Response to Anonymous Referee #1

We thank Referee #1 for a detailed review of our paper, which certainly helped in improving the paper. In the following, we respond to the individual points raised in italic letters (in blue), the reviewer comments are left in normal font (in black).

This study investigates the performance of a regional inversion of anthropogenic CO_2 emissions using synthetic experiments varying the prior fluxes, transport model and optimization method, and inversion setup. Conclusions are drawn regarding the optimal number of regions to be optimized, and the relative importance of transport model uncertainties. However, as will be explained further below, it is not clear what we learn in the end.

Authors' response: This is a model description and evaluation paper. What we learn from this study is that it is very possible to produce flux results with large estimation errors and unrealistic spatial patterns (negative fluxes where they should be positive) when the differences between the prior model concentrations and observations are large and/or biased.

The message that we would like to convey was that an extensive evaluation was necessary for any inverse modelling study (often not done) before the application of the model to real observations. Systematically increasing the number of unknowns (e.g. sub-regions as in our study) in combination with variations in the major model components (e.g. prior fluxes, model transport) is one way to check the stability of the inversion setup and the robustness of the results, as well as identifying the suitable setup for real observation inversion. Understanding the sensitivity of the flux estimates to prior error assumptions, region definitions, flux error, transport error, and optimization procedures could identify where to focus our effort on improving the component that was the most important in the inversion system. We will highlight what we have learned in the Abstract and Conclusions sections more clearly.

Most of the findings that are presented will depend on the specific setup of the inversion that is chosen. However, since this setup is far from realistic – the practical significance of the results for regional inverse modeling of CO_2 using real data remains unclear. In my opinion, this point will have to be addressed clearly in the revised version of the manuscript to make this manuscript acceptable for publication.

Authors' response: it was not the intention for the authors to claim that the current version of the inverse model was ready to estimate regional CO_2 emissions from fossil and/or biospheric sources. We will make it clear in the current version of our inverse model can be used to estimate slow varying monthly emissions as long as the transport error is small. In fact, we used fossil fuel CO_2 and anthropogenic CH_4 as two different priors in this study to show our inversion approach could work for both of these slow monthly varying and positive only emissions.

This comment regarding our model setup and the practical significance of the results is expanded to many detailed comments below. We will address these detailed comments in sequence and make clear that the inversion setup examined here is not intended for use to infer the total regional CO_2 fluxes, please see below.

In addition, to improve readability I recommend that the authors focus only on the main findings, which could substantially reduce the size of the paper.

Authors' response: The descriptions of MCMC and CFM will be moved to the supplementary section. The Introduction, Section 2.2 the description of the prior fluxes and wherever possible will be shortened. In fact, a sensitivity analysis of the estimates to the prior flux (10%, 100%) and transport (10%, 30%, 100%) error assumptions was intentionally left out to reduce the size of the paper. As we found in this study, the specifications of the errors were not the determining factor.

GENERAL COMMENTS

The authors recommend that an inversion system is first tested in a synthetic environment before it is applied to real data. I fully agree to this, however, since the outcome of such an experiment depends on the specific details of the setup it should be realistic in the sense that the same setup could directly be applied to real data. This is clearly not the case for the setup that is presented here, since it only addresses anthropogenic emissions of CO_2, ignoring the natural component of the regional carbon cycle.

Authors' response: The model evaluation (and first real data inversion) was to start with the simple case of source only (positive fluxes, no sink or negative flux) to facilitate the evaluation. A suitable chemical species is wintertime CH_4 as noted, when the wetland sources, soil sinks and OH chemical reaction (over the 5-day model integration period) are small or negligible in a regional sense.

We considered using CarbonTracker- CH_4 (Bruhwiler et al., 2014) as the target result. However, the low resolution of CT-CH_4 transport model (TM5 with 4 deg latitude x 6 deg longitude) would not be sufficient to fully resolve the four measurement sites or more than a few sub-regions in our inversion domain (~ 10 deg latitude x 20 deg lon longitude).

A more suitable choice was CarbonTracker fossil fuel CO_2 (CarbonTracker, 2010, 2011), with positive fluxes and varying on monthly timescale (or without the added complexity of diurnal variations, the typical assumption for CH_4 inversion studies). CarbonTracker CO_2 has the higher resolution of 1 deg x 1 deg for North America (including our domain and surrounding).

The prior flux spatial distribution of CT-ff is similar (not identical) to the anthropogenic CH_4 (our target first real observation inversion, in progress) with some point sources (large facilities). Therefore we used CT-ff as the target for flux and mixing ratios which allow for the testing of prior flux errors, transport errors, background mixing ratio errors, and so on.

These explanations will be included in the revision of the manuscript.

In addition, the boundary conditions of the regional domain that is optimized are assumed to be known exactly. Since the regional biospheric fluxes are expected to be the main uncertain component, it is unclear to me how this inversion is supposed to work when applied to real data.

Authors' response: Although we did not optimize the boundary conditions due to the fact that there were not enough observations and the estimation errors were already large, we made a note in the text that there could be errors associated with the boundary conditions (estimated baseline values). Therefore we made no distinction between boundary error and regional transport error. A general term of "transport error" was used to encompass all errors associated with transport modelling.

It is mentioned that the methodology could be applicable to wintertime CH_4 fluxes. But if this is the application that the authors have in mind then why perform a test inversion for fossil CO_2 instead of CH_4?

Authors' response: At the time of conducting this study and model development, the finest possible horizontal resolution of predicted greenhouse gas concentrations was provided by CarbonTracker CO_2 at 1°x1°, whereas the output resolution of CarbonTracker CH_4 was 6°x4°. It was not possible to use CT-CH_4 as the target.

The authors comment on the estimated posterior uncertainties in relation with the actual deviations from the predefined true fluxes, concluding that the estimates are too optimistic. However, this conclusion depends on how the assumed a priori flux and data uncertainties reflect the actual errors. It seems that no effort was made to analyze the statistics of the difference between for example CT2011 and CT2010 before specifying the a priori flux covariance. The same is true for the model-data mismatch.

