

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 #2:

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.

1 Major comments

- Suitability of the site is a major concern. Urban sites are rarely used for regional scale inversions, due to the difficulty of simulating near-by sources in a regional model, and concerns about how representative the measurements are of the regional scale. Further details should be given about the sampling to address this. For example, is the instrument sited on a tower (where local source influence may be smaller), or in a street canyon (where local sources could be dominant)?

A key aspect of the original study was estimating the concentrations of mercury in Milwaukee and comparing these with the rural site near Devil's Lake. By locating the measurement site in the urban area, we are able to detect the impact of local sources. The inverse method shows that in addition to these sources, we are able to recover information about sources that are further away.

The sampling intake is 1.5m above the roof of a 2 story building. The site is on the southern edge of the campus of the University of Wisconsin - Milwaukee. There are some large structures to the north, but the site is otherwise surrounded by a residential neighborhood. This was added to the text.

Can we really believe that the variations seen are due to the interception of different large-scale air-masses, or the transport of local sources around the urban area?

The time scale analysis described in Section 4.2 shows that there are features in the data across the whole range of time scales, from the seasonal to the intra-day. The purpose of the work is to see if an inverse model is able to recover information from this. Clearly there are uncertainties in the method. Bootstrapping was used to evaluate these and provide histograms or ranges on the emission estimates.

In reality, I suspect that few alternative sites exist for this work?

This is indeed the case.

However, if this is the case, I would suggest that more work should go into identifying times when the measurements can be considered representative of regional-scale air. For example, Manning et al., (2011) try to identify potential local contamination (even at background sites) by identifying days with low boundary layer ventilation.

At a background site with well-defined urban areas in the vicinity it is indeed a good idea to separate the data into different episodes due to air-mass transport. In our case, the measurement site is in the urban area and the concentrations are impacted by sources from very different spatial scales in different combinations at all times. The inverse method is a way of dis-entangling this information, and the more data the better. We experimented with isolating episodes, but this did not significantly alter the results although it did reduce the number of data points available, and hence increase the uncertainty in the results.

The authors note that the local emissions may be causing significant discrepancy between the observations and model. However, the apparent assumption that the local sources can simply contribute to a 'background' offset added to the measurements seems unlikely to be the whole story (P12952 L23): local sources usually cause high variability and large pollution events under certain conditions, and concentrations that are close to background levels are possible even in urban areas.

Yes, local sources can be expected to cause high variability events. These pose a particular challenge for numerical simulations. It is for this reason that we supplemented the analysis using the inverse model with a wind-rose analysis, a Concentration Field Analysis and a time scale analysis. In particular, the time scale analysis showed that about one third of the variability was at short time scales that were not captured by the model. For this reason, we suggest that local sources are responsible for a significant fraction of the residual. The outline of the paper was expanded to discuss these issues and put the different analyses presented in this paper in context.

- Clarification of the inverse method is required to ensure that the results are robust. In particular it is very difficult to discern what constraints are actually being used in the inversion as it is presented:

1. As I understand the procedure, forward simulations of an Eulerian model (CAMx) are run for certain sources (fires and lakes), and a scaling factor is sought to multiply these reference concentrations. In addition to this, emissions from each grid cell in the domain are also derived using WRFLEXPART residence time analysis (RTA). The results then discuss the scaling factors as a measure of the increase or decrease to be applied to the lake or fire emissions. However, I am confused as to whether the RTA also covers the lake and fire emissions regions? If so, the multiplying factors will be correlated with the emissions derived from the 'RTA part' of the inversion, and so the real lake/fire emissions would have to be increased or decreased accordingly.

Yes, this explanation is correct, and yes, there is overlap of the regions. However the temporal profiles of the emissions are very different and so the time series are not correlated. If they were, we would indeed see emissions estimates in the RTA grids corresponding to the CAMx simulations and the results would have to be presented as a sum of the two. As it is, there is little overlap in the results and they can be presented separately.

2. A Bayesian inverse framework is derived in which a cost function containing measurement and prior emissions constraints is derived (Equation 4). Much of the next page or so discusses how the constraints on the observations or emissions are combined. However, on P12951 L8, the author states that a priori information is not used. I do not understand how the framework outlined in Section 2.6 can be used without a priori constraints? If a 'prior-free' inversion were desired, why not use a least-squares approach? In fact, the solution described on P12947 looks more like a tapered least-squares approach to me than a Bayesian method, in which case all of the discussion of measurement and model covariances etc. discussed in the previous page has been discarded. In which case, why mention it? Apologies if I'm confused by this section. A much clearer description of the inversion approach is required with clear justification for the choices made and how they are statistically robust.

