

Reviewer #1

General Comments

More Widespread Discussion of Error Estimates:

We agree that we could have done a better job discussing our results in relation to input and output uncertainties. We went through the text and figures and tried to improve this. Uncertainty is often neglected in inverse model studies, and we feel it's an important diagnostic.

Text:

We noted the estimated uncertainties throughout the sections that discuss the emissions estimates, including the Abstract and Conclusions. Where possible, we tried to interpret our results in the context of the posterior uncertainties and how they change from priors. We also added a short paragraph to the end of section 2.1 about the posterior covariance:

As noted above, the posterior covariance matrix is approximated by using the posterior parameter deviations. Temporal covariance is limited to the period spanned by the assimilation window. Therefore, time aggregated quantities, such as annual uncertainties will likely be overestimates since information about temporal covariations will be limited. Furthermore, as with any inversion, the error covariance matrix ultimately reflects the relative weighting between the model-data mismatch errors and prior emission uncertainties that are specified.

Figures:

Regarding the zonal average figures (8-11), the only error bars are on the observed curves and they are very small. They are calculated using a bootstrap method that quantifies the effect of spatial distribution of observation sites on the global (or zonal) average. We pointed this out in the captions.

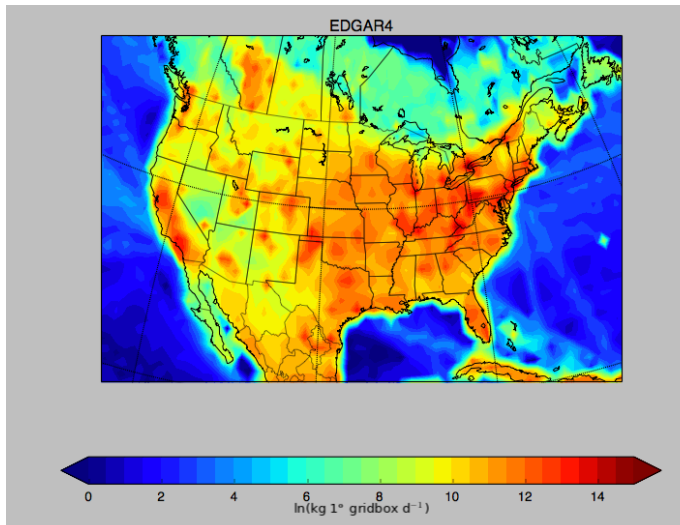
For the histograms (or bar graphs) we felt that adding the error bars for the total emission estimates would further confuse what are already not very easy to interpret plots. We mentioned the average uncertainty estimate in the caption so that readers can compare to the bars. Also, we pointed out that the errors represent the 1-sigma error bounds in each caption.

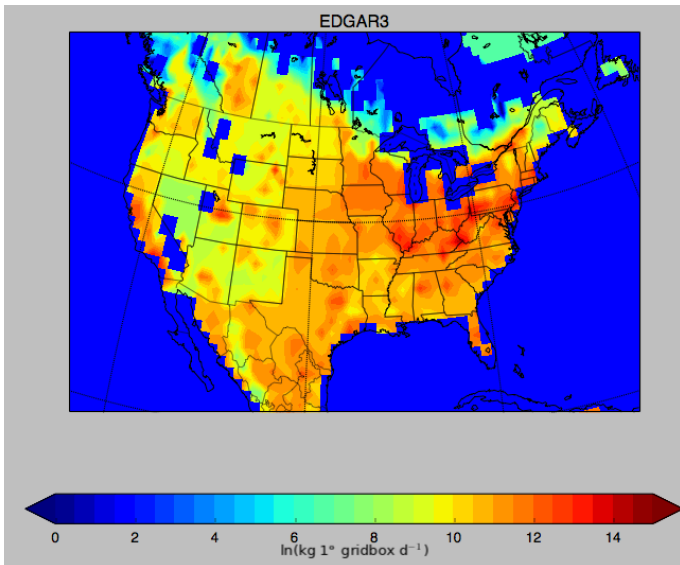
We added error bars to Figure 18 that shows the tropical flux anomalies. While it's true that the error bars are quite large, they are not large enough to disqualify what we say in the text.

