Atmos. Chem. Phys. Discuss., 14, C8058–C8066, 2014
www.atmos-chem-phys-discuss.net/14/C8058/2014/

Atmospheric Chemistry and Physics
Discussions

Interactive comment on "The potential for regional-scale bias in top-down CO_2 flux estimates due to atmospheric transport errors" *by* S. M. Miller et al.

Anonymous Referee #1

Received and published: 15 October 2014

This paper describes a study of transport errors in CO2 inverse modeling studies that result from imperfect meteorological models to determine the CO2 transport. The aim of the research is to examine how errors in CO2 transport in meteorological models affect top-down estimates of CO2 fluxes. The authors use a global meteorological model-data assimilation system (CAM-LETKF) to present two case studies. The first case study examines how biased regional CO2 fluxes would need to be before that bias is detectable above the meteorological uncertainties estimated by CAM-LETKF. The second case study examines which meteorological factors are associated with persistent, month-long deviations in atmospheric transport using a synthetic tracer with

constant global emissions. The authors find that the largest uncertainties in 6 hourly CO2 transport are localized to regions where the magnitude and/or diurnal variation of the fluxes is largest. In contrast, uncertainties in month-long average CO2 concentrations are more spatially distributed; CO2 transport errors correlate over longer periods of time in remote areas. This is the first study I have seen that presents an estimate of transport errors based on an actual ensemble of (assimilated) weather realizations, and the paper is very suitable for publication in ACP. Moreover, the methodology of this paper is original and the modeling experiments are mostly well executed. The writing is very clear and the authors deserve much credit for explaining the many difficult concepts in this study, while at the same time making interesting connections to the wider field of research. I recommend its publication, but only after the authors can address one major, and a few minor concerns I have after reading this manuscript.

(major) My main criticism if focused on case study #1 and the framing of this experiment as a way to quantify the bias detection limit of carbotracker (CT). But CT does not work on a month by month basis lie your lambda scaling factoor, it does not consider signals site-by-site like here but instead a whole network, and it does not scale flux signals locally at each site like in your FSSR but over a large spatial area that is also seen by other sites. I thus fear that the final numbers of this analysis reflect very little on flux biases in CT. Luckily, I also believe the experiment does not necessarily need this direct connection to the CT fluxes, as after all determining a bias detection limit for one particular inverse model in the field is not so relevant for a wider audience. Framing this experiment as a way to determine the balance between large-scale flux influences and transport errors is in that sense more appropriate, and I think describes better what was actually done.

Specifically, your comparison of transport noise (SSR) and flux biases (FSSR) is done in squared residual space which only measures the magnitude of a signal, but does not account for its sign. A bias in fluxes would typically manifest itself as a consistent over- or underestimate of the true concentrations observed and even if these are small (say 0.5 ppm) compared to the more random transport uncertainties (say 3 ppm), their consistency in sign over longer periods of time would make them detectable. In fact, in a Bayesian inversion the system would try to overcome this small bias as by design it strives for zero mean residuals even in the presence of large observation error covariances. This argument of the sign becomes even stronger when we consider more sites at once, which is typically what is done in an inverse system: the coherency of the bias over many sites drives the fluxes towards unbiased even more quickly. This power of the network is reflected in the covariances in residuals between sites which are also not accounted for in this analysis.

To overcome this criticism, I would suggest one of two approaches:

(1) is to try and change the metric so that it includes more sites at once and includes also spatial covariances between residuals. The new metric then also needs to account in some way for the sign of the residuals.

(2) Is to write the question of this case study differently and to say that you'll try to estimate to what variation in flux magnitude the meteorological uncertainty corresponds for each site given a realistic surface flux from CT. This also means that most of the use of the word "bias" gets replaced by "flux signal".

Finally, it could be interesting to ask the carbotracker team (there are some BOAA co-authors) to proof your suggestion that a ∼20% flux bias would remain undetected if transport errors from your system were prescribed. I realize that this might be too much work though.

(minor) I find the discussion section a bit too short, and would like to see some more connections to other studies in this field. For example, some reflection could be added on the LETKF methods used by these authors in the past, and about the possible gain of co-simulating CO2 and transport errors. Also, there is room for some reflection on the covariations of CO2 surface fluxes, and those that shape the weather conditions (water and energy and momentum fluxes). What would the next step with this type of

system look like when surface fluxes also become a function of the weather variables?

