Response to Reviewer Comments:

We thank the two Anonymous Reviewers for their thorough comments; we think they have helped improve the content of the manuscript.

In February (a few months after the manuscript was submitted) we discovered an issue with the treatment of pressure in the model. Our model assumed the pressures supplied by the NASA GMAO office were dry pressures when, in fact, they were moist. We have fixed this error and redone many of the simulations.

The figure below shows the impact of this correction. We found the largest impacts at the surface and the impact on the total column to be relatively small (~1.2 ppbv; with respect to CH_4 and other biases) because water vapor concentrations drop off rapidly with increasing altitude. Additionally, the bias in the total column is, for the most part, spatially uniform. This means it would have been folded into our bias correction shown in Figure 2 of the manuscript. We also tested to ensure our updated simulation was still conserving mass.
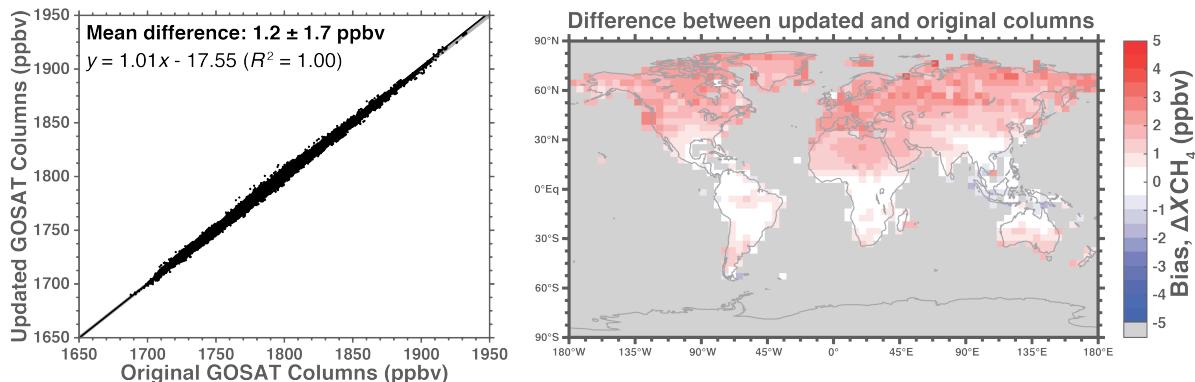

Figure 1: (**Left**) Scatterplot comparison of the original and updated GOSAT columns. (**Right**) Spatial differences between the original and updated GOSAT columns.

Furthermore, the TCCON retrievals have been updated to use the GGG2014 version and we have added additional comparisons to the NOAA/ESRL Tall Tower Network.

Reviewer #1 Comments:

1.) Attribution: As currently written, the manuscript implies stronger attribution than actually is possible in the observing-model framework presented. In the abstract and conclusion it is very strongly stated that "We attribute ....". In fact, the model-inversion framework does the attribution. Importantly, this relies on the prior distribution of source types, and the assigned uncertainty. As the author's state later in the manuscript depending on how one constructs the uncertainty either oil/gas or livestock are 'attributed' as the largest discrepancy in the US. The author's should clearly state in the abstract and conclusion that attribution is purely a model product, and relies on accurate spatial distribution of prior sources (or at least distribution of dominant sources) and is dependent on uncertainty assigned by the author's with expert judgment.

We feel that the attribution from this paper is actually more conservative than past work using similar observation-model frameworks and is a fair attribution for the sources. In the abstract and throughout the manuscript we present a range of attributions (e.g., 29-44% of anthropogenic emissions to livestock, 22-31% to oil/gas, etc.) that is based on inversions with varying specifications of the prior uncertainty. Before presenting the attribution in the paper we explain the underlying assumptions:

Lines 197-200: "We inferred the contributions from different source types to our posterior emissions by assuming that the prior inventory correctly partitions the methane by source type (see Appendix and Table 1) in each 4°×5° grid cell. This does not assume that the global distribution of source types is correct in the prior, only that the local identification of dominant sources is."

Further, these different specifications of the prior uncertainty provide a possible reconciliation for the different attributions in past work (e.g., Miller et al., 2013 and Wecht et al., 2014) that we discuss at length (Lines 327-336). The main conclusion and an avenue to address this limitation are copied below.

Lines 336-340: "With current prior knowledge it is thus difficult to conclusively attribute the US EPA underestimate to oil/gas or livestock emissions. This limitation could be addressed by a better prior knowledge of the spatial distribution of source types or by use of correlative information (e.g., observations of ethane originating from oil/gas) in the inversion."