Use of Different Versions of EDGAR:

We agree that the underlying distribution of anthropogenic emissions is an important factor in the result regarding apparent increases in fossil fuel emissions for North America. Below we show comparisons of EDGAR 3 and EDGAR4.2 for Jan. 2000 (recall that we use constant emissions). It's clear that the emissions are higher for EDGAR 4.2, especially in the area of interest, and the global total is about 30 Tg/yr higher for EDGAR 4.2. We will likely explore use of different anthropogenic emissions in future versions of CarbonTracker-CH₄, however, our aim in this paper was not to test bottom up emissions inventories, but to see what the effect of keeping emissions constant would be. We were especially interested in the sensitivity of our observing system to quantifying anthropogenic emissions. The lower global total prior we use (along with the fact that it doesn't grow over time) plays a role in the low bias we have globally, but there are also significant uncertainties in natural emissions and the chemical loss.

We added a statement about the possible implication of using low prior anthropogenic emissions to section of the paper that describes North American emissions.





The Shift of Emissions from High Northern Latitudes to the Tropics and Southern Mid-latitudes:

Work on improving the N-S transport in TM5 is ongoing and some possible improvements are being tested. The possibility that TM5 underestimates emissions at high latitudes because it traps emissions at the surface has been a concern, however, comparisons of inversions included in Kirschke et al. (2013) show that results obtained using the TM models are not that different from results obtained using non-TM models.

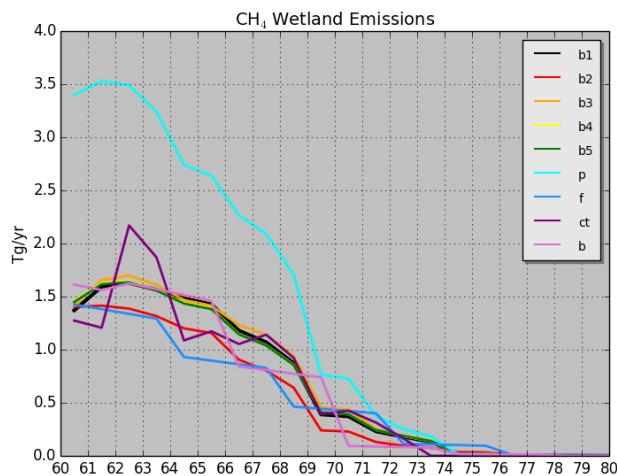


Figure – Comparison of inversions included in Kirschke et al (2013) for high northern latitudes. Note that the TM inversions are violet and purple, while the other inversions use non-TM5 models. (The x-axis is latitude).

ERA-I vs OD Meteo fields:

Early in the development of CarbonTracker (CO₂) a decision was made to use the OD met data. At that time, we did not have ERA-I and other reanalyses did not cover the time span we were interested in. Also, we hoped that we could keep the assimilations very current, although time lags in the availability of observations in practice have still meant that we lag real time by at least 1 year. For example, we don't receive samples from the South Pole during austral winter. Now that ERA-I is available, it will be our first choice for future simulations, however, comparison forward simulations suggest that differences between ERA-I and OD for CH₄ at surface sites is very small, both before and after the change in the vertical levels. Assimilations run with both met data sets for CarbonTracker (CO₂) produce virtually indistinguishable results in estimated fluxes (these results may be found on the CarbonTracker (CO₂) web site). On the other hand, comparisons with high altitude measurements obtained using the aircore technique suggest that the high-altitude structure of CO₂ is represented better with ERA-I, so this product will be a better choice for multi-decadal inversions.

Initialization of the Assimilation:

We neglected to mention how we initialized the assimilation, so we thank the reviewer for pointing this out. We don't believe that the initial conditions propagate very far into the time series of estimated emissions based on synthetic data tests. We added the paragraph below to the end of the first part of section 2:

We initialized the assimilation using an equilibrated distribution produced by another TM5 run that was scaled to match observed zonal average CH₄ mixing ratio for the year 2000. The north-south gradient therefore should represent the observed atmospheric gradient at the surface. Sensitivity runs using synthetic data (not shown) suggest that spin-up effects are restricted to within in the first half year of the assimilation.

