

Revision of ACP manuscript acp-2016-254

“Atmospheric CO₂ observations and models suggest strong carbon uptake by forests in New Zealand”

by K. Steinkamp, S.E. Mikaloff Fletcher, G. Brailsford, D. Smale, S. Moore,
E.D. Keller, W.T. Baisden, H. Mukai and B.B. Stephens

1 December 2016

Dear Editors,

The co-authors and I would like to thank the anonymous referees #1 and #2 for providing constructive reviews. They helped us to improve and clarify the paper. We have responded fully to both reviews below, and we would be delighted to submit a revised manuscript.

We thank you for your continued consideration of this manuscript.

Sincerely,

Kay Steinkamp

Anonymous Referee #1 – Interactive Comment

Although the involved changes in the manuscript are sometimes minimal, my comments have been in most cases sufficiently well responded, and I recommend publication after a very small number of further minor changes (see below). I have read with great interest the comments of referee 2, as my own expertise is more limited, but won't comment on the related improvements in the manuscript. I note however that the new figure 14 is interesting and seems reassuring about the eventual systematic errors caused by the diurnal course; but the weakness and apparent randomness of the effect is somewhat surprising.

We thank the reviewer for his or her constructive comments. When we set up the inversion, we created a very small region around each site to minimise the impact of local variability at the sites on the larger scale flux estimates. Much of the bias created by the diurnal cycle in our synthetic data experiment is captured by our Lauder local area region, as discussed in the paper and in our response to reviewer #2 below.

There is one thing I would wish to be improved: in Table 1 (originally on page 44) an unclear division of both Islands into two regions is used; in their answer to this comment, the authors state that the definition of the division can be found by using Figure 10 etc.; but, first, the division of the

North Island is not described there; and, second, the South Island is apparently divided into two regions which overlap each other, which is odd but which is not further explained.

Some of the region labels in Figure 10 were indeed wrong, in particular such that the South Island West and East regions seemingly overlapped. We have corrected the labels to correctly reflect the division. Thank you for catching this.

A second point: Concerning my earlier comment on page 19 lines 4-29: The change proposed by the authors is good, but it has not been executed in the new version (see page 20 line 18-22 in that version).

That change was indeed not executed – we apologize and have executed it in the revised version

Anonymous Referee #2 – Interactive Comment

Steinkamp et al. have applied major changes to their manuscript "Atmospheric CO2 observations and models suggest strong carbon uptake by forests in New Zealand". I am not entirely satisfied with the answers they gave to my comments but I acknowledge that given the interesting material brought by this study, some of these comments could sound like a bit too meticulous. My main remaining issue is that I feel that the text (including the new and revised parts) is still often confusing and it often lacks of structure and rigor. I also feel that the authors sometimes provide weak answers in order to avoid accounting for my comments and to avoid new analysis and discussions (in particular new analysis of the modeled vs measured concentration timeseries). Finally, a significant number of my comments (either comprised within my major comments or in my list of minor comments) were misunderstood. In conclusion I would still push for a major and general improvement of the text but the scientific issues I would still raise are minor.

We thank the reviewer for this feedback. We regret that the reviewer feels we have attempted to avoid new analysis or discussion in response to his or her comments. This was not our intention. For example, the addition of an OSSE represented a major piece of new work for the paper, which we felt was appropriate given the reviewers serious concerns about bias from diurnal cycles.

In response to Referee #2's ongoing concerns about the quality of the text, we have asked a colleague who is not a co-author on the paper to review our manuscript with fresh eyes and provide suggestions about where the text could be improved. His suggestions have been implemented in the revised manuscript, in the absence of further feedback from the reviewer.