We have changed the text accordingly:

Lines 16-18: "Using prior information on source locations we attribute 29–44% of US anthropogenic methane emissions to livestock, 22–31% to oil/gas, 20% to landfills/waste water, and 11–15% to coal."

Lines 366-368: "On the basis of prior emission patterns we attribute 22–31% of US anthropogenic methane emissions to oil/gas, 29–44% to livestock, 20% to waste, and 11–15% to coal."

2.) Representation error: GOSAT has a footprint with a diameter of ~10.5 km. The GEOS-Chem model is either being run at 4x5 or 1/2 by 2/3 degree resolution (~50 x50 km). The author's never discuss how they address the mismatch between the simulated column enhancements and the observed. Importantly, in regions of strong sources (and topography) and therefore strong XCH4 gradients, observations at 10.5km can and will see signals averaged out at 50x50km (even worse at 4x5 degrees). How are the authors dealing with this? What are the implications of this? In particular given the non-smooth nature of methane emissions from anthropogenic sources, the resolution of these model runs may not be sufficient to constrain fluxes at the uncertainties reported.

We have attempted to account for the representation error. The observational error variances (which include representation error) are specified based on the residual standard deviations (cf. Heald et al., 2004) unless the retrieval error reported for a given observation is larger than the residual standard deviation, in which case we use the reported observation error. Figure 2b from the manuscript shows a plot of the residual standard deviations for the

global inversion and we discuss this in the updated text:

Lines 165-172: "Observational error variances are estimated following Heald et al. (2004) by using residual standard deviations of the differences between observations and the GEOS-Chem simulations with prior emissions, as shown in Fig. 2b. This method attributes the mean bias on the 4°×5° grid to errors in emissions (to be corrected by the inversion) and the residual error to the observational error (including contributions from instrument retrieval, representation, and model transport errors). If the resulting observational error variances are less than the local retrieval error variances reported by Parker et al. (2011) then the latter are used instead. This is the case for 58% of the observations, implying that the observational error is dominated by the instrument retrieval error."

3.) Transport error: When linking atmospheric observations to fluxes, atmospheric transport is the key integral ingredient. Many regional, continental, and global studies go to great lengths to attempt to quantify the potential role of transport error and its impact on the inverted fluxes. This has not been addressed by the author's in this manuscript adequately for inverting fluxes of greenhouse gases. GEOS winds are not perfect, and the representation at 4x5 degrees and 50x50km might greatly impact the analysis and interpretation. What confidence can we have in the transport accuracy? Can we quantify the uncertainty in that term and include it in the inversion? Can we test if a bias error is possible? Many of the cited regional studies over California or North America make efforts to compare different transport as a proxy to understand transport error.

The residual standard deviations will account for some transport error, provided the error is not systematic. We have updated the text accordingly:

Lines 165-169: "…residual standard deviations of the differences between observations and the GEOS-Chem simulations with prior emissions, as shown in Fig. 2b. This method attributes the mean bias on the 4°×5° grid to errors in emissions (to be corrected by the inversion) and the residual error to the observational error (including contributions from instrument retrieval, representation, and model transport errors)."

Minor Comments:

1.) Abstract: Line 1-2. The GOSAT observations do not constrain the inversions, but are used in an inversion framework to constrain fluxes please correct wording here.

We have updated the text accordingly:

Lines 1-3: "We use 2009–2011 space-borne methane observations from the Greenhouse Gases Observing SATellite (GOSAT) to estimate global and North American methane emissions with 4°×5° and up to 50 km × 50 km spatial resolution, respectively."

2.) Line 3: It is confusing to see degrees and then kilometers for resolution. Since the model is run at 1/2x2/3 at higher resolution, that should be stated here (and ~50km can be in

parentheses, but the ~ is needed since it is only approximate).

We feel that kilometers are more intuitive at that spatial scale. Additionally, we phrased the latter resolution as "**up to** 50 km × 50 km" because we use the RBFs for the state vector. So we do not need the "~" because we have the "up to".

3.) Line 9-10: This makes it appear there is some circularity in the analysis the aircraft/surface data is used to address a bias and is then again used as an independent check. I don't think you can still call this an independent check at this point.

We have updated the text accordingly:

Lines 183-185: "Since we used them previously to justify a bias correction in the comparison between GEOS-Chem and GOSAT, they do not provide a true independent test of the inversion results."

4.) Page 4497, Line 12: As you cite later, this result has been found and reported previously, so should not be stated in the abstract as if a new, novel finding it is consistent with other studies.

