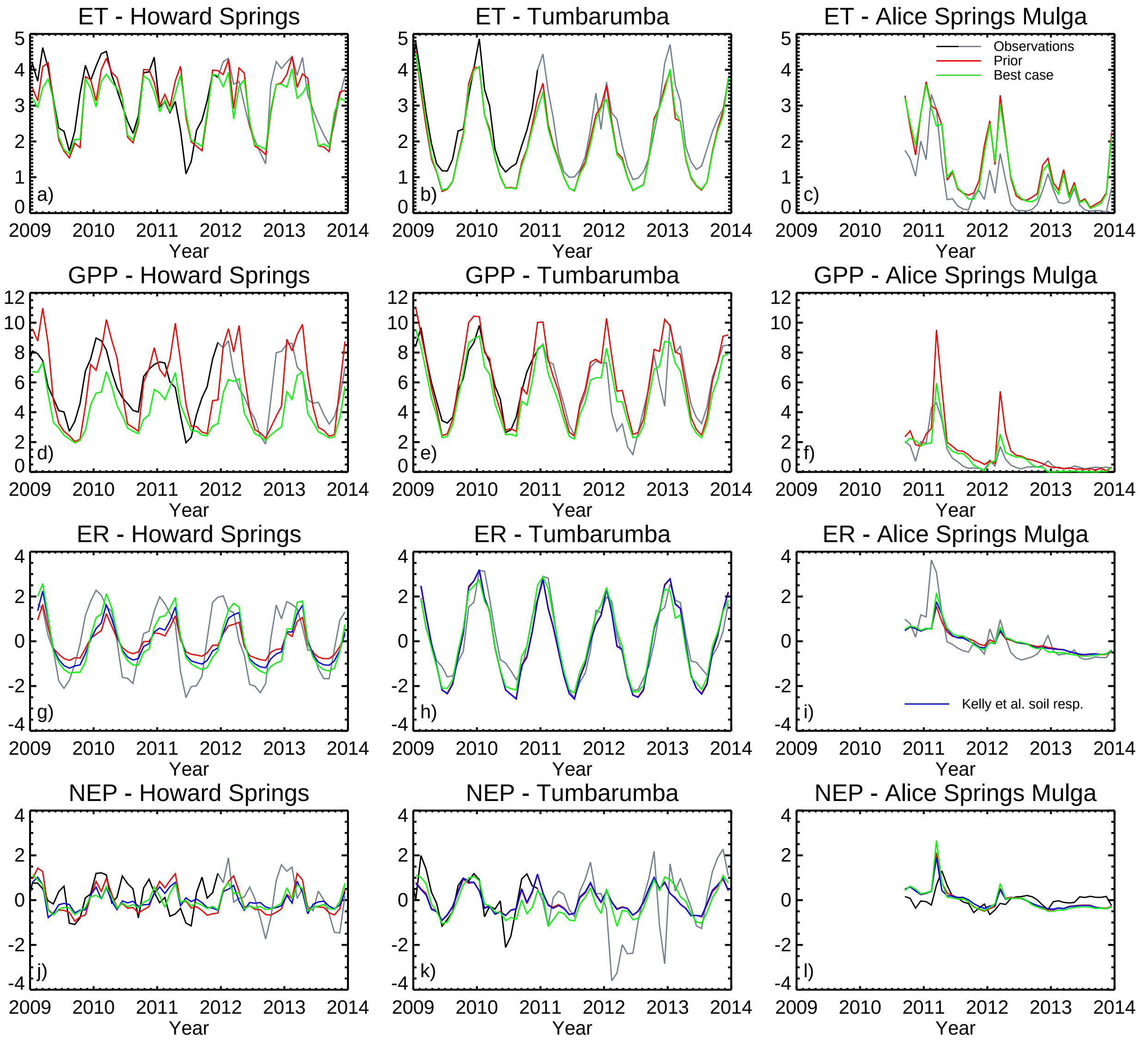

Figure 3. Monthly timeseries of ~~evaporation (~~ET ~~, (~~(mm d^{-1}), GPP (gC m^{-2} d^{-1}), ecosystem respiration (ER, gC m^{-2} d^{-1}) and NEP (gC m^{-2} d^{-1}) at Ozflux sites Howard Springs, Tumbarumba and Alice Springs Mulga. Black lines show observations used for calibration, grey lines show observations left for validation. Green lines show modelled quantities for optimised parameters corresponding to the lowest combined Φ for both CABLE and CASA-CNP. Blue lines (for ecosystem respiration and NEP) show the case using the Kelly et al. (2000) soil respiration function. Red lines show model quantities corresponding to prior parameters. Note that the ER and NEP from CASA-CNP calculated with prior parameters have used inputs from CABLE with optimised parameters, to indicate the effect of optimisation of CASA-CNP parameters only.