As for why the low bias compared with observations occurs, we believe that this is due to prior emissions being too low (indeed this is by design in the case of the anthropogenic emissions) or the loss being too high. In other words, the prior sources and sinks will cause a lower equilibrium value than what the atmosphere would approach if sources and sinks don't vary.

Specific Comments

P2179, line 10: There is no mentioning of the time dimension of the state vector. I presume the 121 refer to a single month?

We added a sentence describing the weekly time step and the assimilation window of 5 weeks.

P2179, line 19: What is meant by satellite observed "hot spots"? Fire counts, burned

area, ..?

We clarified this statement to read: “The final terrestrial emission category is biomass burning, which is treated as a separate category due to the existence of strong spatial constraints coming from satellite observations of locations of large fires.”

P2180, line 11: This argument is more often used to justify short assimilation windows. I wonder, however, if there is any evidence of transport model errors accumulating over time. One may argue also that errors representing synoptic scale variations may dissipate on longer time/spatial scales that are better resolved by the course resolution transport model. Much of the observational constraint that inversions make use of come from larger scale mass balances. By reducing the response functions, this signal may end up being aliased to shorter scales. It is difficult to quantify the significance of this, but a more careful formulation seems needed here.

We added a reference to this issue of accumulated transport errors. We tend to agree with the reviewer that transport errors will likely cancel over long temporal and spatial scale, however, this issue has been the subject of heated debate within our group! CarbonTracker was originally developed to treat a dense observational network, however, budget issues ultimately meant far fewer sites that we had originally hoped for. In meantime, computational limitations and changes in input met fields discouraged us from lengthening the assimilation window. Future versions of CarbonTracker-CH4 will have longer assimilation windows, however.

P2183: Since Bergamaschi et al 2007 refers to an inversion, a reference is needed of where natural wetlands emissions come from that where used in that study (or the model that was used to generate them).

We used the wrong reference here – it’s Bergamaschi et al. (2005). We also added more details about this prior: the wetland prior is based on the distribution of Matthews and Fung (1989) and the emission model of Kaplan (2002).

P2187, line 15: Which global model is ‘a global model’?

We obtained the OH fields from Krol et al., and these were produced by a full-chemistry version of TM5 that was adjusted to agree with methyl chloroform. We clarified this in the text.

P2188, line 18: The model resolution of 6x4 degree seems more relevant here than the 1x1 degree of the emission inventories. Besides this, the inversion doesn’t allow changing small-scale emission patterns. It makes me wonder how valid it is to include tall tower measurements in the analysis. An additional error on top of the representation error seems needed here.

We realize that the resolution is an issue (not only for this inversion, but for many others that simulate transport at relatively coarse resolution). The same issue about representing local sources with such a transport model also applies to the use of continental air samples taken using flasks since these are essentially point measurements and not even afternoon averages. Using the background sites is safer, but then the observational constraints in the inversions would be restricted only to the largest scales. We chose the model-data mismatch errors to be large enough to account for uncertainty coming from the transport model and the assumed local distribution of sources based on both forward runs and posterior differences from observations. We find that

with the exception of sites that are near strong local sources, we generally get posterior residuals (simulated – observed) that are within our model data mismatch. Our web site shows figures at all sites, but these results are summarized in Table 2 in the ‘bias’ column.

Some figures are either quite small (7 and 10). This is true also for Figure 9, but that one doesn't seem to provide much information and could probably be left out.

We separated Figure 7 into two separate figures and revised Figure 9.

Technical Comments:

We thank the reviewer for pointing out these problems, and we particularly regret our oversight with the Figure 5 caption! We fixed all of these problems.

Reviewer #2

General Comments

1. Discussion of Recent Previous Work:

We included discussions of the Bergamaschi et al. (2013) and Houweling et al. (2013) studies throughout our paper. It is difficult indeed for us to keep up with the excellent work of these authors. We added some text to our introduction about how our study differs from these studies.