Authors' response: We did not go into the details to study the statistics of the difference between CT2011 and CT2010 before specifying the prior flux error covariance. Only monthly simple statistics were calculated as presented in Table 2.

Instead, we tested different a priori flux and model-data mismatch uncertainties using 10% and 100% for both $(\sigma_{prior})^2$ and $(\sigma_e)^2$ and found that the flux estimates were not sensitive to a priori and model-data uncertainties on the provincial and annual time scales. This result could be included in the revision, but it was intentionally left out to shorten the manuscript.

When the transport model error is small, it was not necessary to concern with the a priori uncertainties as shown in Figure 5 (flux error case). When the transport model error is large, specifying different prior uncertainties unfortunately cannot remove such systematic model error.

In practice, we often do not know the true prior flux and model-data mismatch errors and the spatio-temporal structures, and most important, the interaction of the flux and model-data mismatch errors as all sources of errors are folded together in the modelled mixing ratios. Therefore, we focus our effort on carrying out a detailed assessment of the sensitivity of the estimates to different assumptions and setups for inversions.

Figures C1 and C2 below show that 100% prior flux error used throughout should be liberal enough to allow the parameter estimates to be calculated on a wide possible range for the seven region definitions used in our study.

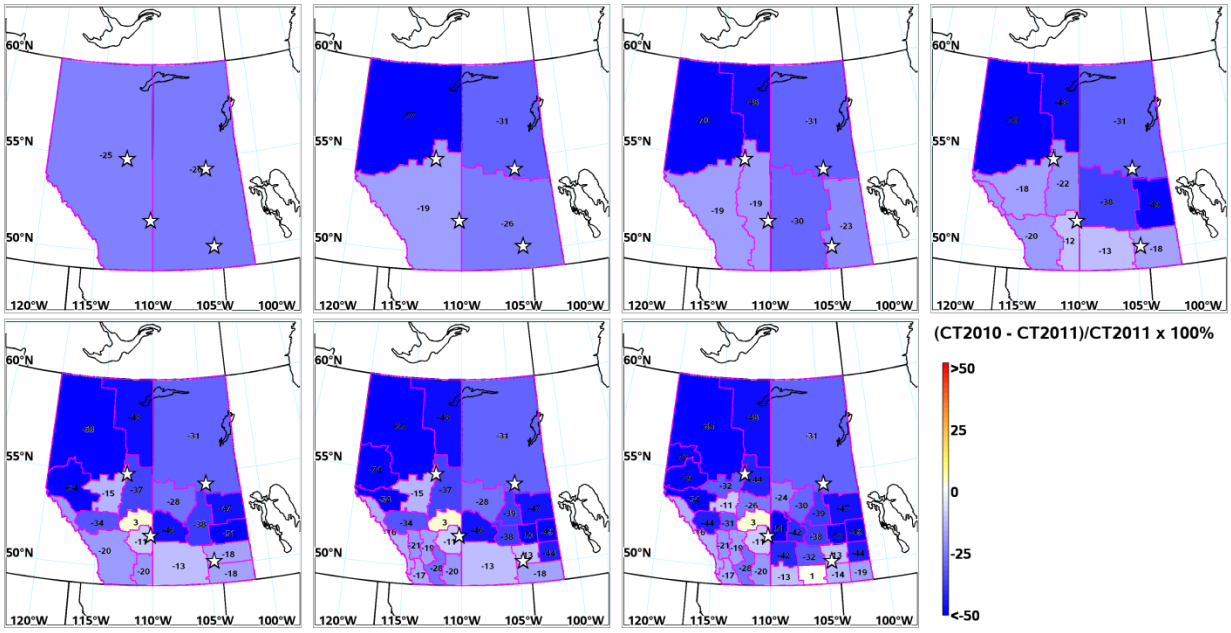


Figure C1 shows the relative percentage difference between CT2011 and CT2010 for the AB+SK inversion domain. The prior flux errors range from +3% to -79% as shown on the map depending on the region definitions.

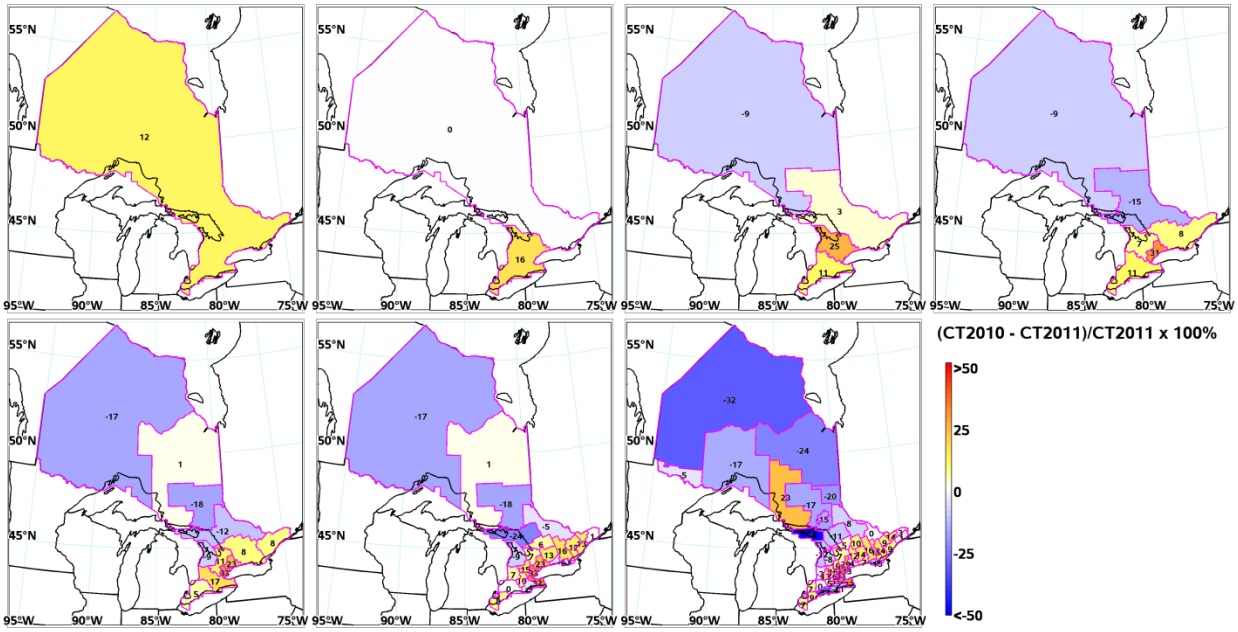


Figure C2 shows the relative percentage difference between CT2011 and CT2010 for the ON inversion domain. The prior flux errors range from +33% to -51% depending on the region definitions.