Here is a list of specific problems I see in their answers and in the

corresponding corrections to the manuscript:

- Regarding the test of sensitivity to the NEE diurnal cycle: the most straightforward test we could have expected was a new experiment using the real data and the reference inversion set-up, but the prior estimate of the NEE with diurnal variations. Comparing the reference inversion to this one would have provided a clear characterization of the impact of this diurnal cycle in the reference inversion results, and an evaluation of the confidence in these results. I do not say that the authors had to make this specific experiment, but their OSSE and of their analysis to make a demonstration of the weak impact of the diurnal cycle can sometimes be a bit puzzling:

The reviewer seems to misunderstand our inverse modelling framework. We minimise differences between weekly prior fluxes and posterior fluxes in the cost function (Equation 2). Since the a priori flux constraint is applied on a weekly basis, adding diurnal variability to the prior would have no impact on the results of our inversion, unless the weekly mean flux were also changed.

While the reviewer did not request we undertake an observing system simulation experiment (OSSE), we felt this was the most comprehensive way to evaluate potential biases introduced by neglecting diurnal variability.

** I am not 100% sure about the way they derived the diurnal cycle of the synthetic NEE: "the diurnal variation in GPP is based on hourly solar insolation": does it mean that it is exactly set-up based on the relative diurnal variations of the solar insolation ? I cannot see numbers illustrating the amplitude of this cycle (so it is difficult to check that it is "exaggerates" actual diurnal variations).*

The reviewer is correct that in our experiment the GPP varies exactly according to the diurnal variation in solar insolation. This exaggerates the diurnal variation in NEE fluxes because, in reality, photosynthesis rates reduce at midday and respiration rates increase with temperature so that the peak amplitude is lower than would be implied by the effect of solar insolation alone. The amplitude of our simulated diurnal cycle is 1331 TgC yr⁻¹ in summer (Dec-Feb average) and 690 TgC yr⁻¹ in winter (June-Aug average) integrated across all New Zealand land regions. We also show it in figure 1 below, but have not added this figure to the manuscript in order to avoid excessive length.

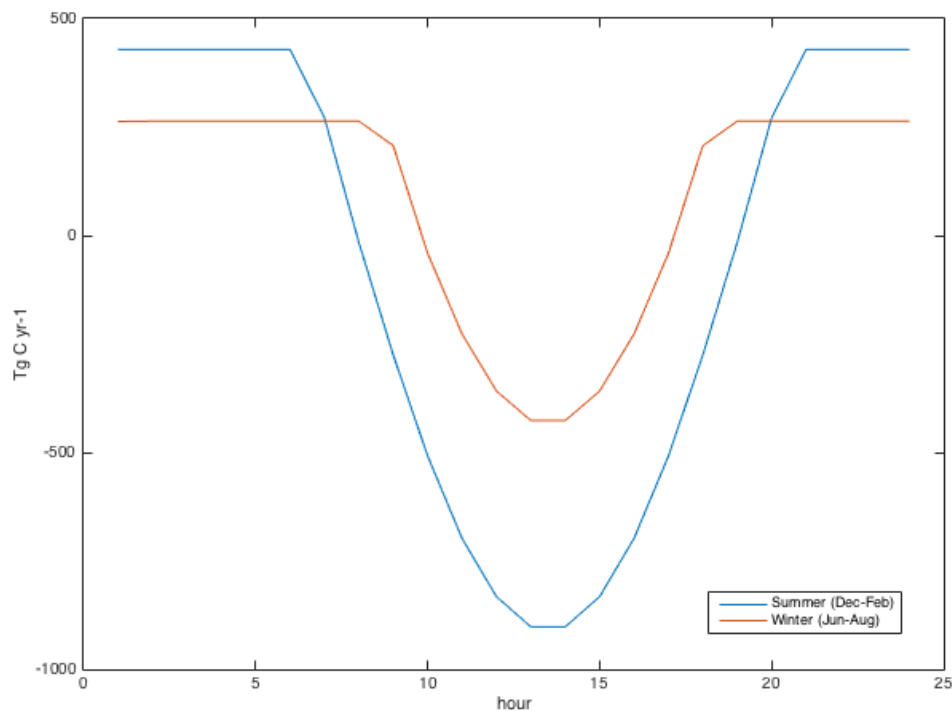


Figure 1. Diurnal cycle of CO₂ fluxes during summer (blue, Dec-Feb average) and winter (red, Jun-Aug average) integrated across all New Zealand, estimated for our synthetic data experiment. Uptake is negative.