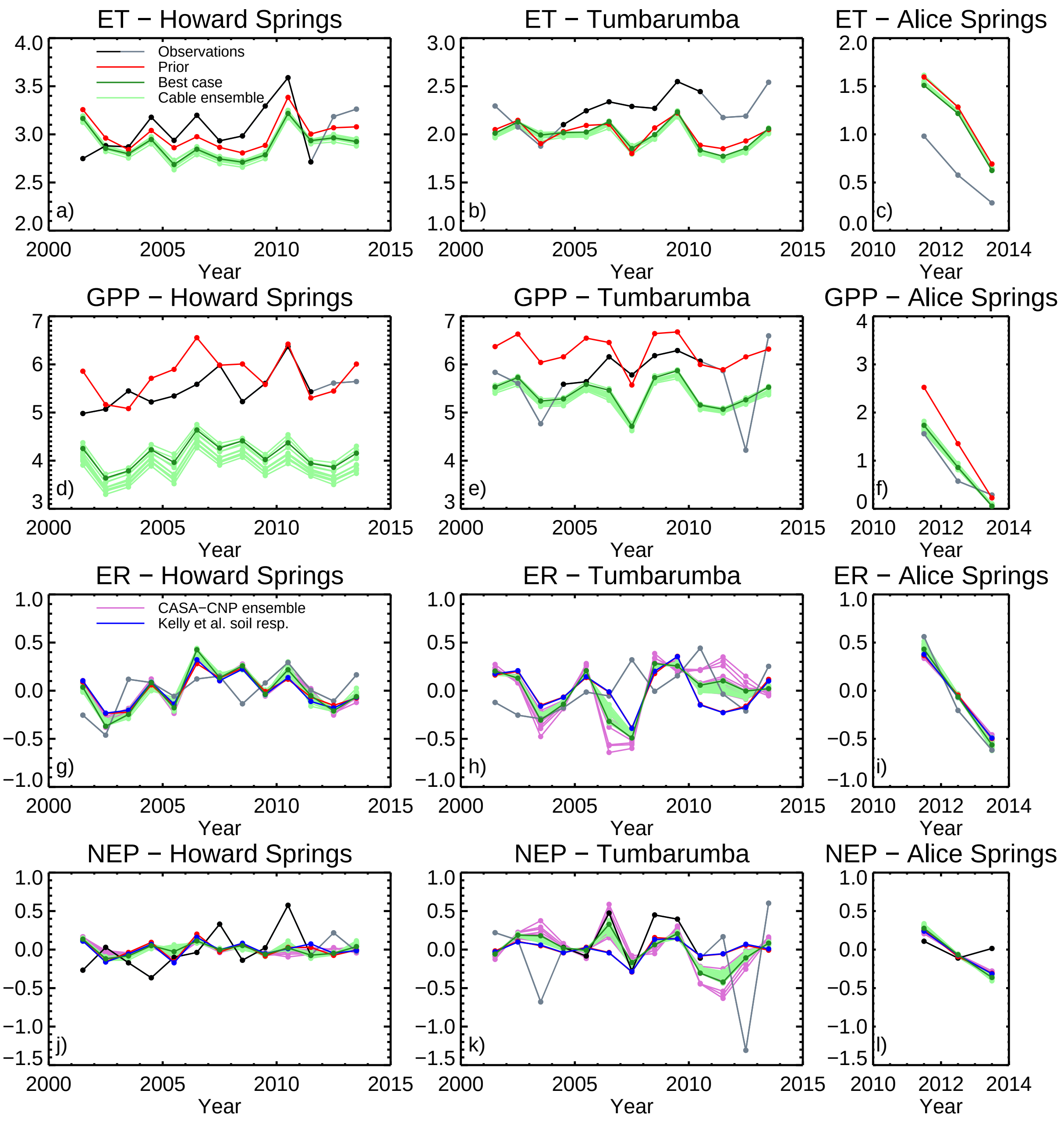

Figure 4. Annual mean timeseries of ~~evaporation (~~ET ~~,~~ (mm d^{-1}), GPP (gC m^{-2} d^{-1}), ecosystem respiration (ER, gC m^{-2} d^{-1}) and NEP (gC m^{-2} d^{-1}) at Ozflux sites Howard Springs, Tumbarumba and Alice Springs Mulga. Line colors are as in Figure 3 with the addition of light green lines for the CABLE parameter ensemble and pink lines for the CASA-CNP parameter ensemble.