We don't think our use of the same transport model as Bergamaschi and Houweling is a big concern since the results coming from inversions are very sensitive to how the problem is set up. Choices of priors, observations and their input uncertainties are important factors in determining solutions. Analysis of results and insights into what they mean can also distinguish studies that appear to be similar. Biases in transport could mean that all inversions using a particular model or meteo product are subject to the same biases in emissions, and this may be the case with TM5 as well.

2. Limitations of the Modeling System:

Aggregation error and representation of spatial distribution of emissions in priors:

We added a paragraph to the end of the section describing the inversion that addresses aggregation error (p 2179) along with the Kaminski reference.

In this study we estimate emissions for continental scale source regions, and although we rely on the prior spatial distribution of the prior emissions to distribute the emissions, the use of large source regions can lead to aggregation errors as shown by Kaminski et al. (2001). An alternative would be to solve for many more sources, possibly at grid scale. However, without significantly more

observational constraints, our solution would be very dependent on not only the prior emissions, but also their assumed spatial and temporal covariance. Ultimately, use of space-based observations might be the preferred solution. At present, significant issues with space-based emissions still exist, such as quantification of biases that vary with space and time (e.g. Houweling et al., 2013). On the other hand, as discussed by Bruhwiler et al. (2011), the global network can constrain certain aspects of the budgets of greenhouse gases, even with its bias towards background atmospheric sites.

Regarding possible inaccuracies in the spatial allocation of emissions by the EDGAR dataset, we point out that this sort of information is part of what we hope to evaluate by doing these inversions. Indeed, with more observational constraints inversions may one day be useful for assessing the impacts of emission reduction policies. At present we lack adequate observational constraints that can help verify emissions, and in the current work, our relatively sparse global network of mostly remote sites is not likely to be very sensitive to regional errors in emissions. We do have more observational constraints for North America, and this issue is discussed in more detail with regard the bias at the continental site in Oklahoma that is near rapidly expanding oil and gas extraction. We also note that there are not currently many independent alternatives to the EDGAR emission database. We added the following to the section on fossil fuel prior emissions (page 2184).

In some cases, the spatial distributions of priors may not be accurate since they may be based on simple assumptions like population. For other emissions, there may have been changes in the spatial distribution of emissions over the decade, oil and gas drilling in North America for example. The atmospheric inversions allow the possibility of diagnosing these problems in the underlying prior emission datasets and may lead to improvements in methodology.

Model-Data Mismatch:

We added a discussion of how we came up with the model-data mismatch numbers (page 2188). It was an oversight not to have this because it is a critical component of the inversion, so we thank the reviewer for pointing this out.

Model-data mismatches were determined by assigning each site to a particular category; marine boundary layer (7.5 ppb), terrestrial (30 ppb), mixed marine and terrestrial (15 or 25 ppb), tower (25 or 30 ppb) and hard to fit sites (75 ppb). The model-data mismatch values were based on evaluation of forward runs and experience gained from CarbonTracker (CO₂, Peters et al., 2005). We forced the assimilation to closely match remote marine background sites while some sites were given a very large model-data mismatch because they are likely

influenced by strong local sources. A complete list of sites and their model-data mismatches is shown in Table 2.

Smoother Window Length:

On page 2180 we have the text:

The length of the smoother window is restricted to five weeks for computational efficiency. Although the posterior flux estimates in relatively densely sampled regions such as North America were found to be robust with a window this short, regions with less dense observational coverage (the tropics, for example) are likely to be poorly constrained even after more than a month of transport and therefore not well resolved. As pointed out by Bruhwiler et al. (2005), a smoother window of at least 3 months is likely to make maximal use of remote network sites, however this may come at the expense of accumulated errors in transport. Even without the problem of a short smoother window, the sparseness of the observational network makes it difficult to resolve under-sampled regions such as the tropical terrestrial biosphere (Bruhwiler et al., 2011).