Furthermore, these findings can nicely be connected to the error budgets presented in Pino et al., (2011) and in Williams et al (2011). Both take a look at the driving forces behind variations in CO2 in the PBL, one from a local and one from a larger perspective.

Some more detailed comments are given in the annotated PDF and repeated as a list below. Note that I remain a bit puzzled on the implementation of the SSR vs FSSR metric in equations 4 and 5 + the explanation in the supplement and would like to see some clarification.

Notes: p.23681: we find that CarbonTracker CO2 fluxes would need to be biased by at least 29 %, on average, before that bias were detectable at existing non-marine atmospheric CO2 observation sites. – Highlighted 14 oct 2014

p.23684: the range – Highlighted 13 oct 2014

p.23684: I could not find where the range is actually applied instead of the SDV – Written 14 oct 2014

p.23684: What is the temporal resolution of these fluxes? To interact properly with the dynamics it needs to be at least 3-hourly or better I would say. Monthly mean fluxes probably do not give a right view of the errors associated with a growing boundary layer and entrainment. – Written 13 oct 2014

p.23684: any subsequent diferences in CO2 among the model realizations are due entirely to meteorological uncertainties. – Highlighted 13 oct 2014

p.23684: So this means that the feedbacks of meteorological errors on carbon exchange are not accounted for? In other words, different weather does lead to different water exchange, but not other carbon fluxes. Okay, I got it. – Written 13 oct 2014

p.23684: – Highlighted 13 oct 2014

p.23685: Larger than most means more than 32 if k=64 members? – Written 13 oct 2014

p.23686: So this suggests that for p to get to 0.05, there must be 64*0.05 = 3.2 elements in A (eq 5). And when there are four or more SSRs in the set that are larger than FSSR then you have proven the null-hypothesis that bias in fluxes is indistinguishable above transport uncertainties. This seems quite strict to me.

Oh wait, I think there might simply be a typo here and you actually meant 0.5 instead of 0.05? Sorry, I spotted this kind of late because 0.05 is such a typical p-value in statistics... – Written 13 oct 2014

p.23686: Can you elaborate in the main text how this temporal covariance is accounted for. I am sure the Supplement gives info but I'd rather like to understand it here. – Written 14 oct 2014

p.23687: This suggests you indeed used fluxes including a diurnal cycle – Written 13 oct 2014

p.23688: I think this is an absolutely wonderful conclusion to draw, and hope it will get a prominent place in the abstract and conclusions – Written 13 oct 2014

p.23688: First, observation sites that are far fromlarge fluxes are more likely to produce a biased CO2 budget than sites near to large 15 surface fluxes. These "remote" sites see a lower CO2 signal from surface fluxes, and the transport errors at these locations are generally correlated over longer periods of time. Second, most existing top-down studies will underestimate the uncertainties in estimated CO2 fluxes. Existing inversions rarely account for error correlations in CO2 transport and most likely underestimate the posterior uncertainties as a direct result. – Highlighted 13 oct 2014

p.23688: I do not think this case study uses an appropriate question, as your test is not a correct metric to determine the minimum size of flux biases that are detectable through atmospheric CO2.

Specifically, your comparison of transport noise (SSR) and flux biases (FSSR) is done in squared residual space which only measures the magnitude of a signal, but does not account for its sign. A bias in fluxes would typically manifest itself as a consistent over- or underestimate of the true concentrations observed and even if these are small (say 0.5 ppm) compared to the more random transport uncertainties (say 3 ppm), their consistency in sign over them would make them detectable. In fact, in a Bayesian inversion the system would try to overcome this small bias as by design it strives for zero mean residuals.

This argument of the sign becomes even stronger when we consider more sites at once, which is typically what is done in an inverse system: the coherency of the bias over many sites drives the fluxes towards unbiased even more quickly. This power of the network is reflected in the covariances in residuals between sites which are also not accounted for in this analysis.

I am especially critical of this because the question is coined as a way to put a lower limit on biases detectable with CT. But CT does not work on a month by month basis, it does not consider signals site-by-site but a whole network, and it does not scale flux signals locally like in the FSSR but over a large spatial area that is also seen by other sites. I thus fear that the final numbers of this analysis reflect very little on flux biases in CT.

Instead, I would suggest one of two approaches:

(1) is to try and change the metric so that it includes more sites at once and includes covariances between residuals. The new metric then also needs to account in some way for the sign of the residuals.