In terms of the difference between the prior model and target mixing ratios, Table below shows the relative percentage difference between the prior model and target mixing ratios for the transport error only case (no flux error) averaged over January to December 2009. The baseline mixing ratios were subtracted from the individual mixing ratios. The relative percentage difference between the prior and the target ranges from 0% to -48%.

Region	*Station*	*Prior (ppm)*	*Target (ppm)*	*(Prior-Target)/Target x 100%*
AB+SK	*BRA*	*0.8*	*1.1*	*-27*
AB+SK	*EST*	*0.9*	*1.1*	*-18*
AB+SK	*ETL*	*0.3*	*0.3*	*0*
AB+SK	*LLB*	*0.8*	*1.1*	*-27*
ON	*EGB*	*2*	*2.7*	*-26*
ON	*FRD*	*0.6*	*0.6*	*0*
ON	*DOW*	*4.5*	*8.7*	*-48*

We used 30% for model-data mismatch error for all stations simply because a sensitivity analysis was done by using 10%, 30% and 100% as the model-data mismatch errors. The results indicated that the flux estimates were not sensitive to such error specification using MCMC (not included due to the length of the manuscript). The 30% prior model-data mismatch is comparable to other real observation-based regional inversion studies, e.g. Gerbig et al. (2003), Zhao et al. (2009), etc.

Model-data "mismatch" errors do not give a clear indication of how well a given station can constrain the surface fluxes. A small model-data mismatch can either mean that the station does not provide any constrain or the prior fluxes are correct.

Flux covariance appears to cause little impact on the results if the transport is perfect as shown in our flux error only case in Figure 5. Similarly other issues (different sub-region definitions, different inversion time window, etc.) seem unimportant if the transport is perfect. This seems to be a general or robust result (east and west).

In this case, how can the method of calculating posterior uncertainties be blamed of inconsistencies? I wonder also how representative the difference between CT2011 and CT2010 is for uncertainties in fossil fuel CO_2 emissions. They can certainly not be considered independent estimates of fossil fuel fluxes.

Authors' response: It is true that CT2010 and CT2011 ff fluxes would have errors compared to the true ff CO_2 emissions and are not to be considered as independent estimates of fossil fuel fluxes. We will clearly note in the revision that CT2011 (target) was used here as a proxy (or a suitable example) for a flux with distribution resembling ff CO_2 and anthropogenic CH_4, and CT2010 was used as the prior with known difference from the target CT2011 in our pseudo-obs inversion.

Conclusions are drawn regarding the relative performance of Bayesian cost function minimization and the MCMC method. However, how can different methods to find the solution of an inverse problem be compared if they are applied to different inverse problems? A clearer distinction should be made between optimization method and inversion setup. If the MCMC method was applied to the same optimization problem, one would expect to find the same solution – unless one method fails to find the optimum, e.g. because of non-linearity. In this case, at least the CFM inverse problem seems linear, so it should be capable of finding the right solution. The difference between MCMC and CFM seems more in the assumed statistics (inverse Gamma versus Gaussian). But then if MCMC performs better it is probably because of a different weighting of outliers. Further analysis is needed to gain understanding of what is causing the difference between the two methods.

Authors' response: This study is comparing solutions of different inverse problem, similar or analogous to studies like Peylin et al. 2013, Gourdji et al. 2012, etc. However, we need to be more careful in the usage of the terms 'optimization method', 'inversion setup', 'inversion problem' and 'optimization problem' as noted by the reviewer above. We will clarify in the revision that:

- *This study examined different inverse problems, including cases with prior flux errors only, transport errors only, both prior flux and transport errors, etc.*
- *The different inverse problems are 'solved' (to find the scaling factors for the sub-regions) using different 'inversion setup' including different number of sub-regions, different inversion time windows or periods (1 month or 3 months), and different optimization methods.*
- *MCMC solves for the joint distribution of the scaling factors and CFM minimizes the standard cost function.*

We found that MCMC and CFM results converged as the amount of observations increased. However, MCMC is less likely to generate negative or physically unrealistic flux results. Following the review suggestion, we will note the differences in the results are partly due to the intrinsic differences in the optimization methods (or differences in the 'optimization problem').

There are many potential sources of errors for an inversion model. Often the prior assumptions are not satisfied, such as observations are approximately independent, errors are approximately normally distributed, transport has negligible biases, etc. The assumption violations in the inversion model could lead to poor inversion results and unrealistic uncertainties. Hence the need to evaluate each inversion setup as fully as possible to help quantify the posterior uncertainty, as the results cannot be generalized or assumed applicable elsewhere or for other chemical species.

The discussion about aggregation errors and the optimal number of regions seems to have been influenced by the treatment of a priori flux uncertainties. If a region is split up in two equal parts, then care should be taken to specify the uncertainty of the separate regions such that they don't alter the uncertainty of the combined region. In this study, it seems that 100% uncertainty is assumed regardless of the size of a region.

But then if the individual fluxes are assumed to have independent uncertainties, the aggregated uncertainty of the whole region will go down as it is split up into a larger number of sub-regions. This is

because the errors of the sub regions will partially cancel out in the regional integral (with the square root of the number of regions). Unless this issue is dealt with carefully, it will confuse any inferences about aggregation errors.

Authors' response: Kaminski et al. (2001) explained that 'the aggregation error arises from biased (or inhomogeneous) sampling of the small-scale, unresolved source structure. We stress again that we generally will not know this small-scale source structure perfectly and we are trying to avoid the impact of errors in this structure.'