The ideal situation would be one in which we have a dense network (like for weather forecasting!) and then information about sources would propagate quickly to the nearest site. For this type of situation we would not need a very long lag. Unfortunately, the surface network has not evolved in a way that provides enough observational coverage even for relatively-densely sampled North America. We are limited by inadequate computing resources at the current time, so 5 weeks was chosen CarbonTracker (both CO₂ and CH₄) as a good trade-off between computational expense and propagation of signals.

ERA-I vs OD Meteo fields:

At the time we ran these calculations we did not have the ERA-I fields, and CarbonTracker-CH₄ used an older version of TM5 that was not able to use the ERA-I meteo. The new version of CarbonTracker-CH₄ we are developing will use ERA-I meteo. In the meantime, we ran forward simulations using both OD and ERA-I and did not find significant differences at surface sites. Also, we did not see differences across the vertical level change.

Work done by the CarbonTracker (CO₂) team showed differences in CO₂ inversions between the two meteo products that were insignificant and some of these results may be found at <http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/>.

More recent comparisons of upper air observations with simulations with both OD and ERA-I suggest that the upper atmospheric transport of ERA-I is likely somewhat better than OD. Multi-decadal inversions and inversions that use upper air observations will need to use ERA-I.

Use of a network that changes over time:

The use of a changing network has been a topic of frequent debate since the first attempts at doing global inversions. We have taken the approach of using all available observations even though this introduces noise into the inversion (e.g. Bruhwiler et al., 2011). We try to keep the changing network in mind when interpreting the results of our assimilation (for example we mention this with respect to SGP). The alternative would be to use a network comprised of sites that are present over the entire inversion, and would amount to neglecting observational constraints coming from newer sites, many of which are seeing terrestrial signals. In the extreme, use of the background sites only would mean that we are stuck being able to see the N-S gradient and not much else.

We also point out that using the data points, rather than a smoothed and filled time series (e.g. GLOBALVIEW) as was done in the early days of atmospheric inversions, also introduces noise. Pretty much everyone uses the actual data these days.

It's certainly possible that satellite observations with their vastly increased data coverage will be a solution to the problem we now have of a sparse and changing network. On the other hand, satellites also change over time. For example: the pixel degradation of SCIAMACHY. Also, it is questionable that satellites will allow us to see multi-decadal trends because of their limited lifetimes. As we learned from trying to quantify trends from space-based ozone instruments, instrument changes can be limiting. Use of in-situ observations is still a useful approach as long as the noise from network changes is kept in mind. In an ideal case, we would have so many observations that they would be redundant, as is often the case with weather assimilations!

3. Zonal Averages:

We added the following text to explain in more detail how we calculated the zonal averages.

Here we follow the approach taken by Bruhwiler et al. (2011) that uses the same sampling, filtering and smoothing procedure used to produce the observed global and zonal CH₄ abundances for both data and model output (see Masarie and Tans (1995) and web updates at esrl.noaa.gov/gmd for a description of the data extension procedure). Zonal averages are constructed using mainly marine boundary layer sites by removing a long term trend approximated as a quadratic function, deseasonalizing by subtracting an average seasonal cycle, and using a low-pass digital filter with a half width of 40 days. Importantly, the model is sampled at the same times as the observations and missing data are filled in the same way for both the observations and simulations. The simulated and observed zonal averages are therefore comparable.

In response to the reviewer's request, we added a figure showing the zonal average methane compared to observations for the southern temperate zone. We didn't originally include this figure because we suspected that it mainly reflected problems with sparse data in the tropics that we discussed in detail. We also re-numbered the figures so that the zonal averages are Figures 7,8, and 9.