In order to clarify this point, we have changed the text on p.30 lines 9-10 as follows.

"The diurnal variation in GPP is directly proportional to the relative amount of hourly solar insolation, and HR is assumed to occur at a constant rate spread evenly throughout the day. There are a number of aspects of plant physiology and ecosystem biogeochemistry that cause actual diurnal variation in NEE to be more muted than the solar-radiation driven pattern we have modelled. These include reductions in photosynthesis during the middle of the day and afternoon as stomata close due to drought or leaf temperature stress. Similarly, reduced respiration can be expected at night due to cooler temperatures. "

** the authors connect this OSSE to the discussion on the model-data mismatch at LAU for 15:00-16:00, but in order to assess whether uncertainties in the NEE diurnal cycle impacts these mismatch, they could had looked at the 15:00-16:00 concentrations modeled when using the flat NEE vs. the one with diurnal variations (instead of trying to read such a comparison out of the inversion results). As a consequence, they assimilate 15:00-16:00 data in the OSSE instead of both 13:00-14:00 and 15:00-16:00 as in the reference experiment. This and having several other parameters that are different in the OSSE compared to the reference inversion does not help evaluate the robustness of the reference inversion with regard to uncertainties in the NEE diurnal cycle.*

There are two ways to look at the impacts of the diurnal cycle in the framework of our synthetic data experiment: 1) compare the modelled mole fractions with and without the diurnal cycle (data space); 2) compare the posterior fluxes with and without the diurnal cycle (flux space).

We chose to undertake the analysis in flux space, because it allows us to quantify potential biases in our flux estimates, which are the central results presented in the paper. We appreciate the referee's suggestion that we undertake this comparison in data space instead, but we feel the flux comparison is more relevant to our final result. In order to satisfy the reviewer's request without making the paper excessively long, we present analysis in data space (below) but omit it from the revised paper.

In Figure 2, we show the difference that the diurnal cycle makes to simulated atmospheric CO₂ at Baring Head and Lauder. At Baring Head, the bias is small compared to the variance, and there is some seasonal signal, with lower XCO₂ in (austral) summer, when the diurnal cycle is resolved. In flux space, this should translate into 1) little overall bias in regional annual mean fluxes and 2) slightly increased CO₂ uptake in summer. Both of these conclusions have also been reached at in our discussion in the manuscript. Furthermore, by doing the analysis in flux space, we could go into a bit more detail, such as pinpointing the regions in which such a seasonal impact is recognizable.

At Lauder, we can see a clear positive bias along with a very small seasonal signal. The fact that the bias is positive and not negative, as one would expect when including diurnal flux variations (at least in summer), is discussed in the manuscript as well. That bias translated to a positive flux anomaly in flux space. As seen from Fig. 14 in the manuscript, that positive anomaly is mostly contained in the local region around Lauder (region 14). Capturing such local biases was part of our motivation for including the local regions around measurement stations.

To summarize, we believe undertaking the analysis in flux space provides more detail and a more direct link to the results of the reference inversion, while also capturing the information obtainable from a comparison in data space.

As for the second part of the comment, we did not include a comparison for 13:00-14:00 data due to the computational cost associated with the new, hourly footprints on top of an already substantial piece of additional work. We chose to use the 15:00-16:00 time slot for our analysis, as it is closer to the observed minimum XCO₂ in the afternoon and should be affected more strongly by the lack of diurnal flux variations in our reference model.

We also would like to emphasize that for our synthetic experiment, we kept as many parameters as possible the same as in the reference case. However, it was unavoidable to make some changes, e.g. the new dataset of hourly fluxes does not have the same weekly averages as the prior

used in the reference inversion, so we had to use the new flux map in both synthetic runs for consistency.