Therefore in our 'Prior flux error case (I)' (with only errors in the source structure, no transport error), the posterior error is mainly the 'aggregation error'. Thus the posterior error is a better measure of the 'aggregation error' than the 'posterior uncertainty' (as it could be smaller than the 'aggregation error' in our results). Increasing the number of sub-regions can increase or decrease the total (sum of the sub-regions) posterior error, and similarly increase or decrease the posterior uncertainties. The optimization procedures can behave in a complex way with the number of sub-regions.

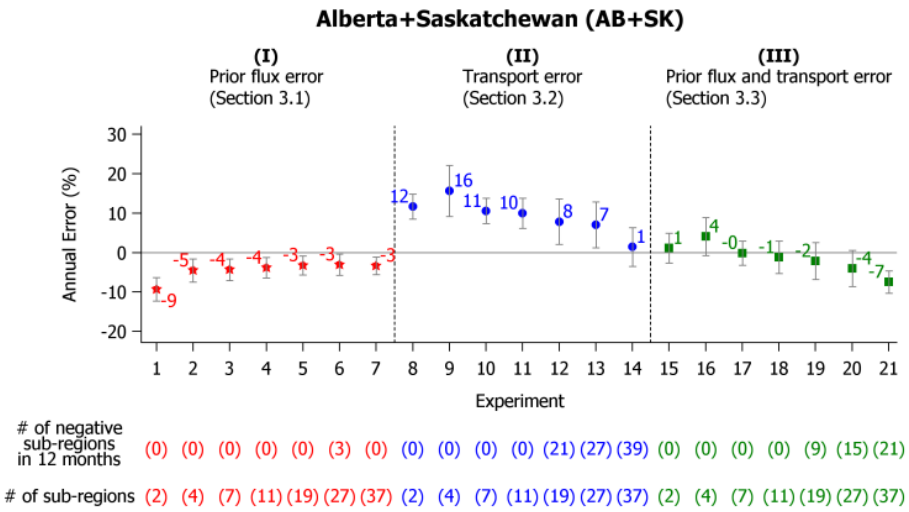
For example, it is possible for the atmospheric transport to correlate the flux signals from different sub-regions. If a parcel trajectory passed through different sub-regions with non-zero fluxes over the 5-day period, then the flux signals from the different sub-regions could be mixed (correlated) within the parcel. Integrating the contributions from all the parcels, the flux errors for the different sub-regions in combination with transport errors could be correlated in complex ways (particularly when the errors are biased).

As it is likely not the case in our model to have 'independent uncertainties', the 'aggregated (posterior) uncertainty of the whole region' in our results could go up or down as it is divided up into a larger number of sub-regions (and not always 'go down').

These characteristics in the posterior uncertainties can be seen more clearly in the revised version (see below) of Figure 5 in which the scales of the y-axes have been changed to improve readability. The number of negative sub-regions for the corresponding number of sub-regions is shown on the 2nd row from the bottom in the figure. These explanations will be added to the manuscript revision.

Figure 5 (re-plotted to improve legibility)

a ***MCMC***



CFM

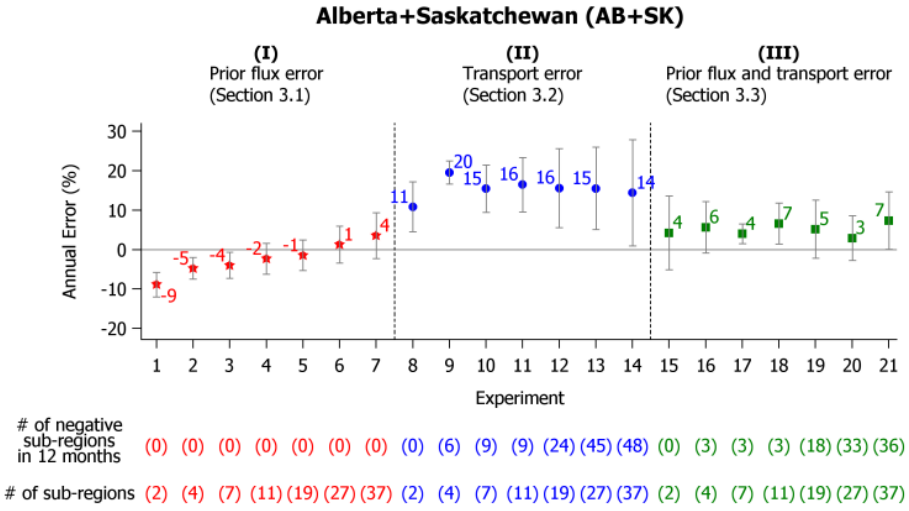
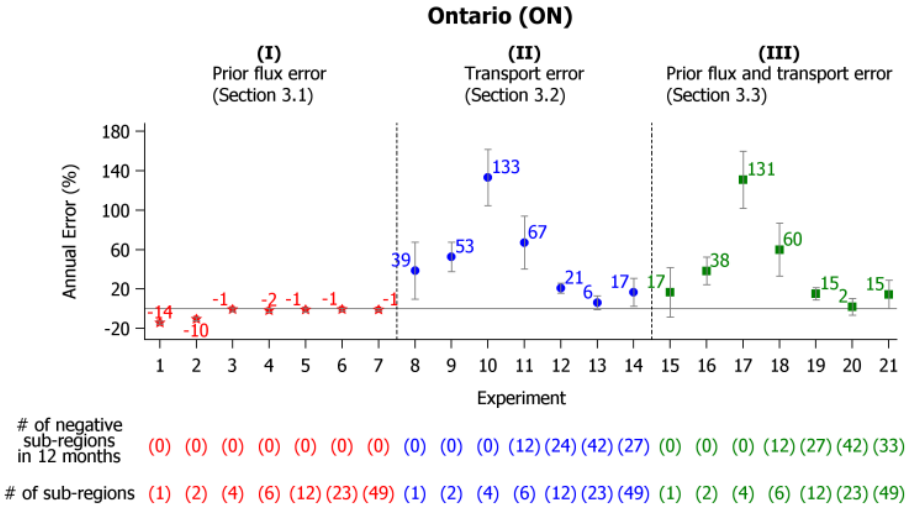


