Atmospheric
Chemistry
and Physics

Discussions

Interactive comment on "Atmospheric CO$_2$ observations and models suggest strong carbon uptake by forests in New Zealand" *by* K. Steinkamp et al.

Anonymous Referee #2

Received and published: 30 May 2016

The paper "Atmospheric CO2 observations and models suggest strong carbon uptake by forests in New Zealand" describes a regional atmospheric inverse modeling framework for the estimate of the weekly to annual mean CO2 land ecosystem fluxes in New Zealand. It evaluates the results using comparisons to estimates from the national inventory.

- General comments

The scope of the manuscript, the atmospheric measurements, the modeling and inversion frameworks as well as the detailed discussion on the comparison between the inversion results and the inventory should, in principle, make this study worth a publication in ACP.

However, there is likely a critical issue regarding the lack of account for the diurnal cycle of the CO2 land natural fluxes both in the prior estimate of these fluxes and during the inversion. The authors need to demonstrate that the impact of this lack of account is not problematic, clarify that they had actually accounted for this if it was the case, or should run new experiments to account for this (see my main comment A)). Furthermore, some analysis are rather weak and the quality of the text critically needs to be improved.

Mainly

A) The actual temporal resolution of the prior estimate of the CO2 land natural fluxes is 1 month. The inversion adjusts 1-week mean fluxes. From my understanding, there is thus no account for the diurnal cycle of these fluxes both in the experimental framework nor in the analysis and discussions of the text. This can be a critical issue given that the system assimilates afternoon data only, and given that, due to the configuration of New Zealand, these data should be primarily influenced by the afternoon, and, to a lesser extent, morning fluxes. By ignoring the diurnal cycle of the CO2 land natural fluxes, the prior would underestimate the natural sink during the afternoon. Consequently, the assimilation of data during the afternoon only would lead to a strong increase of the afternoon sink, which, due to aggregation errors, can be artificially extrapolated into a abnormally high increase of the weekly to annual mean sinks and of the seasonal cycle. Since this is exactly what is observed in the inversion results, this can strongly weaken the confidence in these results. An annual sink of 0.1PgC for NZ is a bit surprising so the authors will definitely need to better support such a number. If the authors actually accounted for this, they should clarify it. If not, they should investigate it and it may be necessary for them to rerun the inversion by separating the adjustments for night-time and daytime fluxes. This problem of the lack of account for the diurnal cycle of the CO2 natural fluxes can be connected to the lack of account for past CO2 regional inverse modeling studies in the introduction and method sections (see my major comment on the introduction: the paper seem to rely on techniques and knowledge from

CH4 inverse modeling experiments, ignoring the potential specificities of the inverse modeling applied to regional CO2 natural flux estimation). In this context, the text at lines 16-20p18 is really embarrassing (in addition to being very confusing).

B) The logic, order and rigor of the text needs to be strongly improved; it would be difficult to detail all the issues corresponding to this comment but I try to list some representative examples below. The order of the figures is not consistent with that of the references to these figures in the text and it often looks like random.

I also think that there is a significant mistake regarding the mathematical framework and configuration of the inversion since I believe that the chi-test that has been used to set up the model errors is wrong. This should have been 2J/n where n is nobs the number of obs and not nobs-nflux (the authors seem to have confused the chi test for J and that for "Jobs" i.e. the part of J corresponding to the misfits to the obs). I assume that this comes from the fact the authors were confused about the type of cost functions analyzed in the ref (Gurney et al 2004 and Baker et al 2006) that they provide for this test. In any case, the way this test is presented is confusing and lacks of rigor.

C) The different focuses of the text could be better balanced. In particular the evaluation of the inversion is a bit short while there is much material on the footprints of the measurements that is not really exploited for such an evaluation. The skill of the transport model for simulating the concentration at the measurements sites is hardly analyzed even though the topography seems quite complex around the sites and the measurements are taken at 10magl only.