We have updated the text accordingly:

Lines 7-9: "Our global adjoint-based inversion yields a total methane source of 539 Tg a^{-1} with some important regional corrections to the EDGARv4.2 inventory used as a prior."

5.) Page 4497, Line 21: The model framework does the attribution (with greater uncertainty that reported)

See our response to the first major comment.

6.) Page 4497, Line 25: "most important anthropogenic greenhouse gas" is more of a subjective statement better to explicitly refer to its climate-relevant role as done in the IPCC report being cited.

We have updated the text accordingly:

Lines 21-22: "Methane (CH_4) emissions have contributed 0.97 W m^{-2} in global radiative forcing of climate since pre-industrial times, second only to CO_2 with 1.7 W m^{-2} (IPCC, 2013)."

7.) Page 4499, lines 6-14. I am confused a bit here. The large spatial overlap in source types is not a problem you circumvent in fact this also limits you. Furthermore, it is misleading to imply satellite work has only focused on global scales you later cite work where satellite methane observations are analyzed are much smaller scales.

Sources are more likely to overlap at coarse spatial scales (e.g., 4°×5°), by moving to higher resolution some of that overlap can be resolved. We have clarified this in the text.

Lines 57-58: "Spatial overlap is reduced at higher resolution, thus optimizing emissions at high spatial resolution can help improve source attribution."

As for the latter point, here we are focused on "Previous inversions of methane emissions using satellite data". We begin by discussing what has generally been done ("mainly focused on coarse scales) and then discuss work by Wecht et al. (2014a,b) in the latter half of the paragraph that did inversions with satellites at finer spatial scales.

8.) Page 4502, Line 25. This seemingly implicates poor representation of stratospheric CH4 (and/or tropopause height) rather than a GOSAT artifact could you comment more on this? And what are the implications in regions with significant topography, where there is variance in the tropospheric column contribution to the total column?

We have attempted to investigate this issue using multiple datasets. For example, we have been collaborating with Kat M. Saad (Caltech; Wennberg group) who has been looking at the partial columns with TCCON (Saad et al. 2014) to try and diagnose the source of the bias. There isn't a conclusive cause for the bias yet. As such, we have not explicitly attributed the bias to GEOS-Chem or GOSAT and have left it open for future investigation. In any case, we have rephrased this discussion to hopefully clarify this section:

Lines 145-147: "This suggests a potential bias in the GEOS-Chem simulation of methane in the polar stratosphere, which warrants further investigation with observations such as TCCON partial columns (Saad et al. 2014; Wang et al. 2014). In any case, we remove the bias using…"

As for the second point, the model uses a terrain-following hybrid pressure-sigma vertical grid. So the model explicitly accounts for topography. We then regrid the model to the satellite vertical grid and convolve the model with the satellite averaging kernel. Additionally, by running the model at higher resolution we are able to more accurately account for variations in topography and don't need ad hoc topography corrections. Sub-grid scale variations in topography are unlikely to induce a bias because the model uses mean topography. Sub-grid scale variations in topography would lead to increased representation error but that would be captured by the residual SDs (Fig 2b).

9.) Page 4506, line 5. It could be misleading to state the errors are fully characterized. There are a large number of assumptions input including uncertainty levels, lack of systematic errors, and lack of covariances, which are very important in the total error assessment and are not included in the uncertainty range presented.

This simply means we obtain a full posterior distribution instead of the MAP solution. The companion paper (Turner and Jacob, ACPD 2015) discusses this at length. However, we have removed "full" and updated the text accordingly:

Lines 10-11: "…using radial basis functions to achieve high resolution of large sources and provide error characterization."

Lines 70-71: "…with up to 50km x 50km resolution and error characterization."

10.) Page 4507 line 21-22. I'm not quite sure if state-scale is well-defined constraining California does not imply any other states could have their flux quantified independently CA is very large and is on the ocean so it doesn't have much upwind sources. I would suspect a more accurate statement could be made here about the ability of GOSAT to constrain a specific spatial extent defined by the number of 50x50km boxes that can be constrained together.

We have updated the text accordingly:

Lines 266-267: "By using a longer time record and an optimally defined state vector we achieve much better success."

11.) Page 4508, line16-17. This might not be quite equivalent to assuming the prior distribution is correct but it is close and is highly dependent on the prior distribution.

The reviewer is correct in pointing out that this is dependent on the prior, however the latter part of that sentence explicitly states the assumption associated with this: "Again, this does not assume that the prior distribution is correct, only that the identification of locally dominant sources is correct".