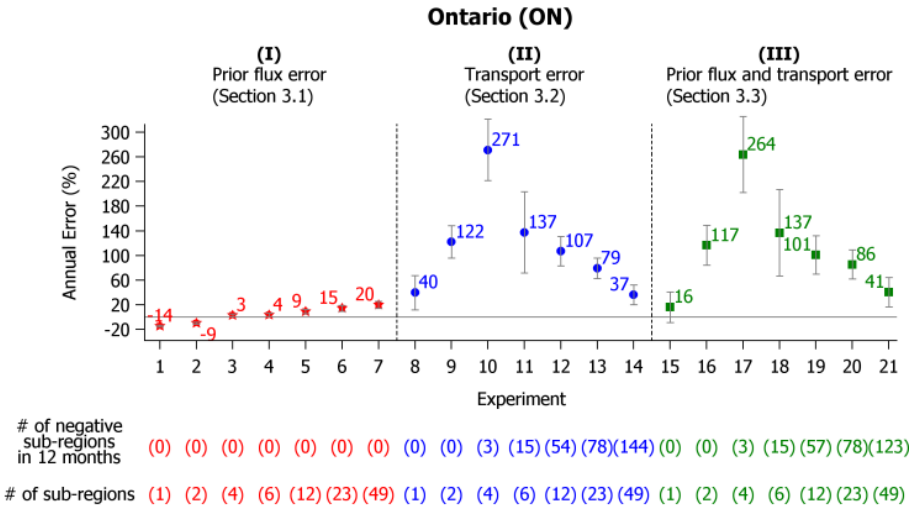
Figure 5 (re-plotted to improve legibility)

b

MCMC



CFM



SPECIFIC COMMENTS

line 10: 'Increasing the number ...' Does this mean that none of the set ups is significantly different from 'unstable' and 'unrealistic'?

Authors' response: 'Increasing the number of sub-regions (unknowns) beyond "optimality" can produce unstable and unrealistic fluxes for some sub-regions, and does not yield significantly different flux estimates overall.'

The sentence is saying: comparing inversion results from few sub-regions (or unknowns) to many sub-regions, the posterior flux estimates could improve (smaller posterior error) as the number of sub-regions increased. But as the number increase more and more, sub-regions with large posterior flux errors appeared. Thus there could be a number of sub-regions we can informally call 'optimal'. Another feature evident is that the posterior flux estimates (summed up for the provinces) could improve as the number of sub-regions resolved by the inversion increased, and then reach a level with little or no improvement with further increase in the number of sub-regions used in the inversion.

We will focus on the main message in the abstract and defer the others to the main text. Revised sentence: Increasing the number of sub-regions (unknowns) could produce large errors in the posterior fluxes (negative fluxes) for some sub-regions.

line 13: 'prior R2 ~0.8' Wouldn't it be better to quantify transport model error using the true fluxes (otherwise it is unclear which part of R2 is due to the prior flux uncertainty).

Authors' response: Yes. The true (target) fluxes were used to produce the synthetic observations which were simulated by CarbonTracker. We will specify this in the abstract.

line 16: 'a poorly simulated station' Why not just mention the station here?

Authors' response: Yes. Will include the name of the station.

line 19: 'improvements are needed with the current inversion setup ...' It seems that the data availability is the problem. This sentence suggests that an improved setup can compensate for missing measurements. It may be that the problem is in the word 'setup', but then this should be formulated more clearly.

Authors' response: Yes. We will replace "inversion setup" with "transport model". In this study, the amount of data is sufficient to yield improved posterior flux estimates when there is no transport error. When there is large transport error, the inversion could generate unrealistic results (negative fluxes).

line 28: 'These atmospheric mole fractions...' Here reality is described as if it is a model. Please change the formulation to avoid confusion.

Authors' response: We will change the sentence to: 'These atmospheric mole fractions are the results of the GHG emissions (sources and sinks) being transported by the atmosphere to the measurement sites (receptors).

line 75: What is a positive definite flux distribution. The term 'Positive definite' refers to a symmetric matrix.

Authors' response: Will remove the word "definite".

line 81: the spatiotemporal distribution of regional CO_2 and CH_4 fluxes are rather different.

Authors' response: Will replace "similar temporal and spatial distribution" with "similar slow-varying monthly variation". We did test using CH_4 spatial distribution as another (rather different) prior to evaluate the performance of the inversion model. The results showed similar level of improvement in the posterior fluxes (when there is no transport error). This result suggests that the inversion model is able to better deal with errors in the prior fluxes than transport errors.

line 94: The sensitivity of the estimation error and uncertainty to what?

Authors' response: "to prior flux, transport model, optimization methods and so on".

line 128: What makes these two estimates suitable as prior and truth?

Authors' response: In this study, we tried to examine how variations in the prior impact the inversion results (in reaching the 'true' or target fluxes CT-2011-ff CO_2) by testing 2 different priors (CT-2010-ff CO_2 and scaled CT-CH_4 fluxes). Which is addressing the question: 'are different spatial emission distributions suitable as a 'prior' to estimate the target (or truth)?' In general, it is assumed that the prior should be not too different from the truth. In our results, the impact from prior flux errors and/or uncertainties appears smaller than the impact of transport errors. Thus more focus should be on improving the model transport or better quantify the transport uncertainty. We note that the results could change for different inversion setups.

line 130: How were the fluxes redistributed to 0.2 x 0.2?

Authors' response: This line will be replaced by "The prior fluxes gridded at 0.2° x 0.2° were produced by assigning the same value (in $gCO_2/m^2/day$) of fossil fuel CO_2 emissions from the original resolution of 1° x 1° to all 25 cells within the 1° x 1° grid."

line 157: Something must have gone wrong with this definition of retroplume.