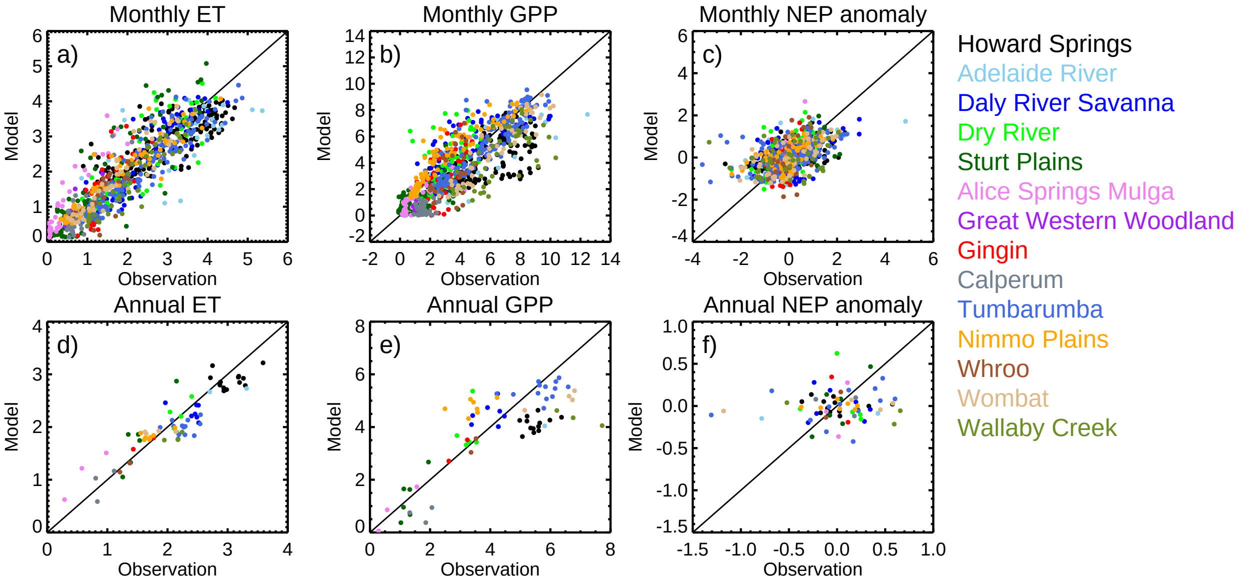

Figure 5. Scatter plots of observed and modelled (best case) monthly and annual ~~evaporation~~ ET (mm d^{-1}), GPP (gC m^{-2} d^{-1}) and NEP (gC m^{-2} d^{-1}) at 14 Ozflux sites. Symbols are color-coded according to site.

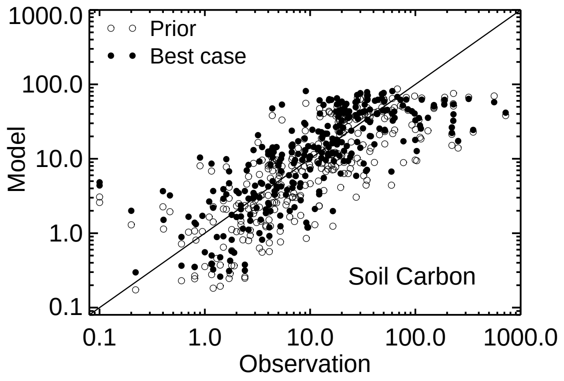

Figure 6. Scatter plot of observed and modelled (best case) long-term averaged soil carbon density in the top 15 cm.