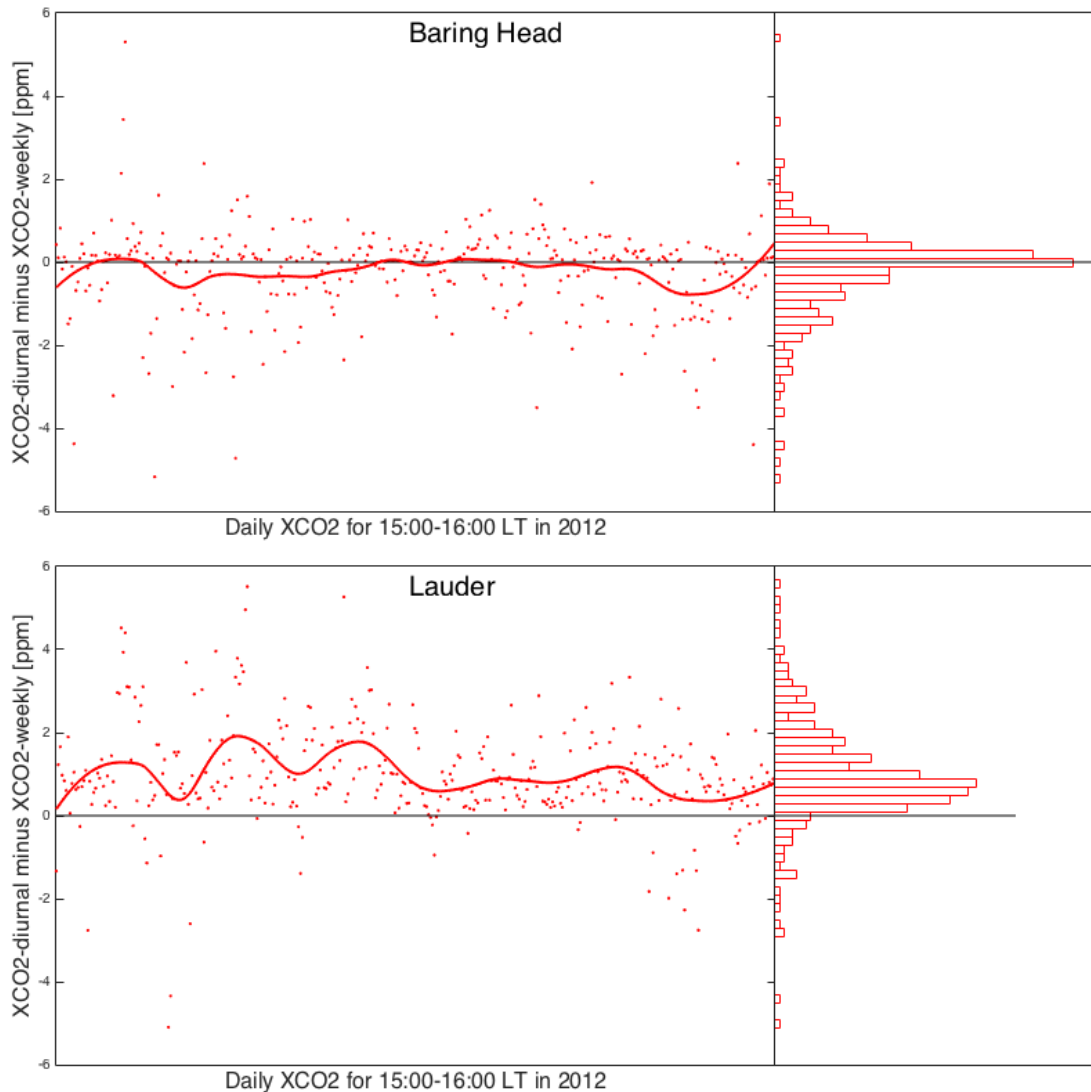


Figure 2. Data mismatch for synthetic experiment with and without a diurnal flux cycle. Solid lines are Loess fits with a 3 month window.

the analysis added in 7.4 to comment on the impact of uncertainties in the diurnal cycle based on Figure 14 are often confusing but yes, Figure 14 seem to indicate that this impact is small at the annual scale. The much larger impact on results at the monthly scale and for the seasonal cycle could have been more emphasized (while the claim by the author that the truth and the inversion agrees within their uncertainty is misleading). At least, for regions 7, 12, 14 and 15 it is clearly significant and often larger than the other sources of uncertainties accounted for in the estimate of the posterior uncertainty, which shows that this uncertainty should be increased to account for this additional source of error at the monthly scale.