D) I feel that there is sometimes a sort of over-confidence in the concept of baseline and in that of constraining the fluxes of different regions using the different measurements sites independently (while a traditional concept of the atmospheric inversion is to exploit gradients between sites to infer fluxes between these sites), and I think that it should be better discussed and weighted when analyzing the results.

The idea of a very slowly varying baseline concentration applying to a very large por-

tion of the model boundaries (the background sectors) may often be unadapted. I do not contest the use of such a concept to solve for the boundary conditions in this study since these conditions are difficult to deal with and this practical solution is among the relevant ones. But the paper could better discuss it and more carefully analyze it in the CO2 data timeseries. CO2 simulations over large domains at high resolution (e.g. https://www.youtube.com/watch?v=x1SgmFa0r04) indicate that synoptic patterns of CO2 can travel over large distances which hamper the concept of baseline. In particular, the influence of CO2 sources and sinks in Australia (not that of the portion included in the domain only) could generate large variations in the concentrations measured in New Zealand at synoptic scales even though it is at more than 2000km from New Zealand. In principle, regional inversion frameworks using dense networks where, for most of the wind directions, parts of the measurement sites are located downwind some of the other sites (with the targeted fluxes located in between) limit the impact of uncertainties in the boundary conditions to the margins of the observation network. However, here, the network seems to have been deliberately set up so that the stations work independently to target different areas of New Zealand (and the text on p3 l. 20 only sees the use of several stations as an "advantage").

The paper considers a baseline uncertainty whose exact computation is quite impossible to understand in the main text (l20-21p13) and in appendix B (B1 does not mention any estimate of uncertainties for the southern baseline, B2 is quite confusing). It is difficult to understand what the authors aim at characterizing with this uncertainty. However, some statistics on the timeseries analyzed for such a computation could help to assess the robustness of the baseline concept. The test of sensitivity to a 0.1ppm bias in the baseline ((i) page 18) does not evaluate the weakness of the concept of baseline that I discuss here as illustrated by the solutions proposed in conclusion for tackling such a bias.

The study could better characterize whether there could be some weeks or month when results could be less robust due to the influence of fluxes from Australia or from

other areas (e.g. by comparing the filtered baseline to the actual measurements when the sensitivity to New Zealand fluxes is relatively low, or when large winds blow from a background sector; and by using the analysis of the footprint and the timeseries of the model-data misfits). All of this should be better introduced when presenting the technique (section 5 is often confusing), and better analyzed and discussed in the result and discussion sections.

- Major comments by section:

1) Abstract

The authors could give some insights on the confidence in the inversion method and estimates; this would help better end this abstract (the present "but some differences are likely to remain" is a bit abrupt)

2) Introduction (1)

At line 30-32, the authors discuss the development and application of regional atmospheric CO_2 inversions. However, they cite papers on CO_2 footprint modeling and inverse modeling for other species, but no paper on CO_2 regional inverse modeling even though there have been a lot studies in this field since more the 5 years. Providing details at the end of page 3 on the papers by Stohl et al. (2009) and Manning et al. (2011), which had to deal with the estimate of sources with very different spatial and temporal patterns compared to the CO_2 natural fluxes, is a bit problematic (and one can hardly see the link between these detailed description on page 3 and the specificities of this study on page 4). Therefore, the introduction presents the principle of regional inversion as a sort of generic algorithm that could be applied similarly to any GHG. And it gives a limited view on the range of techniques that have been used for regional atmospheric inversions. This is emphasized by the implicit assumption that the the concept of "baseline" (in the way it is treated in this study) applies to all atmospheric regional inversion cases, while many regional systems, by using outputs from larger scale model to force their boundary conditions and/or by relying on the

spatial gradients between measurement stations to limit their impact, give a different answer to the problem of solving for the influence of fluxes outside the modeling domain. Opposite to what is said at line 2 on page 3, the use of Lagrangian models is not a requirement for regional inversion. In general, the text which attempts at defining the regional inversions vs global inversion from page 2 to page 3 is a bit confusing and could have been more concise.