*Yes, part of the purpose of the paper is to show that there is a relationship between the Bayesian approach and a least-squares approach as described by Wunsch, 2006:
"Much of what follows in this book can be described using very elegant and powerful mathematical tools. On the other hand, by restricting the discussion to discrete models and finite numbers of*

measurements (all that ever goes into a digital computer), almost everything can also be viewed as a form of ordinary least-squares, providing a much more intuitive approach than one through functional analysis.” (Section 2.4)

In the present work, we can perform the inversion with a least-squares method. We have not discarded the covariances, but rather shown how in practice they simplify to the regularization parameter which can be determined empirically. A further advantage of the least-squares formulation is that it allows us to enforce constraints and do bootstrapping.

Thank you for pointing this out which led us to find the relevant material in Wunsch, 2006. We have added it to the text and clarified the discussion accordingly.

3. If a 'Bayesian' approach was actually followed, then I have a number of concerns. Firstly, an emission covariance appears to be applied to the prior (Rb). This itself provides a weighting between the prior and the observations (which are weighted by Ra). To derive the cost function as written in Equation 4, the assumption is that the prior estimates and the measurements are uncorrelated. This is fine if Ra and x are not determined using the observations. However, on P12947 L17 onwards, it is stated that a scaling factor is iteratively applied that weights the Jobs and Jemis depending on the model measurement mismatch. This means that independence between the two parts of the cost function is lost and in fact there is little point specifying Rb a priori anyway. Similarly, the procedure of iteratively decreasing the a priori uncertainty on grid cells where negative emissions are obtained also removes the independence between the prior and the observations.

As described in Wunsch, 2006, we simplify the inversion to a least-squares approach. The derivation in Lorenc et al., 1986, does indeed rely on independence of the cost functions. In our case however, we do not specify Ra and Rb, but rather show that the problem simplifies to a tapered least-squares method as you point out. As a result, the equations simplify so that we have a regularization parameter alpha (“trade-off parameter” gamma in Wunsch, 2006). In this formulation, we do not make any specific assumptions about the cost function. Wunsch describes this parameter on pragmatic grounds: “It was long ago recognized that some control over the magnitudes of x, n, Cxx could be obtained in the simple least-squares context by modifying the objective function (2.107) to have an additional term” (Section 2.4).

4. The fit between the optimized model and the observations seems poor (Figure 10). In particular, the baseline seems to be higher than the observed background for much of the period (e.g. see Stohl et al., 2009 where the background is almost always below the smallest measurements). If this is true, does this not suggest that the derived emissions could be underestimated?

Yes, there is a high residual and the simulated background is high. The high residual is due to the complexity of simulating mercury concentrations accurately. This is a more widespread problem with mercury modeling compared to other pollutants that are better constrained. The high background arises from the method trying to compensate for the high residual. Because we know what the global and the regional background values are from Rutter et al., 2008, we suggest that most of the excess background is due to regional sources. This is further corroborated by the time scale analysis that shows that the simulations fail to represent the intra-day time scale. The discussion was expanded to include this.

2 Specific comments

Introduction: Examples of previous regional inversions, specifically using Lagrangian Particle Dispersion models should be given.

A paragraph was added to the introduction.

The use of the term 'grid models' throughout is confusing. Does the author mean Eulerian chemical transport models in this case? All models use grids of some sort or other!

This was changed to Eulerian grid models (see also Ref #1 comment).

P12942, L10: Are 1000 particles/hour enough to calculate robust footprints? Typically an order of magnitude more particles are released to reduce noise (e.g. Stohl et al., 2009 use 40,000 every 3 hours).

We performed tests using grids with 250 and 500 particles/hour which yielded the same results as using 1000 particles/hour. The figure below shows the comparison between 500 and 1000 particles/hour. We therefore conclude that 1000 particles/hour are enough in the present case. This is a similar finding to Wen et al., 2011 who use 3000 particles/hour. Note that the larger the spatial scale, the more particles are needed. Stohl et al., 2009 perform a global inversion and consequently need more particles.