We discussed the higher weekly variability in regions 12, 14, and 15 in paragraph 3 of the Diurnal Variability subsection of 7.4 in the previous submission. However, In response to the reviewer's concerns, we have revised the text of section 7.4 as follows. New text is shown in *itallics*.

Unrepresented diurnal variability led to biases in the annual mean flux estimates for some regions in our inversion, but these errors were much smaller than our uncertainty estimates for most regions *on an annual scale* (Figure 14)...

On a weekly time scale, estimates generally agree to within their uncertainties for most regions, with the exception of the Eastern South Island (Regions 12 and 15), the southern Central North Island (region 7), and the local area region around Lauder (region 14.) Regions 12 and 15 (eastern and south-eastern South Island) show an increased sink late in the year (as part of the 2012/2013 summer) but a smaller sink early in the year (as part of the 2011/2012 summer), suggesting that there may be a seasonal bias in our inverse methodology for the eastern South Island. Likewise, diurnal cycle bias leads to significant weekly errors in the central North Island (region 7), although with less seasonal coherence. The Lauder local region (region 14) was created to capture local signals that are not well represented in our inverse model, including diurnal variability, and prevent them from biasing the inverse estimates on larger spatial scales. Thus the larger errors for this region are expected.

- Regarding the need for a general improvement of the text: as I explicitly said, I only listed examples illustrating such a need in my previous review, but I did not conduct an exhaustive listing of the problems. However, the authors "sought to improve the quality of the text in places specifically pointed out by the referee" when they did not just rebut the corresponding concerns. As a consequence, they did not conduct the detailed proofreading which was needed, and which is thus still needed.

We regret that the reviewer feels the revised manuscript was poorly written, as we have made an earnest attempt to put forward a high quality manuscript. We have asked a colleague who is not a co-author on the paper, Dr. Hinrich Schaefer, to provide an independent review of the writing for clarity, rigor, and structure. He has pointed out a number of places where the text could be improved or clarified, and we have implemented these changes throughout the revised manuscript, but he did not recommend any major structural changes to the text. We hope our revised manuscript addresses the reviewer's concerns on this point.

- I am still a bit puzzled regarding the topic of the model error. I am really confused by the authors' explanations regarding the precise computation of the model error in page 16-17 and how it can be consistent with what is said in page 18 regarding the 0.4ppm value. The text does not give the typical value for the model error arising from such computations. Furthermore, the authors refused to conduct some more detailed analysis

of the prior and posterior model vs measured concentration timeseries (which would have been useful to discuss the theoretical value that they derive for the model error), and more generally to conduct some additional evaluation of the transport model, assuming that "a more in depth analysis of transport model bias is outside the scope of this paper".

We agree that the phrasing on page 18 can be improved. As detailed on page 16-17, the 0.4 ppm refer to the minimum uncertainty, and hence are only part of the final uncertainty. The latter is calculated as described further on page 16-17.

We have clarified this in the text on page 18:

Computing the data uncertainty as described above ensures $X^2 \approx 1$, which means (...)

In addition, we have added a sentence on page 16-17 to provide an average value and a range of the final uncertainties:

The resulting uncertainty is taken as the root mean square (quadrature) of both components and has a minimum value of 1.16 ppm. The mean uncertainty is 1.91 ppm and 95% of the values are within the 1.16 to 4.56 ppm range.

-My comment on the weakness of the configuration of the sensitivity test has not really been accounted for even though these tests are used to provide an assessment of the robustness of the inversion results

We respectfully disagree. This point was answered in our original revision and response letter by addressing specific questions the reviewer raised in the response letter and expanding discussion of errors due to transport model uncertainty and unaccounted for diurnal variability in the first revised manuscript.