The reviewer suggests moving the section on zonal average comparisons to after the discussion of residuals and before the comparison to the aircraft profiles, presumably because the zonal average analysis uses observations that are used as constraints in the inversion. Although we are willing to do this, we prefer to keep it after what we consider are the more common evaluation techniques. Although other authors have used the global average time series and sometimes the growth rate of global methane as evaluation techniques, we are not aware that the meridional gradient information we exploit by comparing with the observed zonal averages has been discussed in other publications. For example, recent work by Houweling et al. (2013) and Bergamaschi et al. (2013) show N-S gradients and average CH₄ for zones, hemispheres and the globe, but only for inversion results. Of course, they use space-based observations, and it is an interesting question whether the observed diagnostic quantities like global and zonal averages and growth rates can be constructed from space based observations and how similar these would be to surface derived quantities. We totally agree with the reviewer that these types of diagnostics should be standard tools for evaluating inversions, and that is why we emphasize it in this study.

Aircraft Profiles:

Over the past decade, the aircraft program has experienced significant highs and mostly lows in funding. As a result, some of the 22 sites that we sampled the model for have very short records. 17 have reasonably long data records and are ongoing, and 8 of these are located in central North America, 3 on the west coast and 3 on the east coast. The remaining sites are in Alaska, Hawaii and Rarotonga. Sites located in central North America show similar features with respect to the inversion. Likewise, sites along the coasts are very similar. Rather than showing all sites, we would like to show sites that tell a story. We have chosen to focus on a clean west coast site, a central site, and a site near significant sources of anthropogenic emissions. We expanded the discussion in this section to reflect this. In the section on the tropical results we also mention the Beck et al. (2012) BARCA paper regarding the too-low prior emissions we have in tropical South America and the inability of the network to correct this.

Definition of High Northern Latitudes:

We agree that the use of Transcom regions doesn't easily lend itself to organizing results by zones as we try to do in this paper. Our next version of this inversion will have a different source region distribution. For this study, it was difficult to decide whether to put Europe with the high northern latitudes or mid-latitudes. We decided to provide a discussion of how much of the emissions may come from southern Europe since we

found that most of the fossil fuel and just about all of the wetland emissions come from northern Europe, while the agriculture and waste category is split between northern and southern Europe.

Specific Comments

abstract, page 2176, lines 9/10: "...a result consistent with previous;" I assume previous studies are meant ?

That's supposed to read: "a result consistent with work by Bergamaschi et al., (2005); however, unlike their results, emissions from wetlands..." " An entire line disappeared somewhere in the editorial process. We hope this problem isn't repeated elsewhere!

page 2178, line 12/13 " It contributes about half the radiative forcing of CO₂, 0.48_0.05 Wm² in 2010": according to the cited <http://www.esrl.noaa.gov/gmd/aggi/aggi.html>, the radiative forcing of CO₂ was 1.791 Wm² in 2010.

We meant that the forcing of CH₄ is 0.48, half of non-CO₂ gases. We fixed this so it won't be confusing.

page 2178, line 14-15 " it is 25 times more efficient per mass" I assume that the GWP is meant (defined as the time-integrated RF due to a pulse emission of a given component, relative to a pulse emission of an equal mass of CO₂ (IPCC AR5)) ? Add reference.

Yes, we are talking about the global warming potential here. We added the most recent number along with the correct IPCC reference (2013).

page 2178, line 9. "increase" -> "to increase"

Done.

page 2179: at which temporal resolution are emissions optimized (weekly / monthly ?)

?
We added a sentence describing the weekly time step and the assimilation window of 5 weeks.

page 2179, lines 13-17: from the listing it is not clear which are the single emission processes that are optimized (e.g., is "oil and gas production" one process or two separate processes ?).

We added a statement to clarify that we are optimizing the sum of these.

page 2181, lines 1-2: check sentence

We moved this sentence to page 2175 where we first mention the scaling factors.

page 2183, lines 4-5, "prior flux estimates of Bergamaschi et al. (2007)": In the cited paper 2 different wetland inventories were used - which of these has been used in the present study ?

We used the wrong reference here – it's Bergamaschi et al. (2005).

page 2183, lines 16 / 19: add references for the estimates of CH₄ emissions from oil and gas production, and from coal mining.

We added references for the estimates we discuss in this paragraph.

page 2184, line 4: Why has the old EDGARv3.2FT 2000 inventory been used (and not newer versions, which may have better spatial disaggregation for some processes) ?