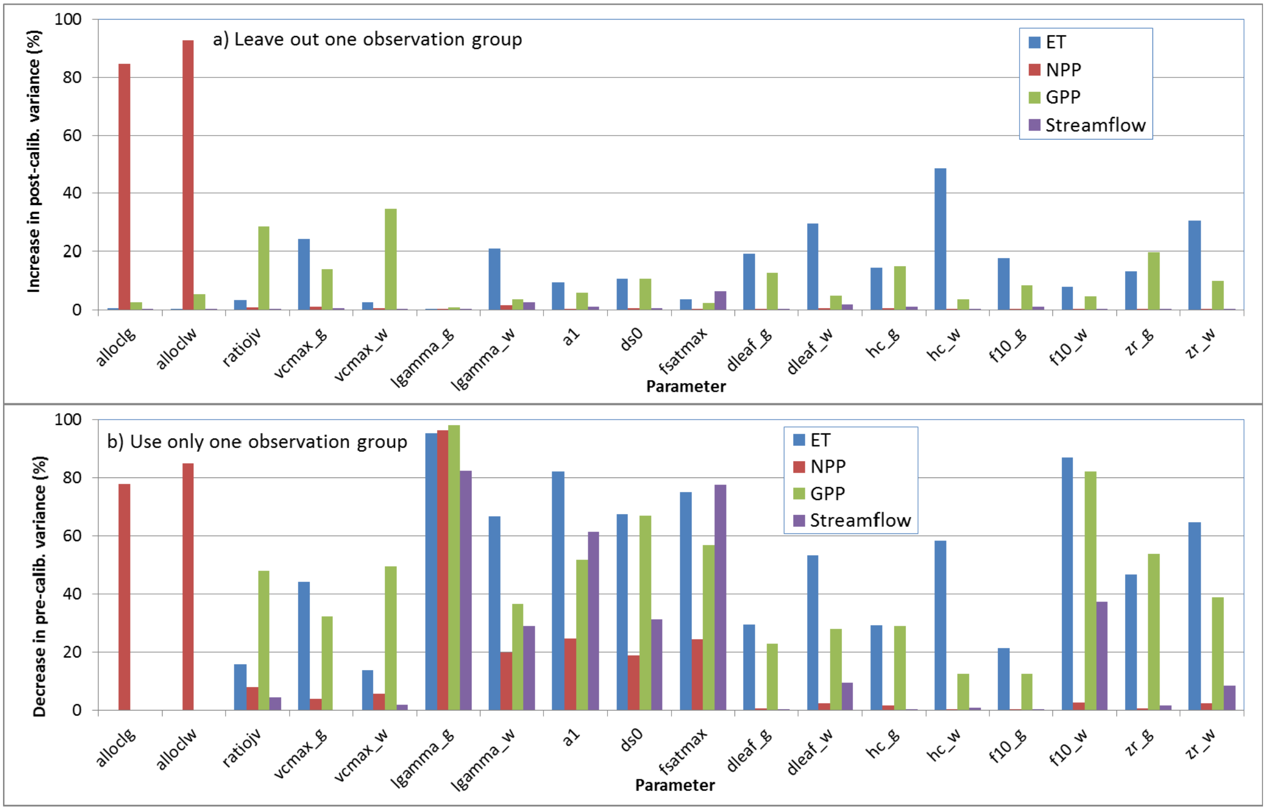

Figure 7. Worth of the different observation groups to the estimate of each CABLE parameter. a) Increase (%) in post-calibration parameter uncertainty variance incurred through loss of observation groups (i.e. leaving out each group in turn). b) Decrease (%) in pre-calibration parameter uncertainty variance incurred through addition of observation groups (i.e. each observation group is used on its own).

