Re: Estimation of mercury emissions from forest fires, lakes, regional and local sources using measurements in Milwaukee and an inverse method, by B. de Foy, C. Wiedinmyer, and J. J. Schauer.

Response to Referee #1:
Thank you for your comments. Please find below our response in italics inserted after the original comments in normal font. A revised manuscript has been prepared thanks to your review.

2 Major comment
• The results of this manuscript show very little contribution from local sources, likely as a consequence of excluding high measurements.

Referring to Table 5, we can see emissions from the grid cells within a 50 km radius of the measurement site and from the Great Lakes. The greater part of the local sources however are in the term for the local and regional background, as discussed in section 3.3. Furthermore, in section 4.2 we use a time scale analysis to evaluate the impact of excluding high measurements. Overall, we conclude that the present study is in agreement with the result of Rutter et al., 2008 that up to one third of GEM in Milwaukee is from local point sources.

– Their approach (47.27) to excluding outliers seems arbitrary. How is it justified?
It would appear to be an important assumption, as it causes the inversion to neglect high-frequency peaks in the measurements and thus estimate low values for local sources.

Least-squares methods are sensitive to outliers in the data. A robust fit can be obtained by removing the outliers or reducing the weight given to them. This process is dependent on the specific problem. We performed tests with different criteria and cut-off points for the outliers and made a selection that yields robust results.

The time scale analysis (Section 4.2) was performed specifically to address this question. High frequency peaks are more likely to be due to local sources. By using filter analysis we provide an alternative method of estimating these sources that is less sensitive to the least-squares fit of individual data points.

The outline of the manuscript has been expanded to point out more clearly the link between this analysis and the limitations of the inversion method, and the explanation of the time scale analysis has been expanded in Section 4.2.

– Previous studies of Wisconsin discussed in the introduction indicated significant missing local point sources, which would not be represented in coarse Eulerian grid models. The relevance of the method and approach in the present work is thus not clear. What can the authors conclude about such missing sources and the ability of their modeling framework to address these? At the very least, somewhere in the discussion and conclusion their findings should be compared to those discussed earlier from the literature.

There are two factors at work here: estimates of local sources using an inverse method are limited in accuracy because least-squares methods are sensitive to outliers, and because it is difficult to match individual peak concentrations exactly in numerical models. We therefore use the time scale analysis to provide an independent assessment of the local source. If we take the rapidly varying intra-day

component of the time series as being representative of local sources, this yields an estimate of 29%. This is consistent with Rutter et al., 2008, who estimate 33%.
The discussion and conclusions were expanded to describe this more clearly.

• 51.8-9: The authors state here that they do not use a priori emissions. This raises a few broad concerns.
– The results of their inversion are presented in terms of scaling factors. If a priori emissions are not used, to what are these scaling factors referring? I have no idea what the meaning of a scaling factor of 0.2 or 4 means given that I don't have an a priori emission to multiply this by in order to infer an actual absolute emissions.

We apologize for the confusion, the explanation has been clarified. We use zero a priori emissions for the backward trajectory grids. In contrast, the forward Eulerian simulations are based on an estimate of the emissions. The scaling factors apply to these emissions (described in Sections 2.4 and 2.5). Absolute a posteriori emissions can therefore be obtained by multiplying the scaling factors from the inversion (shown in Table 4) with the a priori emissions (Tables 1 & 2).

– If the inverse model is started from zero a priori emissions, then the results obtained are not absolute estimates of emissions, rather they are estimates of the emissions which contribute to the measurements in Milwaukee. My understanding in this light is in sync with the author's description of the inversion results as "contributions" and "impacts" in many places. However, they also in many other places (such as comparison to other Hg inventories) refer to their findings as emissions, not contributions or impacts. Yet these results have no bearing on emissions within the modeling domain that lead to Hg concentrations in locations other than Milwaukee. Hence direct comparison to total inventories is misleading, as is concluding that particular emissions sources are fundamentally over or underestimated. I thus have an issue with all of the discussions and conclusions related to absolute emissions, as well as the wording of the title (i.e., replace identification with contribution or impacts).

It is important to note that over the course of a year, Milwaukee is impacted by air masses from all directions. It is therefore possible to obtain estimates of emissions for the entire domain even though at any given time the concentrations are due to emissions from a small part of the domain.

With respect to the output of the inversion method, it is important to distinguish between 2 different products. The first product consists of actual emissions on the polar grid. By multiplying these by the trajectory impacts, one obtains a measure of the contribution of each grid cell to the concentration at the measurement site. The second product consists of the scaling factors on the forward Eulerian model simulations. By multiplying the a priori emissions in the Eulerian model with the corresponding scaling factor, one obtains the a posteriori emissions.

One can then estimate the contribution to the average concentration by averaging the simulated concentrations at the measurement site.

We are happy to change the title to "Estimation of mercury emissions…"

3 Minor comments
• Throughout: suggest Eulerian grid model or chemical transport model rather than 'grid simulations'.
Done, Thanks.

• 40: The overview of previous region studies in areas other than Wisconsin begins to feel a bit extensive and could be condensed.
2 paragraphs shortened and merged.