3) Method sections (2-5)

The logical structure is often confusing. This regularly forces the authors to anticipate for the next sections and thus make redundancies (e.g. the inversion technique is introduced on p8 26-28 and at the beginning of sections 4 and 5.3). Presenting the general frame of the inversion (5.3) could help solve for it.

The mathematical formalism should be based on rigorous notations. Presently, the vector x and matrix T describe completely different things between section 3.2 and section 5.3. The spatial distribution of the fluxes within a region, which is implicitly contained in matrix T in section 5.3, and not in the vector of the prior estimate of the regional budgets x0, are called throughout the text "prior distributions".

The prior uncertainties derived from scientific publications and objective comparisons (section 4.1 and then extrapolated in section 5.3 at the regional scale) seem to be smaller than the artificial "uncertainty component of 50% of the seasonal amplitude" added in section 5.3 (even though it is difficult to understand what l15-17p16 really mean). Therefore, the derivation of the prior uncertainties seems artificial.

The introduction of the term with S in equation (2) raises questions. Why did not the authors rather introduce temporal correlations in the Co matrix to limit the changes in the weekly variations of the fluxes ?

The logic behind the specific choice and configuration of the sensitivity tests on page 18 is not really convincing and does not seem to tackle some of the most critical sources

of errors in the inversion. What is the link between the modeling in NAME (l4p18) and the details given on (ii) later in page 18, which concern the spatial distribution of the fluxes within the regions ? Definitely, a practical assessment of the transport model uncertainties would have been interesting. I do not understand the test of sensitivity to the inclusion of the ocean fluxes in the modeling framework. What does l2p19 mean ? How to connect it to l30p28 ? In principle, the impact of the uncertainty in these fluxes should be correctly accounted for in the reference test. The point maybe be about "biases" but it is difficult to guess, in this section and when analyzing the results, whether the authors systematically make a rigorous use of the term "bias".

4) Result sections (7)

In general, the analysis are a too qualitative.

The robustness of the inversion needs to be better assessed through the analysis of the comparisons between prior and posterior CO2 vs. measurements. I feel that relying on the sensitivity tests to state that the results are robusts (l20-21p1, l8p29) is not really satisfying given that the sensitivity tests do not really sample the most critical sources of errors in the inversion system. The first subsection of section 7 could have been dedicated to such a detailed analysis of the CO2 concentration model - data misfits, providing an opportunity to exploit the long discussions on the station footprints to potentially correlate the highest misfits with specific transport conditions.

In order to ensure that the chi square statistics are right, the authors have to derive a 0.4ppm model error which is surprisingly low for 10magl stations surrounded by a complex topography, and which is strange given that the model data misfits often exceed the projection of the prior uncertainty in the obs space. As said in major comment B, I think that the authors made a wrong chi test but I also assume that this hardly explains such a low diagnostic of the model error. Figure 12 normalize the model data misfits by the prior uncertainty and the text hardly discusses absolute values (in ppm) of these misfits. These absolute values could reveal the skill of the model for simulating CO2 at

the measurement sites. At least, figure 12 reveals large biases (and errors with a high temporal correlation over several months) whose consequences for the confidence in the inversion results should be better weighted.

The confidence in the inversion results at the weekly scale can also be weakened by the quite high week to week variations of the inverted weekly fluxes despite the term with the S matrix in equation 2. However, while the text takes time to discuss this penalty term in section 5.3, it skips the analysis of the week to week variations in the result sections. Does it lower the confidence in the monthly mean results or can it be explained by actual variations or through the variations of the observation constraint ?

One has to trust the authors that their inversion results can be directly compared to a sort of crude NIR total estimate in the abstract and in the introduction. However, the text in 7.3 reveals (even though it is often confusing) many differences between the type of fluxes covered by the inventory and the inversion, and that the inventory provide enough details to filter some major flux components that cannot be accounted for by the inversion, so that a more relevant budget could be derived for this comparison. Therefore, things should be presented differently (i.e. turned the other way around, starting with a presentation of the content of the NIR, and following with an extraction of a relevant budget from the NIR) and, in the abstract,"the sink [derived by the inversion]" should not be compared to "the reported 27 tGco2YR-1" (l. 26-27 p1) but to a more relevant combination of the NIR components.