For this study, we decided to use constant anthropogenic emissions for a year representative of the beginning of the run. As we explain in the text, we wanted to know whether the atmospheric observations are consistent with trends as large as those in the emission inventories. We did do some forward runs with more current versions of EDGAR, but we did not get very significant differences for the marine background sites that are the most influential in our inversion. The continental sites that we use in this study, such as SGP (discussed in more detail later in the paper), exist mostly in North America and distributions of fossil fuel emissions have changed so rapidly, that it's unlikely EDGAR 4 would have captured the variability either. We are currently testing emissions from a variety of other sources, so we intend to try to use the best priors available for future versions of CarbonTracker-CH₄.

page 2184, line 27: add reference of CH₄ emissions from rice agriculture. How is the seasonal variation of CH₄ emissions from rice agriculture modelled (EDGAR provides only annual total emissions).

We were sloppy with our references here and thank the reviewer for noticing this. We used Matthews 1989 for this and it has a seasonal cycle. We added the more detailed reference as well (Matthews et al., 1991).

page 2195 / 2196 and Figure 9: While the Arctic temperature anomalies in 2007/2008 had been discussed in previous papers [Dlugokencky et al., 2009], it would be interesting to extend here the analysis of correlation between CH₄ emissions and temperature to the entire 2000-2010 period.

We looked into doing this and it seemed that it would be difficult to extract a coherent story, so we decided to back off on the claims we made here.

page 2196, lines 13-14, " while fossil fuel and agricultural and waste emissions are distributed mainly in populated areas of Europe" these are mainly from mid-latitudes (see general comment (5))

page 2197, lines 17-18, " High latitude emissions of CH₄ from agriculture and waste are significant only for Europe" same comment as previous

We agree that the Transcom regions do not easily lend themselves to the analysis we're trying to do here. We are moving away from this source region configuration, but we're stuck with it in this study. We added estimates of how much of European anthropogenic emissions are coming from the south vs. the north with the dividing line between these based on the southernmost

extents of boreal North America and Eurasia. Furthermore, we point out that our framework and use of data was not designed to estimate emissions from Europe at high spatial resolution as the studies of Bergamaschi et al. (2005).

page 2197, line 26, "Walter et al. (2007)" : reference is missing

We added this. Thankyou for pointing this out.

page 2197, line 26, "Walter et al. (2007)" : reference is missing

We corrected this omission.

page 2198, lines 1-6, " CH4 is stored in ice hydrates.... The estimates of Shakhova et al. (2010) estimate the size of this source to be 10 TgCH₄ yr⁻¹ for the Arctic Ocean". Shakhova et al. (2010) do not claim that the CH₄ emissions which they estimate for the East Siberian Artic Shelf region are due to CH₄ hydrates.

We went back and looked at both the 2010 and the recent 2013 Shakhova papers. It is correct that hydrates are never mentioned as the source. Our confusion probably comes from oral presentations by Shakhova et al that have mentioned the possibility of hydrates being the source of the ESAS emissions. We thank the reviewer for noticing this, and we revised the text accordingly (also updating to the more recent study by Shakhova et al. (2013)).

page 2200, line 5, "EDGAR (version 4.3)", To my knowledge version 4.3 has not yet been released. Please check version number.

We changed 4.3 to 4.2.

page 2200, line 9, "Fig. 5": I assume this should be Fig. 6 ?

We changed this to Figure 6 (thanks for catching this!).

page 2201, line 18, " EDGAR emission estimates are too high": do you mean "emission estimates" or trend in emissions ?

We revised this section to discuss the Bergamaschi (2013) results in more detail, and in the process clarified that we were talking about trends.

page 2206, line 16, " agree well with observations, generally to within 15 ppb". Should be specified more accurately - do you mean average bias here ?

We substituted the rather vague range with the computed average bias over all sites.

Figure 5: labels at x-axis are missing

We fixed this.

Figure 6: "The growth rate is computed by taking the first derivative of the average mole fractions shown in the top figure": I assume that the concentrations are first deseasonalized

We clarified this in the captions for all figures like Figure 6.