• 41.9: Why is the focus on the GEM measurements alone?
GEM has a long atmospheric lifetime (> 6 months) and so it can be treated as a relatively passive tracer: both deposition and chemical transformation can be neglected for the purposes of estimating sources. Obviously these are very important for estimating actual deposition and ecosystem effects, but in this paper we are focused on emissions alone. Treating GEM as a passive tracer simplifies the analysis and the inversion procedure.
We have expanded the justification in the outline.

• 42.15: What is meant by "eliminate all deposition effects" and why is this desirable? Is deposition not an important sink of mercury?
Please see above. Deposition is an important sink of mercury and is particularly important for calculating ecosystem effects. However because the lifetime of GEM in the atmosphere is at least 6 months, GEM can be treated as a relatively passive tracer for the purpose of estimating emissions.

• 42.25: I don't understand what is meant by the sentence beginning "It is important . . . " In inverse modeling, it is important to have the number of variable parameters be matched by the number of observations plus regularization constraints. If there is one variable per grid cell, then I suppose this aspect could be translated into a statement regarding grid resolution, but that wasn't clearly the intention. The resolving power of an observing system is also not trivial to estimate; how the authors determined what is an optimal grid size or modeling domain (i.e., number of variable parameters) is not apparent.

Indeed, the resolving power is not a trivial thing to estimate. If there were only one data point per variable in the inversion, the algorithm could potentially have a perfect match even though it would not be robust to changes in data selection. We have proceeded heuristically by testing different grids. The configuration used was chosen to balance a desire for high resolution but also to have sufficient constraints on each variable in the inversion. We have 371 variables for 3954 measurements, ensuring a ratio above 10 to 1 of data points per free variable. The description of the grid selection was expanded in the paper.

• 46.10: It wasn't clear here how the inversion parameter vector and sensitivity matrix are constructed. I think after finishing the article I had a better idea, but it might be useful to more explicitly define things here. From presentation of the results it seems that there are variable parameters (scaling factors) for each of several biomass burning regions, one global background region, and many grid-specific sources. But here xCAMx is defined as "scaling factors on the concentrations", so how these are used to invert for emissions scaling factors is not clear. Or are these used to rescale initial conditions? Then what are background values? So I guess I still don't understand what is xCAMx vs the other elements of x.

The background value is a single value obtained from the model. CAMx provides timeseries of concentrations using known inventories. The inverse method then calculates a scaling factor on the concentrations simulated with the prior inventory. We have expanded the explanation of the merged matrix in section 2.6 to be more explicit, thank you for the recommendation.

• 48.17: The wording here is odd, as we don't know what is meant by "times" until the subsequent sentence.
Wording adjusted, thanks.

• 49: This entire section is a bit strange. CFA wasn't really explained much. These results also don't seem to fit into the rest of the paper. They are not mentioned in discussion, conclusions or abstract. Overall, I would suggest skipping this analysis or incorporating it more completely into the paper.

Inverse models can be sensitive to model inputs. We have therefore sought in this paper to supplement the analysis with simpler and more robust methods. For this reason we feel that it is important to show the windrose analysis, the Concentration Field Analysis and the time scale analysis. It is in the same spirit that we applied the bootstrapped method to increase the confidence in the method.
The justification for this section was expanded to better integrate it with the rest of the paper.

• 50: The synthetic inversion test is a great way to assess the capabilities of an inverse model, and I'm glad the authors included this. However, there are a few aspects of their synthetic inversion that were not clear.
– Like the real inversion, were these also performed using zero a priori?
Yes, the test was performed without prior knowledge of the source location. This was clarified in the text.

– Was the location with the X the only source?
Yes, there was just one source. This was clarified in the text.

– This sentence: "Because the model simulates . . . " does not make sense. I don't think the model domain is the single reason. I think the inversion tends to overestimate emissions far away because there are no other observation locations to constrain such sources.
As you point out, if we were to add extra locations, we would have better constraints on the inversion. However in this work we are interested in situations with just one measurement site.
The wording of the sentence was simplified to emphasize that in this sentence we are not dealing with a causal explanation but merely a statement of fact: the overall over-estimation is a result of the non-zero emission values from grid points far from the source.

– The last sentence of the first paragraph at first seemed at odds with the third to last sentence. It took me a while to see where 68% comes from (82 out of 121). Maybe this could be stated more clearly.
This last sentence provides a different way of expressing the same data, as you point out. The sentence was clarified to reflect this.

• White is used in Figures 3, 7, 8, 11, 13 but it isn't in the color bar.
Added to caption.

• 52.17: It would be useful to also plot the time series of the observations that are actually being fit in the inversion.
Done

• 54.23: The text would imply that Table 3 compares model values to those from the inventories shown in Fig 8, but it does not, it only contains model values, and only from fire emissions. Suggest expanding

this table for better inter comparison of model with inventory (assuming this is even possible, given above discussion regarding contributions vs emissions).
We apologize for the insufficient caption. The tables shows a comparison of emission estimates from the inverse model with actual data from the National Emissions Inventory and the Toxic Release Inventory. As discussed above, the inverse model does provide estimates of the emissions for direct comparison. The caption has been expanded for clarity.

• 55.4: would be expected
Changed, thanks.

• 58.17: This assessment is a bit vague; can the authors be more quantitative in their summary?
Conclusions updated to be more quantitative.