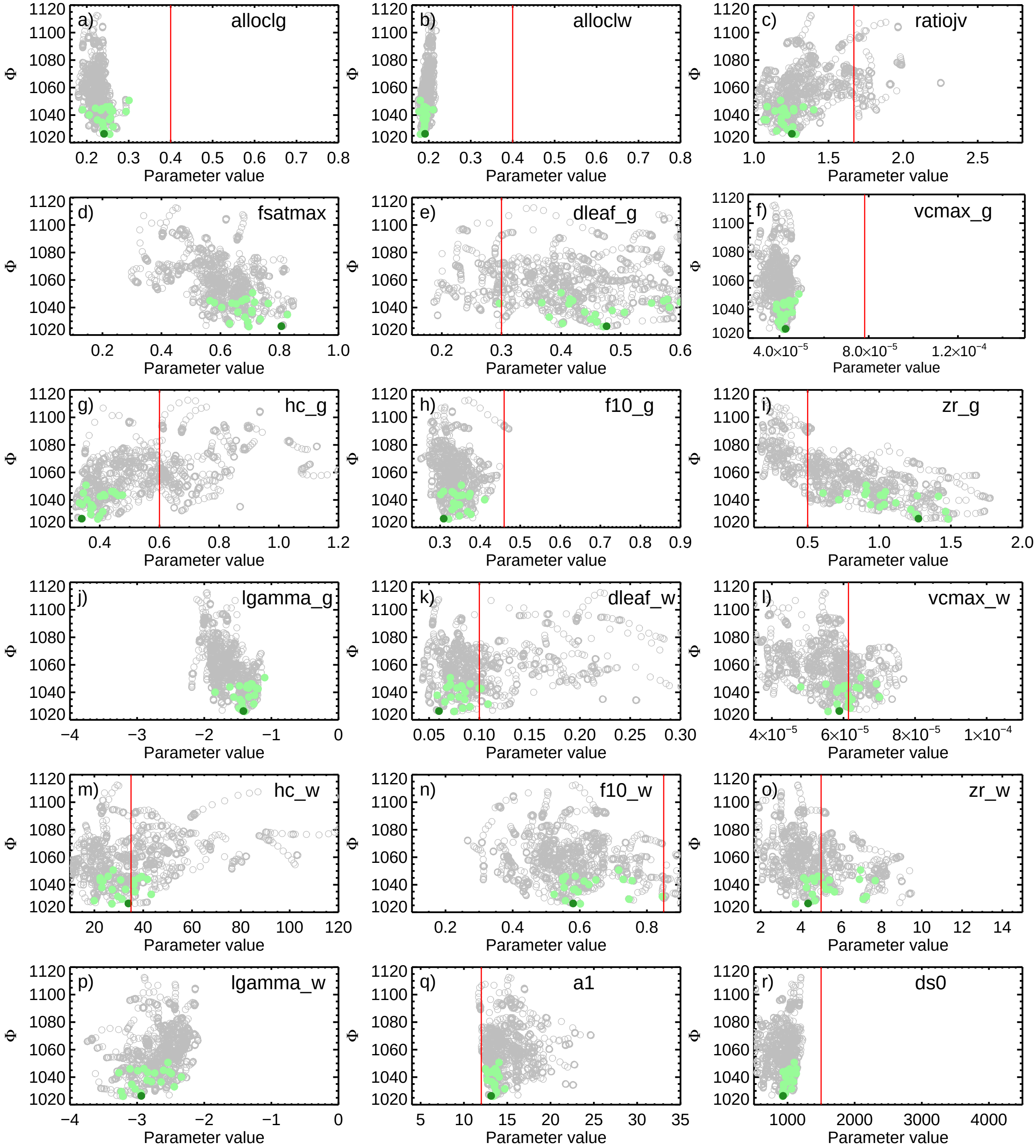

Figure 8. Model-data mismatch for CABLE observations, ~~Φ~~Φ_{CABLE}, plotted against parameter values for CABLE. Grey symbols are all parameter sets tested during the null space recalibration. Light green symbols ~~are~~ show the ~~18~~ CABLE parameter ~~sets selected to represent~~ ensemble plus the ~~parameter ranges and the~~ original CABLE optimisation. The dark green symbol shows the parameter set that gives the lowest combined Φ for both models. Red vertical lines show the prior parameter constraints.

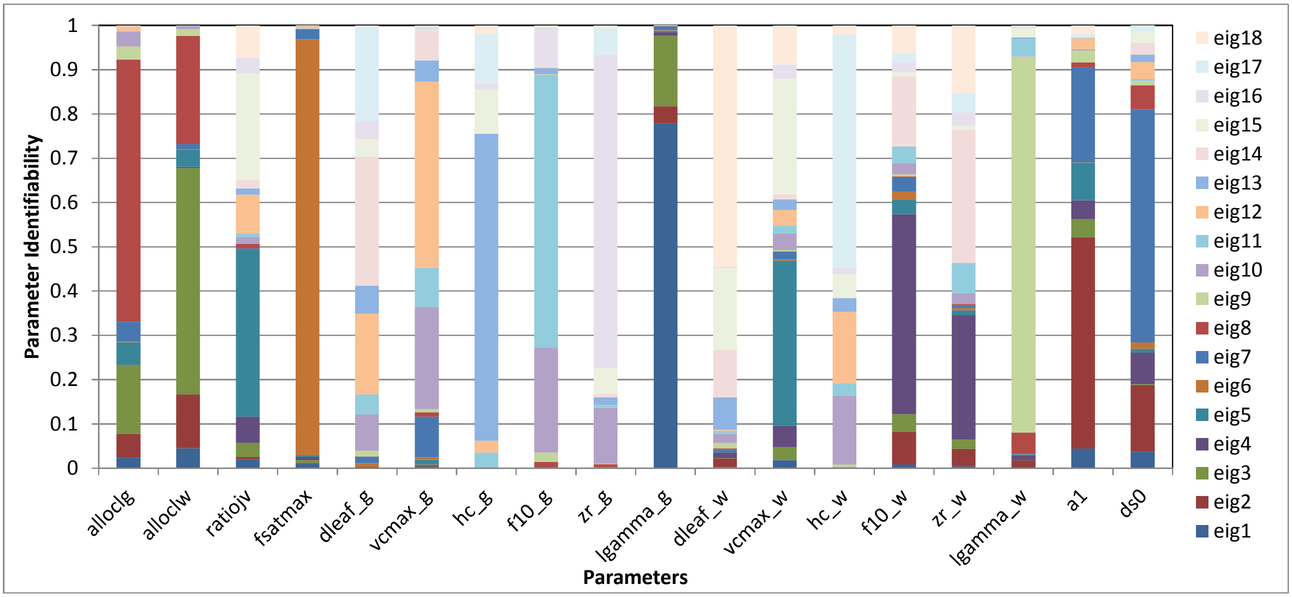

Figure 9. Parameter identifiability for CABLE parameters from PEST's linear analysis tools. Dark colors indicate eigenvectors that are more identifiable than light colors.