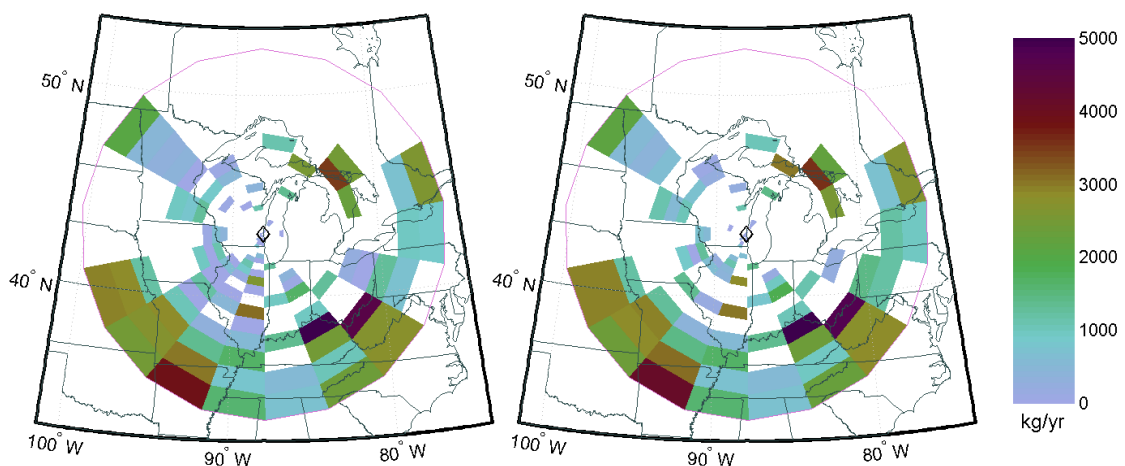


Figure 1: Left: Inversion results using 500 particles/hour, Right: Inversion results using 1000 particles/hour. Note that they are nearly indistinguishable.

Section 2.6: Un-accounted for sources of uncertainty in the inversion should be discussed.

In particular, model-measurement mismatch errors do not seem to be accounted for. Furthermore, aggregation errors are likely to be significant in the time domain, as (I think) annual-mean emissions factors are derived, rather than shorter time periods (e.g. Thompson et al., 2011). Systematic model errors are also likely to be dominant for these types of inversion (e.g. Gurney et al., 2002).

A new paragraph was added to Section 2.6 to discuss these issues.

P12946 L1: Is it simply measurement repeatability that goes into R_a ? Usually some estimate of model-measurement mismatch are also included (e.g. Chen and Prinn, 2006).

As discussed above, R_a is folded into the regularization parameter α . This parameter is determined empirically to balance the two parts of the cost function.

P12946 L1: R_a is the uncertainty covariance corresponding to the measurement (and model prediction of the measurements), not the sensitivity matrix.

Yes, thank you. The text was adjusted.

P12946 L26: Is D the identity matrix in this formulation? Further, I think that the transpose of the two matrices is required on this line (e.g. $y''=(y, x_{zero})^T$).

Both changed, thank you.

P12946 L28: I think that off-diagonal zeros are required for the combined R matrix.

Yes, R is the combination of R_a and R_b with zero terms in the upper-right hand and lower-left hand blocks. The description was changed in the text.

P12947 L25: I think units are required for this scaling factor.

The regularization parameter balances the cost function for the residual and the cost function for the emissions. It therefore has units of $((\text{mass}/\text{volume})/(\text{mass}/\text{time}))^2$ which are proportional to $(\text{time}/\text{volume})^2$. As described in Wunsch, 2006, it is determined empirically by the desire to balance the accuracy of the match with the magnitude of the vector x . In practice, I have never seen it reported with actual units. The text was adjusted to include a mention of this.

P12942, L16-19: Is this residence time analysis carried out for the surface grid cells? This needs to be clarified in this section (although it is mentioned later in the manuscript).

For CFA, it is performed in the entire vertical column. For the inverse model, it is performed in the bottom 1000m. This was added to the text.

P12948, L11: It is stated that a 1000m cell height is used for the RTA so that sufficient particle counts are obtained. Usually a lower height is used to ensure that the particles can realistically be expected to interact with the surface. Could this situation not be improved by again using more particles when running the model?