12.) Page 4509, section 5. It may be useful to discuss the different time frame these studies focused their observation-inversions on. The different studies have been conducted focused on different years, and one could speculate that could contribute to the differences (though I find it more likely different observing network/transport/inverse strategy has a bigger role). Miller et al., 2012 and Wecht et al., 2014 had different years of focus.

Excellent question! We have ongoing work investigating the difference between these studies and is the focus of a manuscript that is in preparation. So it's not something we wanted to belabor in this manuscript.

13.) Page 4509 17-18. This is a bit misleading as stated. The Miller study had little to no sensitivity to many regions with stronger wetland sources (such as Florida), and so pre-subtracting wetland emissions is not really an important component of the analysis in that respect (wetland emissions from Florida could be increased by multiple Tg and it would not affect the Miller inversion this pre-subtraction matters, but it is not a simple Tg subtraction from the net).

This seems to be a circular argument. I fail to see how one can claim to have a US methane budget while simultaneously being insensitive to regions that could emit multiple Tg of methane. The Miller et al. paper report an estimate of US methane emissions. Additionally, we confirmed that the wetland source from their work was 2.7 Tg a-1 (Scot M. Miller, personal communication, 31 March 2015).

Further, the wetland regions are not necessarily confined to regions that Miller et al. was insensitive to. Examination of US wetland emissions from the WETCHIMP (Melton et al., 2013; Wania et al., 2013) shows many regions that the Miller et al. study was sensitive to with non-negligible emissions in the WETCHIMP models.

14.) Page 4510, line 15-16. This is a really important statement that needs to be made clear in the abstract and conclusion as well.

We agree that it is an important statement and that's why we listed ranges for all of the attributions in the abstract and throughout the manuscript. Additionally, the final paragraph of the conclusions reiterates this point:

Lines 372-373: "We find that either oil/gas or livestock emissions dominate the correction to prior emissions depending on the assumptions regarding prior errors."

15.) Page 4511 line 24. Suggest change 'We attribute' to something along the lines of: "The model framework attributes (with potentially larger uncertainty)"

See response to major comment 1.

16.) Table 2: What does the increase in all the mean biased post-inversion mean?

See response to comment #12 from Reviewer #2.

17.) Figure 7: There appears to be a plotting problem with the error bars.

The error bars were plotted correctly. As was mentioned in the figure caption, the error bar on the gray bar for the four sectors (oil/gas, livestock, waste, and coal) is from the sensitivity inversion. We have to clarified the figure caption:

Caption: "Error bars on sectoral emissions for our results are defined by the sensitivity inversion with 30% prior uncertainty for all state vector elements."

Reviewer #2 Comments:

This is an interesting and well-written paper. I have what I think are only small concerns that could use some clarification.

Minor Comments:

1.) P4500, L27 – Looking at Figure 1, I don't understand how the relevant spatial differences can be 10 ppb. It looks like they could be 4 times this number.

Fig. 1 shows the "raw" observations. The spatial patterns are, largely, a convolution of the sources and topography. Higher elevation would mean that a larger fraction of the total column is in the stratosphere (where methane concentrations drop off rapidly). Looking at regions of comparable elevation, we see relevant spatial differences of ~10 ppbv.

2.) P4501, L28-29 – This statement doesn't seem to be backed up with evidence. How do we know what the day to day variability is, and how would it follow that GOSAT can constrain the multi-year average? Please elaborate.

Central limit theorem allows us to "beat down the error" with more observations. So the observations may not be accurate enough to constrain a given grid cell but with many observations we can constrain a mean value. We found the seasonal cycle to be reasonably well represented in the prior, so we can use multiple years of observations to constrain the mean. We have updated the text accordingly:

Lines 95-96: "With a mean single-scene instrument precision of 13.3 ppbv, reducible by temporal or spatial averaging, GOSAT…"

3.) P4501, L14-24 – Do we have to worry that Xco2 produced by LMDZ is biased in places other than near large urban sources? The models can produce significant biases, especially in sparsely observed regions. What sort of errors can we expect to arise from this? Could comparisons with full physics retrievals be shown for other latitudes and locations?