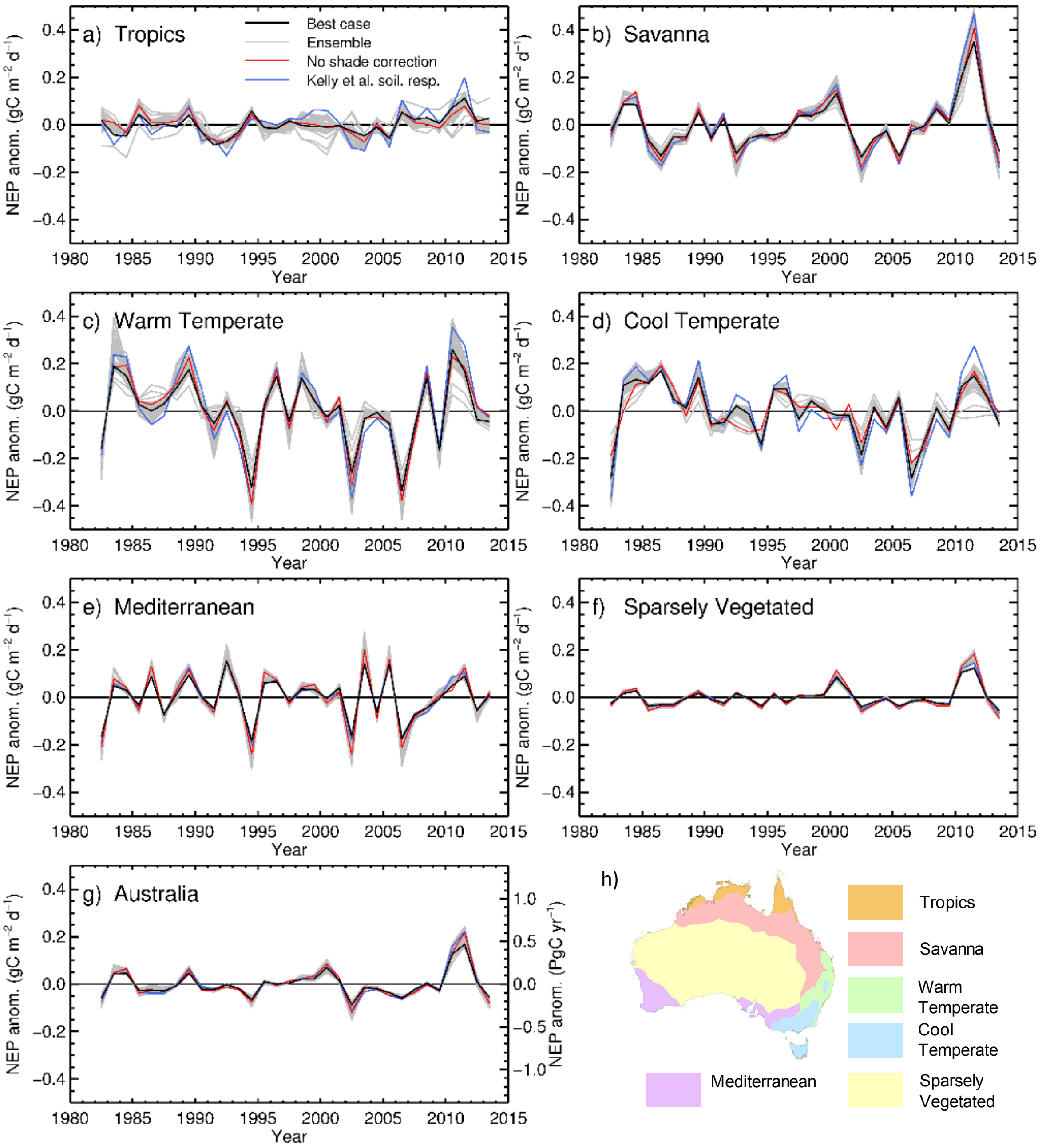

Figure 10. Annual anomalies in net ecosystem production for six bioclimatic regions and Australia. The best case is shown in black, with the other ensemble members in grey to indicate the influence of parameter equifinality. The red line corresponds to the case re-optimised without the correction to the vegetation cover for shaded grass. The blue line corresponds to the case re-optimised with the Kelly et al. (2000) soil respiration function. The y-axis range is the same in all panels. Units are gC m^{-2} d^{-1}, but panel g) also shows units of PgC yr^{-1} on the right. Panel h) shows the location of the bioclimatic regions.