Authors' response: the definition is correct. Same as the one defined in other studies. (e.g. Stohl et al., 2003; Cooper et al., 2005, 2010).

line 161: The unit of the multiplication is kg/kg*m3 i.o. mole fraction

Authors' response: the flux was originally in moleCO_2 $m^{-2}s^{-1}$. The footprint was originally in s m^3 $kgAir^{-1}$. We will rewrite to "CO_2 mole fractions were constructed by taking the product of { ("footprint" in s m^3 $kgAir^{-1}$) , (the footprint layer (100 m)$^{-1}$), (the CO_2 fluxes at 0.2° x 0.2° in moleCO_2 $m^{-2}s^{-1}$) , (the molecular weight of air of 0.029 kg/mole) }, then summed up over all the grid cells…"

line 162: Prior means before here, rather than a priori, right? Please avoid confusion here.

Authors' response: Yes. We will make the changes.

line 174: How were the station specific baseline time series quantified? It sounds like you calculate the baseline contribution from the back plume initializations which you then subtract. However, in reality you don't have the 'true' initial values.

Authors' response: Yes, in reality no one will know the true initial values. Here "initial" we meant the baseline (background) concentrations that would be subtracted from the real observations. In real inversions, we would apply the same approach to calculate the baseline for each station by sampling the 5-day back concentrations from the 3D predicted concentration fields of a global model.

line 209: but you apply an a priori constraint to lambda, which is effectively equivalent to the regularization in CFM.

Authors' response: regularization is the term added to the matrix form of the linear regression model in CFM so that the parameters would be forced to the prior flux if the data-model uncertainty was too big. A constraint (statistical distribution) to lambda in MCMC was done in such a way the lambda became a random variable with wide sampling space.

line 222: If no regularization term is used in MCMC than how can it use the same prior error?

Authors' response: A statistical distribution was used in MCMC. "Assume λ_p follows normal distribution with a mean of 1 and a variance of 1 for $(\sigma_{prior})^2$, which corresponds to a 100% allowable error." The details are described in the Appendix.

line 249: Why is it necessary to compare means? Since in the Gaussian assumption mean and median are the same, you might as well compare medians.

Authors' response: It is possible to compare medians. We did not mean to imply it was necessary to compare means and means only. We will remove "Although the use of mean posterior estimates should be avoided (Tarantola, 2005), it is necessary here to compare the results using MCMC to those using the CFM method."

line 251: Why do you take the average of intermediate solutions in the iterative optimization? Shouldn't the optimum solution be the end point to which the iterative chain converges?

Authors' response: It is the statistical distribution of the chain of the random variables (lambdas) will converge (Appendix), not the chain itself. The use of the average of the converged distribution is the standard practice to ensure stable results [Chib, S. and Greenberg (1995), Congdon, P (2006)]. The end point will only be a single value. In fact, the first 10,000 (unconverged) estimates were excluded for the mean calculations as described in the text.

line 252: Since lambda is defined as time independent this sentence is not needed anymore (better would be to state explicitly that lambda is time independent at the point where it is defined).

Authors' response: We will add "lambda is time independent" at the beginning of section 2.5.1. We will delete "Same scaling factors of every three months would be used to calculate the posterior monthly fluxes."

line 263 - 265: But those scaling factors are intermediate solutions in a optimization process, therefore they are not independent optimal solutions of the inverse problem. For this reason, I don't see how the statistics of the scaling factors could represent the posterior flux uncertainty.

Authors' response: The goal of using all the estimates from the chain is to obtain the statistical distribution of the lambda as opposed to a point estimate.

line 321: I would rather call the prior flux error a disaggragation error, since it is mostly the spatial disaggragation which is different between CT2010 and CT2011.

Authors' response: Kaminski et al. (2001) explained that 'the aggregation error arises from biased (or inhomogeneous) sampling of the small-scale, unresolved source structure. We stress again that we generally will not know this small-scale source structure perfectly and we are trying to avoid the impact of errors in this structure.'

We will clarify the sentence to 'The spatial difference between the prior and target (when aggregated from grid points to sub-regions) could lead to errors in the inversion results, this error in the inversion result is sometimes referred to as the "aggregation error". This spatial difference between the prior and target will be referred to as the "source distribution error" '

line 338: It is still no clear to me what causes the baseline error in this inversion.

Authors' response: If the endpoints of the FLEXPART particles are at the 'wrong' locations (latitude, longitude and altitude) due to errors and biases in the meteorological data fields, then the mixing ratios sampled from the global model could have errors. This is our baseline error.

line 344: 'an example of one inversion experiment', which inversion experiment?

Authors' response: Experiment E21. We will this label.

line 363: Why is the same representation error of 30% used for all sites, when some sites are easier to simulate by the model than others? By the way, 30% is 30% of what? The deviation from the baseline?

Authors' response: Estimating model-data mismatch often include many assumptions and simplifications in practice.

For example, Zhao et al. (2009) defined the prior model-data mismatch Se as:

$S_e = S_{veg} + S_{part} + S_{eddy} + S_{transp} + S_{aggr} + S_{ocean}$

These diverse errors include particle number error (estimated ~10%), PBL mixing uncertainties (~22%), background concentration errors (~15%), etc. All these errors were assumed independent and summed to yield a total 'expected model prediction mismatch error' of 32%.

Similar error estimation for our setup yielded error ~30%. Given the many approximations necessary, in this study we decided to test a range of values for the model prediction mismatch error. The values 10%, 30% and 100% were tested. The inversion results were not very sensitive to the assigned value, and we presented the case of 30% model prediction mismatch error to be comparable with other studies.

line 374: 'representS'

Authors' response: change to "represent".

line 471: This paragraph refers to the same figure as the one before, but why then do you explain the figure here and not before?

Authors' response: This paragraph and Figure 6 will be moved to earlier section following Figure 4 in the revised manuscript.

line 483-484: I think it is clear that an improved fit to assimilated data is not the right way to validate inversion-estimated fluxes. What is done, however, is to test whether the optimized model does a better job simulating independent data (i.e. that were not used in the inversion). You could do this test as well, which would yield a more meaningful answer regarding inversion validation.

Authors' response: In a pseudo-observation inversion, the target is known and we were able to provide the posterior errors directly in this study. We did examine the constraining power for the different sites. This was done in a sense by dropping one station at a time in Section 5 (Figure 10).

line 493-494: It is not clear how the seasonal variation can become larger if the state vector is time independent.