5) Conclusion (8)

Given all my concerns that are detailed above, I think that the discussion misses the critical needs for the improvement of the inversion in the last paragraphs, and that it is too optimistic regarding the results in the first paragraph and at l1-3p31.

- some questions

* why do not the author assimilate 14:00-15:00 data ? in order to save computations ?

* p6 l3 vs l7: it seems that this site should be strongly influenced by the emissions of Wellington. Can the authors provide more details (e.g. statistics) on this topic than at l23-24p20 ?

* l31p11, l12p2, l12-16p24: if focusing on fossil fuel emissions, what is the difference between EDGAR and NIR ? can we assume that NIR has a far more precise estimate than EDGAR ?

- Sample of minor issues illustrating some of the general comments above

* the text often forgets to be precise about the fluxes that are discussed (l13p1-> natural, l.12 p2: CO2 emissions -> anthropogenic accounting for land use or fossil fuel only ?; first sentence and line 13 of abstract and 1st paragraph page 4: precise that you target natural fluxes; l27p8: the system solves for natural fluxes; l2-3p9 sinks and sources: natural sinks and sources...); on the same topic, lines 12-14 p24: as it is, these sentences do not make sense since land use change emissions are not fossil fuel emissions.

* discussing RBM in section 2 and 5.2 is a bit strange but the authors are embarrassed with the fact that the partitioning of the fluxes accounts for the future inclusion of RBM data in the system for future experiments.

* examples of abusive shortcuts: l13 p1 (from measurement records), l16 p2 (locally present vegetation), p822-24, l7p17 (which quantity does this number corresponds to ? C is a cov matrix)...

* examples of awkward and confusing sentences: l2-3p3, l10-11p3, all sentences from l14 to l21p3, l6p4, l23-24p9 (there is a long discussion on the biomes and land use maps on page 9 and 10 but we hardly understand how they will be used), the first paragraph of 5.1, "monthly standard deviations" l20p13, l4p15,l15-17p16, l20-21p16, l4-6p17, l10p17,l14-15p18, the whole paragraph (ii) on p18, l21p21,12-14 p24, l20p24, l21-23p26, l22-25p27, l26-28p27....

* l18p5: the link between the PBL and the horizontal extent of the footprint of the measurements is a bit confusing

* the logic behind the model representativeness error computation is not obvious (p5l28-32), in principle, the STD of the concentration variability at the 5-min scale does not correspond to the skill of the model to represent hourly measurement averages; the assumption underlying this computation should be explained

* p8 l8-11: the normalization (if I understand it correctly) does not make sense and just loses information.

* the text often forgets to associate numbers with a time or space scale (e.g. l31-32p10, 3rd paragraph of page 11, l7p24...)

* section 6 is poorly connected to the other sections, and the part of section 6 before the start of 6.1 sometimes sounds like a summary of the subsections 6.1 and 6.2

*l14-17p14 and l25-29p19: the diagnostic assesses the sensitivity to the fossil fuel emissions in the part of Australia that is in the modeling domain, not the sensitivity to all fluxes in the whole Australia; in other places, and especially in the conclusion, the paper will state that potential errors from Australia need to be handled.

* the logic behind the justification of the partition of the ocean regions in the second half of p14 is not really clear.

* l5p16: it is not the "uncertainty in the inverse modeling system" (which is something that would be difficult to quantify), it is the likely the uncertainty in the transport model

* l7-10 p23 are poorly connected to the analysis above

* l17-20p27 are wrong, if posterior uncertainties are low, negative correlations between these uncertainties do not mean that the inversion is unable to distinguish the corresponding fluxes; in such cases, it would just mean that the residual uncertainty in the corresponding fluxes is due to such a problem of separation

* labels in the figures are often difficult to read

*l22-24 p21: at this stage of the manuscript, we do not know how the additional uncertainty from the sensitivity tests is included in the figure (this will be explained at l21-25p28 which are quite confusing)

C11