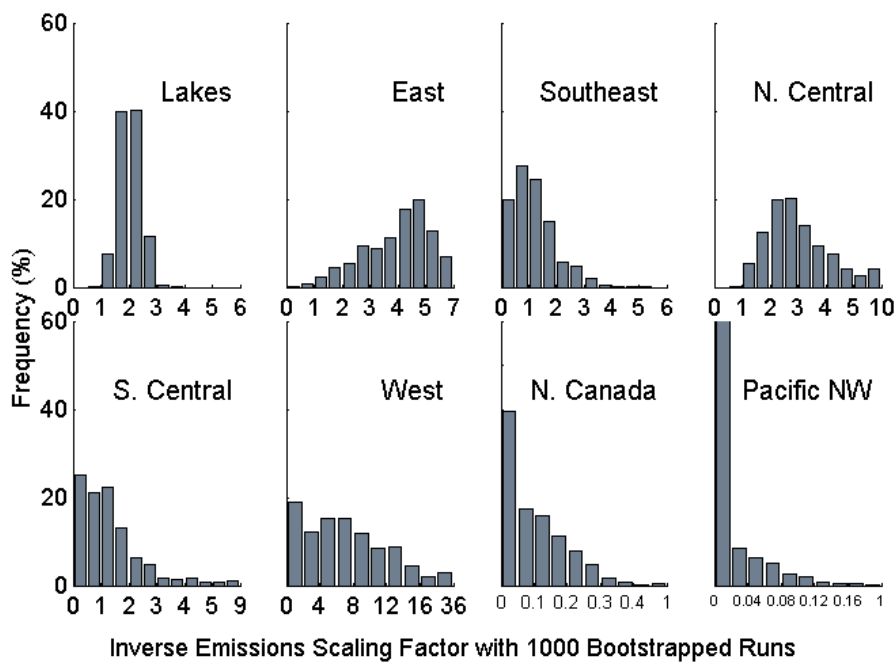
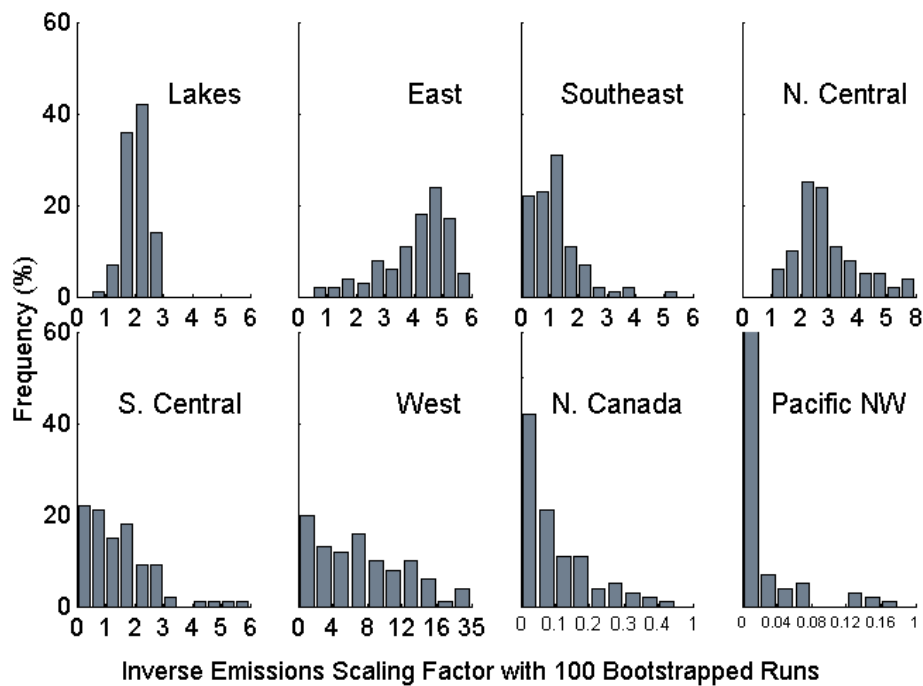
There is clearly a tradeoff between computational cost and particle coverage, but there is also a balance between the size of the grids and the capability of the model to resolve fine grid scales. After extensive testing, we found that for the current study 1000m was a robust choice.

P12948 L21: What is the basis for the 75

This was a practical expedient. In practice it made no difference and in the end we did not use it. The text was adjusted accordingly.

P12951 L5: Are 100 bootstraps sufficient to calculate robust statistics here?

We performed additional tests with 1000 bootstraps and this did not change the results. The two figures below shows results from a new set of runs with 100 and 1000 bootstrapped simulations. Text adjusted accordingly.



P12952 L29: Does the seasonal cycle in the fire emissions not stem from the emissions dataset used, rather than the inversion, since a single factor multiplies the CAMx entire time series?
Yes, this was clumsy wording. The text has been adjusted.

Figure 4: There appears to be a step-change in the time series after a break in the measurements towards the end of March 2005. Can the authors provide some meteorological explanation for this apparent increase in the baseline, or could this indicate instrumental issues?

The measurements were thoroughly quality controlled before publication of Rutter et al., 2008. We are confident that this is not a measurement issue. The apparent step-change occurs at a time of increasing biomass burning impacts and is most likely a feature of the fluctuating concentrations at that time.

References:

Lorenç, A. C.: Analysis Methods for Numerical Weather Prediction, Q. J. R. Meteorol. Soc., 112, 1177–1194, 1986.

Wen, D., Lin, J. C., Meng, F., Gbor, P. K., He, Z., and Sloan, J. J.: Quantitative assessment of upstream source influences on total gaseous mercury observations in Ontario, Canada, Atmos. Chem. Phys., 11, 1405–1415, doi:{10.5194/acp-11-1405-2011}, 2011.

Wunsch, C.: Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications, Cambridge University Press, 2006.