Authors' response: The scaling factors were estimated every three months. Within each 3-month period, it is correct that we assumed the scaling factors are constant or time independent, but the scaling factors can change from one three-month period to the next. Therefore, this will translate into change in the seasonal variation.

line 517: But in set I the CT2010 fluxes were used.

Authors' response: Section 4 is a separate section. The results show that the inverse model works again in the western region even though anthropogenic CH_4 spatial distribution and 50% less than the target flux were used as the prior flux when there was no transport error. We will clarity this in the revision.

line 546: Whether or not the results can be considered significantly different obviously depends on the spatiotemporal scale over which fluxes are integrated. The scale should be specified more clearly.

Authors' response: We will specify provincial and annual time scales in the text, the spatiotemporal scale that we are interested.

line 547-549: This formulation is too vague and needs to be supported by actual numbers.

Authors' response: numbers will be added as follows: "Errors tend to be more positive using either the truncated normal (ranging from -13% to -2%) or the lognormal PDF (ranging from -8% to 15%) than those using the normal PDF (ranging from -15% to -2%) setup…"

line 553-551: It is not clear why going from 1 month to 3 months is increasing the observational constraints. What is expressed on the y-axis is the annual flux error. Whether this is composed of 4 block of 3 monthly fluxes or 12 blocks of monthly fluxes doesn't make a difference regarding the number of data that are used. The only difference would be difference in the temporal degree of freedom of the fluxes, but this is not the way it is explained in the text. This comparison needs to be explained more clearly.

Authors' response: Yes, we have fewer temporal degrees of freedom in the 3-monthly inversion case (the number of scaling factors per year is 4 x number of sub-regions), compared to the monthly inversion case (the number of scaling factor per year is 12 x number of sub-regions). We will rephrase the sentences to:

If transport model error is "unbiased", increasing the inversion time window for example, from 1 month (Figure S1) to 3 months (Figure 5) should effectively increase the observational constraint per parameter (total number of scaling factors for a sub-region decreased) and reduce the error or uncertainty. This is in fact the case, the estimation errors for AB+SK were substantially reduced by ~60% and stable results were obtained consistently for the two optimisation methods used in this study. However, because the transport error was large and likely not random for ON, regardless which optimisation method was used, increasing the inversion time window did not improve or stabilize the results.

line 559: It is unclear why the transport error statistics would be so different for two regions that are not very different regarding transport.

Authors' response: The difference appears to be the complexity of simulating near the land-water contrast within FLEXPART for the DOW site (~20 km from a large body of water, Lake Ontario). Thus we had larger differences in model-data comparison (Figure 4b). In contrast, the western region (AB+SK) has no large body of water and consequently the simulated transport appears to be relatively consistent between CarbonTracker (TM5) and FLEXPART.

line 561-563: Another possible strategy to do what? Please explain more clearly.

Authors' response: Correction: "Another possible strategy to reduce estimation error…"

line 572-573: But if the prior flux is the truth, then increasing the observational constraint is expected to increase the posterior flux error (an 'inversion' without any observations will yield the correct flux).

Authors' response: We tested the ideal case of no prior flux error and no transport error (not included in the manuscript). The inversion (essentially minimizing the least square errors) was able to yield the same

posterior fluxes as the prior fluxes. Therefore we found the observational constraint would only increase the posterior flux error if there were other errors in the inversion model (e.g. transport error).

line 610-612: Statistically it is not expected that the true flux is always within the 2 sigma interval. If the actual error exceeds the posterior uncertainty this could simply mean that the prior flux uncertainty doesn't properly reflect the prior flux error, or that the model-data mismatch doesn't properly account for transport model error. Both of these options are likely, given that ad-hoc assumption on these uncertainties were made in the inversion set-up.

Authors' response: It is correct that the true flux is not always within the 2 sigma interval (only ~95% probability). In Figure 7c, the estimates for the whole year in the Ontario case are outside the 2 sigma interval. We believe this is mainly caused by the transport bias error for the DOW, consequently our statistical optimization methods were unable to deal with such biases and yielded results with large errors for the posterior fluxes and uncertainties. We will explain this as well as the assumptions on the uncertainties we used in the revision (and in answers to comments from the reviewers).

line 622-624: Flux error contributions occasionally cancelling each other out are not a sign of non-linearity. If the inversion is linear, as seems to be the case here, you would actually expect the error contributions to add up. If they don't, it raises the question why this happens.

Authors' response: in the absence of the transport error, it appears the sub-regional flux errors are independent. In this case,

In general:

$C = MX$ *(where C=concentration, M=transport, X=flux)*

$C+dCx = M(X+dx) = MX+Mdx$ *(dCx=error in C from flux error)*

$C+dCm = (M+dm)X = MX+dmX$ *(dCm=error in C from transport error)*

$C+d(Cmx) = (M+dm)*(X+dx)$ *(dCmx=error in C from flux and transport errors)*

$D(Cmx) = dCm+dCx+dmdx$

Therefore, if M and X are correlated then their covariance (dmdx) is nonzero and errors are not linearly additive. This is likely the case for biased errors.

To illustrate this and other possible non-linearity clearly will require more detailed set of experiments. To avoid confusion, we will not mention non-linearity.

line 630-631: Unrealistic results for some months and sub regions are expected when increasing the degrees of freedom beyond the point that can be resolved by the data.

Authors' response: Yes, we will add: "This model evaluation (like the present study) is helpful in identifying the limits (how many degrees of freedom or sub-regions) for defining the inversion setup to avoid poor solutions."

line 635: A reference is needed here.

Authors' response: To avoid the confusions caused by the poor wording, the paragraph will be rephrased:

Two other possible sources of errors which include the representation and aggregation errors, and their impacts on the interpretation of the results in this study will be discussed next. Their estimated magnitudes compared to other sources of errors will give some measure of their importance.

line 646: But systematic differences in simulated concentrations during nighttime are probably not just caused by horizontal resolution.

Authors' response: We will remove this sentence. We agree our explanation was not complete. Horizontal resolution may not necessarily be the only determining factor. Other factors such as the vertical mixing could be quite significant.

line 635: If the representation error is not a concern, does this mean that the 30% uncertainty that is assumed was too large?