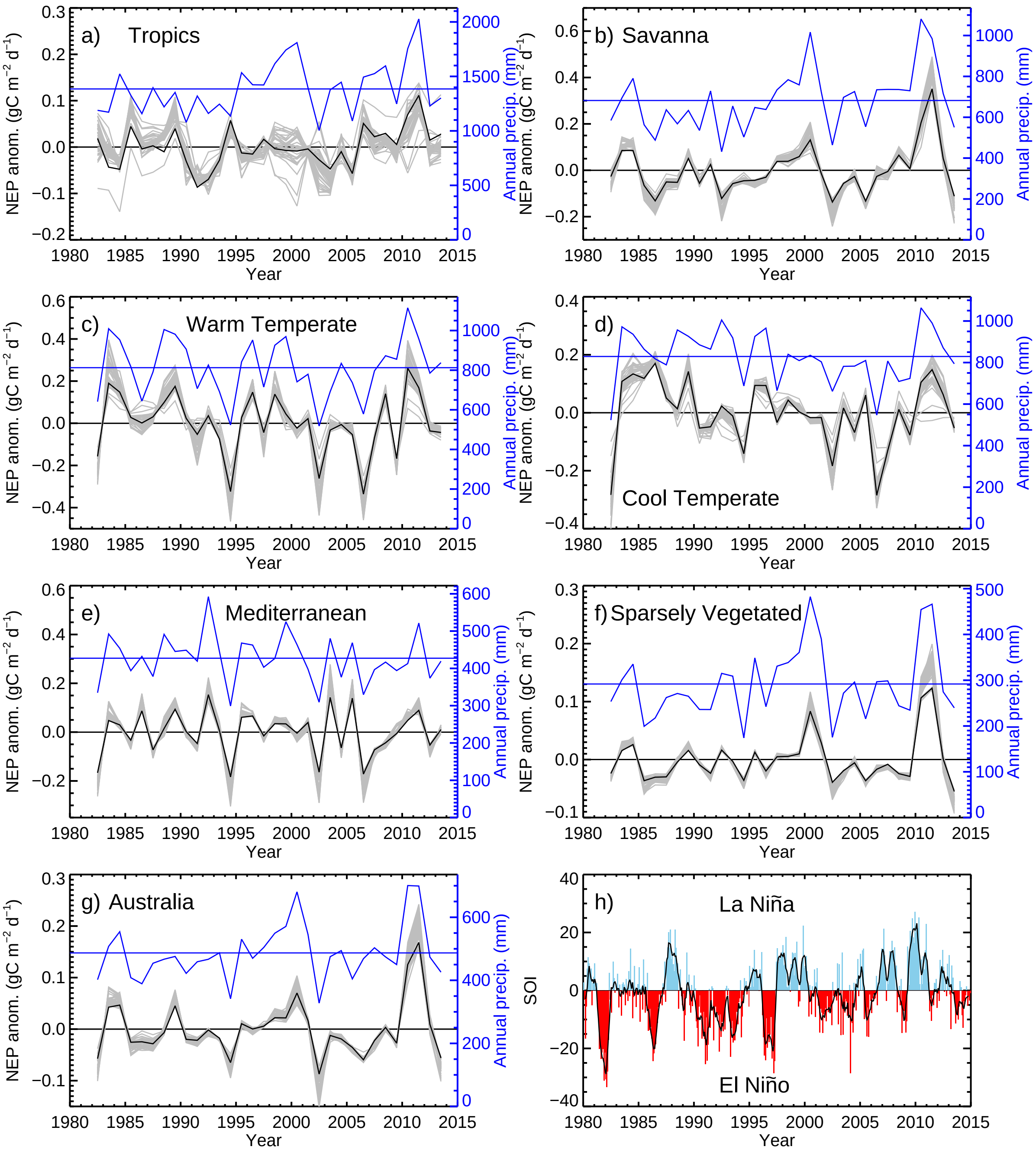

Figure 11. a) to g) Annual anomalies in net ecosystem production for six bioclimatic regions and Australia. Black lines show the case with lowest Φ and the grey band is the range due to parameter equifinality excluding the two outlier cases. Blue lines show annual precipitation (mm) for each region. Note that the y-axis range is different for each region. h) Southern Oscillation Index from http://www.bom.gov.au/climate/current/soihtm1.shtml.

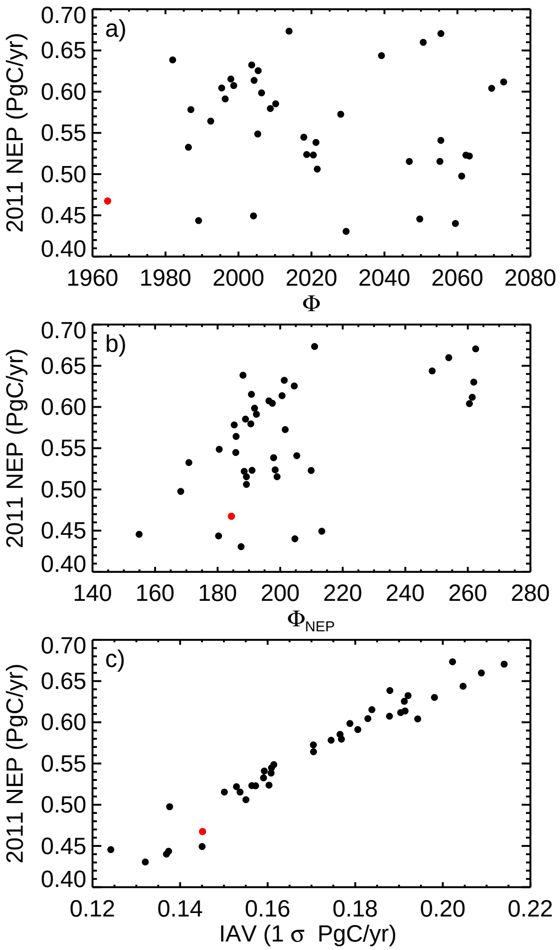

Figure 12. a) Australian NEP anomaly in 2011 relative to 1982-2013 for the ensemble of estimates plotted against the value of the total Φ from both models. b) Australian 2011 NEP anomaly plotted against the value of Φ from just the monthly NEP flux observations (Φ_{NEP}). c) Australian 2011 NEP anomaly plotted against the IAV for Australia over 1982-2013, expressed as 1σ. In all panels the best case (lowest total Φ) is shown in red.