(2) Is to write the question of this case study differently and to say that you'll try to estimate to what variation in flux magnitude the meteo uncertainty corresponds for each site given a realistic surface flux from CT. This also means that most of the use of the word "bias" gets replaced by "flux signal"

– Written 13 oct 2014

p.23688: This effect of measurement bias was explored by Masarie et al., (2011), please reference. – Written 14 oct 2014

p.23689: greater or equal to 0.3 – Highlighted 14 oct 2014

p.23689: What does the number 0.3 represent? – Written 14 oct 2014

p.23689: uncertain these – Highlighted 14 oct 2014

p.23689: Since this point is now mentioned a second time, a reference to Pino et al., (2012) is in place as he already showed such PBL-CO2 error relations. – Written 14 oct 2014

p.23689: Again, your analysis is very nice but this conclusions is not correct. Since one of the authors is associated with the CT group at NOAA, perhaps a synthetic inversion could be done to prove this statement beyond my doubt? – Written 14 oct 2014

p.23689: Among other results, we find that CT would need to be biased by 29 %, on average, before that bias were detectable above CO2 transport uncertainties15 at terrestrial, atmospheric observation sites. – Highlighted 14 oct 2014

p.23689: This second part is very nice. Can you speculate how this conclusion might change if the interactions between the meteorological variables and the CO2 fluxes themselves were included in a follow-up study? – Written 14 oct 2014

p.23690: I find the discussion section a bit too short, and would like to see some more connections to other studies in this field. For example, some reflection could be added on the LETKF methods used by these authors in the past, and about the possible gain of co-simulating CO2 and transport errors.

Also, there is room for some reflection on the covariations of CO2 surface fluxes, and those that shape the weather conditions (water and energy and momentum fluxes). What would the next step with this type of system look like when surface fluxes also

become a function of the weather variables?

Furthermore, these findings can nicely be connected to the error budgets presented in Pino et al., (2011) and in Williams et al (2011). Both take a look at the driving forces behind variations in CO2 in the PBL, one from a local and one from a larger perspective. – Written 14 oct 2014

p.23693: The Supplement provides analogous figures for daytimeor nighttime-only model output – Highlighted 13 oct 2014

p.23693: You could compare these to the posterior flux uncertainty in CT and show that they are at least as large indeed. – Written 13 oct 2014

p.23694: What do the letters below the x-axis indicate? – Written 14 oct 2014

p.23694: Why do we only see the land CV? Was the constant flux also only applied over land? This was not clear to me from the description yet. – Written 14 oct 2014

p.23695: The variables 1,2 and 4 look very similar as one would expect from meteorological principles. In the same way, 5 and 6 are closely related. What is perhaps more interesting is that (1) the PBL height which in the end is most directly related to the CO2 mixing ratios is not shaped the same as these primary drivers. This stresses the need for a meteorological model to calculate the (co)variances of transport errors rather than to just use some simple proxy. And (b) is that the CV of temperature and CO2 are very similar which is because they are shaped by the same large scale synoptic systems. This is also discussed in the Williams et al., (2011) paper, and the driving power behind the LETK methods shown by Kang, Kalnay, Liu, and Fung (co-authors here). Perhaps this is worth to mention in the discussion. – Written 14 oct 2014

Supplement:

p.1: Thanks, this is a helpful Supplement and nicely keeps extra information out of the main text.

p.19: This 5% I guess corresponds the p=0.05 probability stated in the main text. That suggests this was not just a typo, and I remain confused on equations 5 and the use of this test.

p.19: This is a nice illustration of the properties of the SSR, which I think correctly assumes transport errors to be normally distributed around a zero mean. But the problem I have is in the comparison to FSSR, which for a biased flux would not just be a residual around some mean, but an actual signal with a sign and a spatial pattern. See for instance the figures S9, S11, and S14 that both represent winter conditions. A shift of the fluxes by 10% upwards would lift both lines for the ensemble mean upwards by 0.5-2.0 ppm and reveal a systematic offset (if the model mean was a bit more unbiased which it is not without data assimilation of the fluxes) at three locations.

Please also note the supplement to this comment:
http://www.atmos-chem-phys-discuss.net/14/C8058/2014/acpd-14-C8058-2014-supplement.pdf

---

Interactive comment on Atmos. Chem. Phys. Discuss., 14, 23681, 2014.