Authors' response: The 30% uncertainty for model-data mismatch error included many other possible errors (in addition to representation error) as explained above (regarding comment on line 363). We will explain the model-data mismatch error more clearly in the revision.

line 685: Given the difficulty to separate the contribution of aggregation errors from other errors, how do you know that using CH_4 as a prior causes the largest aggregation error? Where has this been shown?

Authors' response: Reviewer #2 had a similar comment. We will include an explanation in the revision as follows:

The source distribution (SD) errors for the two inversion regions using CT2010 and CT-CH_4 priors are defined as the following equations. The first equation below calculates the source distribution (SD) error resulting from the difference of the fossil fuel CO_2 spatial distributions between CT2010 and CT2011. The second equation calculates the error resulting from the difference of the spatial distributions between

CT-CH_4 (scaled to represent fossil fuel CO_2 as described in Section 4) and fossil fuel CO_2 CT2011. CT-CH_4 represents the anthropogenic CH_4 fluxes provided from CarbonTracker Methane.

$$\frac{\sum_{grid}|CT2010\text{-}CT2011|}{\sum_{grid} CT2011} \times 100\%$$

$$\frac{\sum_{grid}|CT\text{-}CH_4 - CT2011|}{\sum_{grid} CT2011} \times 100\%$$

Table below shows the source distribution (SD) errors for the two inversion regions investigated in this study. This result shows that even if the source distribution (SD) error (the cause of aggregation error in the inversion) is large (>100% by this measure), it does not play an important role in the inversions as discussed in Sections 3.1 and 4.

Inversion region	*Prior flux*	*Target flux*	*SD Error*	*Section*
AB+SK	*fossil fuel CO_2 CT2010*	*fossil fuel CO_2 CT2011*	*28%*	*3.1*
ON	*fossil fuel CO_2 CT2010*	*fossil fuel CO_2 CT2011*	*29%*	*3.1*
AB+SK	*Anthropogenic CH_4 CT-CH_4*	*fossil fuel CO_2 CT2011*	*182%*	*4*

line 698: What is meant with ’degree of spatial resolution’?

Authors’ response: Change “best degree” to “optimal”.

line 726: What do you mean by optimization procedure error?

Authors’ response: the optimization procedure errors of the MCMC and CFM methods.

line 776: What is the difference between equation 1 and A1? (same question for 2 and A2)

Authors’ response: Yes. A1 and 1, A2 and 2 are the same. A1 and A2 in the Appendix will be removed.

line 874-877: This is not right. The most commonly used inverse modelling methods define the state vector elements as random variables.

Authors’ response: We were trying to explain the CFM method solved for the scaling factors by computing the matrices defining the vector of scaling factors in Equation 5. This is a single computation. In contrast, the MCMC method requires large number of simulations to find the posterior estimates and the associated statistical distribution. But lines 874-877 are incorrect as the reviewer noted and will be removed.

line 917-919: This is not right. The size of matrices in analytical inversions is limited by computer memory, but this happens for state vector sizes that are much larger than ’only a few parameters’.

Authors’ response: This last paragraph is poorly worded and confusing. The whole paragraph will be removed.

line 925-927: In many cases the inverse problem is approximately linear, and the statistics not far from normal. In this case, you won’t have multi modal distributions and the use of means, or medians together with an estimate of the width of the distribution is perfectly fine.

Authors’ response: This last paragraph is poorly worded and confusing. The whole paragraph will be removed.

TECHNICAL CORRECTIONS

line 5: ’analysis’ i.o ’analyses’

Authors’ response: corrected.

line 52: remove ’with known bounds’

Authors’ response: removed.

line 129: ’are summarized’

Authors’ response: corrected.

line 188: The GMD formatting policy is to use bold roman for matrices.

Authors’ response: corrected for all matrices.

line 315: ’using’ i.o. ’used

Authors’ response: corrected.

References

Bruhwiler, L., Dlugokencky, E., Masarie, K., Ishizawa, M., Andrews, A., Miller, J., Sweeney, C., Tans, P., and Worthy, D.: CarbonTracker-CH_4: an assimilation system for estimating emissions of atmospheric methane, Atmos. Chem. Phys., 14, 8269-8293, doi:10.5194/acp-14-8269-2014, 2014.

CarbonTracker: CT2011 results provided by NOAA ESRL, Boulder, Colorado, USA, available at: http://carbontracker.noaa.gov, 2010, 2011.

Chib, S. and Greenberg, E.: Understanding the Metropolis–Hastings algorithm, Am. Stat., 49, 327–335, 1995.

Congdon, P.: Bayesian Statistical Modeling, 2nd edn., John Wiley & Sons, New York, USA, 2006.

Cooper, O. R., et al.: A springtime comparison of tropospheric ozone and transport pathways on the east and west coasts of the United States, J. Geophys. Res., 110, D05S90, doi:10.1029/2004JD005183, 2005.

Cooper, O. R., Parrish, D. D., Stohl, A., Trainer, M., Nedelec, P., and Thouret, V.: Increasing springtime ozone mixing ratios in the free troposphere over western North America, Nature, 463, 344–348, 2010.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, J. Geophys. Res., 108, 4756, doi:10.1029/2002JD003018, 2003.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO_2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457-475, doi:10.5194/bg-9-457-2012, 2012.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions, Biogeosciences, 10, 6699-6720, doi:10.5194/bg-10-6699-2013, 2013.

Stohl, A., C. Forster, S. Eckhardt, N. Spichtinger, H. Huntrieser, J. Heland, H. Schlager, S. Wilhelm, F. Arnold, and O. Cooper: A backward modeling study of intercontinental pollution transport using aircraft measurements, J. Geophys. Res., 108, 4370, doi:10.1029/2002JD002862, D12, 2003.

Zhao, C., Andrews, A. E., Bianco, L., Eluszkiewicz, J., Hirsch, A., MacDonald, C., Nehrkorn, T., and Fischer, M. L.: Atmospheric inverse estimates of methane emissions from Central California, J. Geophys. Res., 114, D16302, doi:10.1029/2008JD011671, 2009.