This was something that we discussed with other co-authors prior to submission. We tested the impact of tested the impact of potential XCO_2 gradients or biases by simply replacing the model XCO_2 with a full physics XCO_2. So usually we obtain the proxy retrievals as:

$$XCH_4^a = \frac{XCH_4^*}{XCO_2^*} XCO_2^{\mathrm{model}}$$

where the asterisk is a non-scattering retrieval and the XCO_2 comes from a model. However, we can instead get a proxy retrieval using an XCO_2 from a full-physics retrieval:

$$XCH_4^b = \frac{XCH_4^*}{XCO_2^*} XCO_2^{\mathrm{full-physics}}$$

Taking the difference and rearranging gives us:

$$XCH_4^a - XCH_4^b = XCH_4^a \left(1 - \frac{XCO_2^{\mathrm{full-physics}}}{XCO_2^{\mathrm{model}}}\right)$$

We have plotted this difference in the figure below. We can see there are some differences
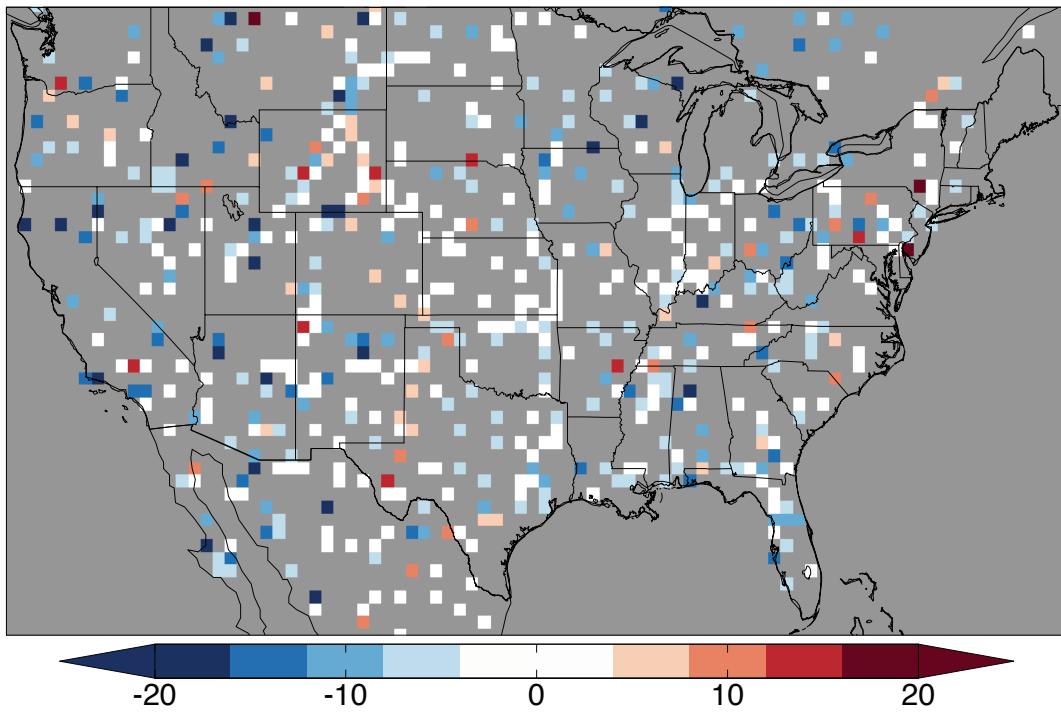
but we don't see any major systematic biases.

Figure 2: Difference between the proxy retrievals using a modeled CO_2 and the full-physics retrieval for CO_2: ($XCH_4{}^a - XCH_4{}^b$).

We have updated the text to reflect this.

Lines 115-117: "We determined the extent of the bias by replacing XCO_2 in Eq. (1) with (sparser) XCO_2 data from a full-physics GOSAT retrieval. This indicates a 14 ppbv low bias in Los Angeles but much weaker biases on regional scales."

4.) P4501, L25-27- It seems odd to say that the GOSAT will be compared to surface and aircraft obs. by using a model (and not even one that has emissions optimized with the obs.!). I guess the idea is that GEOS-Chem agrees with the obs, so therefore GEOS-Chem is a proxy for the obs. On the other hand, Figure A3 seems to imply a bias in GEOS-Chem that can be as large 20-25 ppb vs flasks. Is this considered unimportant? If so, please explain because this difference could be as large as the 10 ppb spatial difference mentioned above.

Good question! There is a lot of variability in the comparison of the model with surface flasks. A 10 ppb difference is quite small. We have added 1-σ error bars to the middle panel of A3. The uncertainties of the running medians all bound zero.

5.) Figure A3 - While I'm on the topic of Figure A3, I'd like to suggest that the N-S gradient figures be shown using a vertical axis that ranges from 1700 to 1900. It's hard to see the differences on this scale. The same goes for the middle figure, which could have a vertical axis ranging from -10 to 30 or something like that.