- Some technical points:

** the mathematical notations are still problematic, e.g. T_g and T are multiplied by the same vector on page 9 or the dimensions of x and T change from page 9 to 17*

In order to prevent potential confusion regarding equations defined on the model grid with those defined on aggregated regions, we now distinguish between fluxes \mathbf{x}_g (on the grid scale) and \mathbf{x} (aggregated regions). The respective text on page 9 now reads:

With \mathbf{x}_g being a vector containing all grid cells and \mathbf{c} a vector containing the concentration (unit $\text{g CO}_2 \text{ m}^{-3}$) for all 1 h periods, this is written as $T_g \mathbf{x}_g = \mathbf{c}$. Given T_g and the measured concentrations \mathbf{c} , the aim is to solve for the CO_2 fluxes \mathbf{x}_g using a Bayesian inversion, i.e., a statistical model that balances information from measurements with *a priori* knowledge about the fluxes (section 6).

Instead of solving on the grid scale, the fluxes in \mathbf{x}_g are pre-aggregated into a set of regional fluxes (section 5.2), \mathbf{x} , and a transport matrix T is created by aggregating grid cells in T_g to reflect the regions in \mathbf{x} ,

$$Tx = c \quad (1)$$

The Bayesian inversion developed here solves for \mathbf{x} . In addition, *a priori* flux maps are taken into account for the terrestrial and oceanic portions of the domain (section 4).

**I do not understand the justification for the S term in equation (2) based on a matter of "transparency".*

As we described as part of our response to the original comment about the S term, we think the ability to quantify the strength of the smoother with respect to the other terms in the cost function is a transparent way of introducing this additional constraint. We do not say that imposing temporal correlations instead cannot be transparent, but merely that formulating the smoothing constraint as an explicit term in the cost function makes a transparent interpretation of its impact straightforward.

**I think that the authors misunderstood my point regarding their estimate of the representativeness error. I just think that strictly speaking, the 5-min scale variability of the measurements does not reflect the spatial variability of the hourly mean measurements in model box.*

We agree on this. We did not intent to suggest the spatial variability in the model can be represented by the 5-min data variability. Rather, this 5-min variability can be used to estimate error related to differing temporal resolutions of model (1 hour) and data (5 minutes). We have rewritten the respective sentences on page 6:

For both stations, one standard deviation of a 5-minute measurement interval is taken as random data uncertainty for the hourly mean. This uncertainty is generally much greater than the measurement precision, as it reflects real atmospheric variability, and is instead intended to capture representativeness errors such as different temporal resolutions of model and data or the model failing to represent the specific conditions at an individual location.

** Regarding the meaning of the correlations between posterior uncertainties in two flux components: the author insist in writing something wrong i.e. "strong negative correlations between two regions would indicate that the inversion has difficulties to distinguish their individual flux components with the available data". Again, if the uncertainty reduction for the flux in each region is about 99% but the remaining (posterior) insignificant uncertainties in each region have a -0.9 correlation, we would still have to say that the inversion managed to distinguish their individual flux components.*

We have difficulties following the reviewer's logic here; suppose a posterior correlation matrix for two flux components has (i) a -0.9 off-diagonal value, (ii) a -0.1 off-diagonal value. No matter what the variances are, the sum of both flux components will have a lower uncertainty in case (i) than in case (ii). We agree that with stronger uncertainty reduction the significance of the correlation diminishes in the context of distinguishability of the flux components, which is why we weakened our statement in response to the original comment. However, it is still a valid statement. The reviewer seems to suggest there is some kind of threshold, i.e. that once the uncertainty reduction is large enough (the reviewer mentions 99%) the distinguishability issue disappears completely, which is not true. In any case, in the context of our study, typical uncertainty reductions are in the range of 30% to 60%, except for the local regions around the sites where it can reach 80%, arguably not high enough to justify the claim that the inversion can resolve all regions perfectly. One reason we added that statement in the manuscript was to make it clear that we do not claim our inversion to perfectly distinguish all flux components.

** in general: the authors easily use the term "bias" when discussing random or varying errors*

We have examined each use of the term bias in the text, and replaced it with 'error' where appropriate.