Table 1. CABLE parameters to be tuned.

Parameter	Description
alloclg, alloclw	Allocation of C to leaves (grassy and woody)
ratiojv	Jmax/Vcmax
fsatmax	Multiplier for litter depth
dleaf_g, dleaf_w	Leaf length (grassy and woody)
vcmax_g, vcmax_w	Maximum RuBP carboxylation rate to leaf (grassy and woody)
hc_g, hc_w	Canopy height (grassy and woody)
f10_g, f10_w	Fraction of roots in top 10 cm (grassy and woody)
zr_g, zr_w	Maximum rooting depth (grassy and woody)
lgamma_g, lgamma_w	(log10 of) parameter in root efficiency function (grassy and woody)
a1	Parameter in stomatal conductance function
ds0	Sensitivity of stomatal conductance to VPD

Table 2. CASA-CNP parameters to be tuned.

Parameter	Description
soilc0_frac	Fraction of soil C in top 15 cm
age_leaf_g, age_leaf_w	Leaf turnover time (grassy and woody)
age_wood	Woody biomass turnover time (yr)
age_clitt1	Base metabolic litter turnover time (yr)
age_clitt2	Base fine structural litter turnover time (yr)
age_clitt3	Base coarse woody debris turnover time (yr)
age_csoil1	Fast soil C pool turnover time (yr)
age_csoil2	Slow soil C pool turnover time (yr)
age_csoil3	Passive soil C pool turnover time
fallocc_w	Fraction of non-leaf C allocated to wood
rsratio_g, rsratio_w	Fine root to shoot ratio (grassy and woody)
q,c,w_0,w_1,w_2,w_3	Six parameters to define function for effect of soil moisture on soil respiration

Table 3. Location and type of vegetation at the OzFlux sites (~~Beringer et al. (2016a)~~Beringer et al. (2016a); Isaac et al. (2016), http://www.ozflux.org.au) used in this study. Data period is for all of the data we have used at each site, with a subset of this data used for calibration.

Site	Coordinates	Ecosystem	Data period	Calibration period	Reference
1. Howard Springs	12.4952^{o}S, 131.1501^{o}E	Open woodland savanna	01/2001–12/2013	2001–2011	Hutley et al. (2005) Beringer et al. (2011)
2. Adelaide River	13.0769^{o}S, 131.1178^{o}E	Savanna	01/2007–05/2009	-	Beringer et al. (2007) Beringer et al. (2011)
3. Daly River Savanna	14.1592^{o}S, 131.3833^{o}E	Woodland savanna	01/2007–12/2013	2007–11	Beringer et al. (2011)
4. Dry River	15.2588^{o}S, 132.3706^{o}E	Open forest savanna	07/2008–12/2013	2008–11	Beringer et al. (2011)
5. Sturt Plains	17.1507^{o}S, 133.3502^{o}E	Open grassland	01/2008–12/2013	-	Beringer et al. (2011)
6. Alice Springs Mulga	22.283^{o}S, 133.249^{o}E	Mulga woodland	09/2010–12/2013	2010–2013	Cleverly et al. (2013)
7. Great Western Woodland		Woodland	01/2013–10/2013	-	
8. Gingin (Gnangara)	31.3764^{o}S, 115.7139^{o}E	Banksia woodland	01/2011–11/2013	-	
9. Calperum	34.0027^{o}S, 140.5877^{o}E	Mallee	01/2010–10/2013	2010–2013	Meyer et al. (2015)
10. Tumbarumba	35.6566^{o}S, 148.1517^{o}E	Cool temperate wet sclerophyll	01/2001–12/2013	2004–2010	Leuning et al. (2005) van Gorsel et al. (2007)
11. Nimmo Plains	36.2159^{o}S, 148.5528^{o}E	Grassland	01/2007–12/2013	2007–2011	
12. Whroo	36.6731^{o}S, 145.0262^{o}E	Woodland	12/2011–12/2013	-	Beringer (2013a)
13. Wombat	37.4222^{o}S, 144.0944^{o}E	Cool temperate dry sclerophyll	01/2010–12/2013	2010–2013	
14. Wallaby Creek	37.4262^{o}S, 145.1872^{o}E	Old growth temperate	08/2005–01/2009	-	Martin et al. (2007) Beringer (2013b)