We have compressed the vertical axis of the N-S gradient figures.

6.) P4502, L10 - "Comparisons over North America with NOAA. . ." It's not clear to me what is being compared here - I think it's the model, but it could be GOSAT. Please clarify.

We have rephrased this:

Lines 128-129: "Comparing GEOS-Chem at 4°×5° over North America with …"

7.) P4502, L16-29 - I don't think I follow this discussion that determines the model stratosphere is the cause of the latitudinal bias between the model and GOSAT. The solar zenith angle is of course determined by latitude (and season), so how can latitude and sza be separated? Also, if it is the model, then how was the stratosphere determined to be the culprit?

Also, why does the model-GOSAT residual still show differences over Greenland (while it doesn't show differences over the Tibetan Plateau)?

I think it would be very helpful to compare the modeled stratospheric CH4 to other data, like stratospheric aircraft or air cores.

We touch on this in our response to Minor Comment #8 from Reviewer #1. We tried examining a wealth of observations and have not been able to conclusively attribute the bias to either GOSAT or GEOS-Chem. Solar Zenith Angle and latitude can be separated in a simple linear model. We construct simple linear models for the bias (read: multiple linear regression), used a model selection metric, the Bayesian Information Criterion, and found that linear models with latitude were a better predictor for the bias than SZA. This leads us to suspect the bias is in GEOS-Chem because, physically, SZA is the parameter that would affect the satellite retrieval, not latitude. Again, this is not conclusive and we have been working with Kat M. Saad (Caltech; Wennberg group) using their partial TCCON columns to evaluate the model and interrogate a possible stratospheric bias. However we have not yet found the source of the bias and think it is best to leave this for future investigation.

As for the Greenland question, there are very few observations over Greenland and they are quite noisy (see Fig. 2b).

We have rephrased the bias correction paragraph:

Lines 145-147: "This suggests a potential bias in the GEOS-Chem simulation of methane in the polar stratosphere, which warrants further investigation with observations such as TCCON partial columns (Saad et al. 2014; Wang et al. 2014). In any case, we remove the bias using…"

8.) Sections 3, 4 - It would be helpful to be more explicit about what is being estimated. Are parameters in the global inversion grid-box total emissions? Are they time dependent? Or annual totals?

We define the state vector (what is being estimated) in the first paragraph of the global and North American inversion sections:

Lines 159-160: "The state vector consists of scaling factors for emissions at 4°×5° resolution for June 2009–December 2011."

Lines 208-210: "We only solve for the spatial distribution of emissions, assuming that the prior temporal distribution is correct (aseasonal except for wetlands and open fires, see Appendix)."

9.) P4504, L28-29- I don't understand this statement. How does one trust the prior distributions of sources processes at a local scale, but not at global scales?

It's not that we trust the local distributions of sources more than the global scales but we trust the relative numbers more. This is far from ideal but we feel it's the best we can do given the current observing system. Ideally we would use observations of correlated species (e.g., ethane or $\delta^{13}C$) or a better prior. We elaborate on this in the conclusions:

Lines 373-375: "This limitation [source attribution] could be addressed in the future through better specification of the prior source distribution using high-resolution information on activity rates, and through the use of correlated variables in the inversion."

10.) P4505, L21-26 - I like this approach of reducing parameters, but these lines could use a bit more elaboration. What is the rationale for using the adjustments from the global inversions to cluster grid boxes?

Ah, that was a single iteration from an adjoint-based inversion at high resolution. The reason for including that parameter was to account for sources that may be present in the satellite data and missing from other "criteria". We elaborate on this in the companion paper (Turner and Jacob, 2015). We have also updated the text accordingly:

Lines 212-221: "The RBFs are constructed from estimation of the factors driving error correlations between the native resolution state vector elements including spatial proximity, correction from one iteration of an adjoint-based inversion at ½°×⅔° resolution, and prior source type distributions. Including the correction from the adjoint-based inversion allows us to account for sources not included in the prior."

11.) P4508, L16-17 - Here again is the statement about prior source distributions that I don't understand. How can we trust the allocation to different sources, but not the spatial distribution?

See response to comment #9.

12.) Table 2 - Why is the posterior mean bias larger for the posterior than the prior? Shouldn't the fit improve after inversion?

These observations are independent from the inverse modeling framework. Thus, there won't necessarily be an improvement. The posterior did improve the comparison to the independent observations by most metrics we examined.