26 August 2014

Revisions for "Estimating sources of elemental and organic carbon and their temporal emission patterns using a Least Squares Inverse model and hourly measurements from the St. Louis-Midwest Supersite" by B. de Foy, Y. Y. Cui, J. J. Schauer, M. Janssen, J. R. Turner, C. Wiedinmyer, Atmospheric Chemistry and Physics Discussion, 2014.

Thank you for the reviews of our manuscript. Please find below a point by point reply to the comments along with a list of the changes made to the text.
The complete original reviewer comments are in black below, and the responses and modifications made to the manuscript are listed in blue.

Sincerely,
B. de Foy, Y. Y. Cui, J. J. Schauer, M. Janssen, J. R. Turner, C. Wiedinmyer

Anonymous Referee #2, Received and published: 10 June 2014
General comments
This study examines one year of hourly measurements of elemental carbon (EC) and organic carbon (OC) from the St. Louis Midwest Supersite. Using a least squares inverse model and atmospheric transport modelling, the authors estimate the emissions from different source types. In addition, the difference between weekday versus weekend emissions and the diurnal cycle are resolved. The authors find reasonably good agreement of the emissions with the prior estimate, but that open burning emissions are likely significantly underestimated in their prior. I recommend this manuscript for publication after the following comments have been addressed.

Section 2.4, which describes the inverse modelling method, is difficult to follow. In particular, there appear to be a number of inconsistencies in the definitions of the variables in Eq. 1 and the physical units of these. Details are given in the specific comments below. I suggest that the authors review this section carefully to make it clear to the reader exactly what was done.
Specific comments
P12021, L22: Specify whether this is emissions of EC, OC or both.
Text adjusted: "Emissions of EC and OC in the St. Louis region"

P12023, L27: What is meant by "smoking vehicles". Is this synonymous with "vehicle emissions"?
We mean "High-emitting smoker vehicles," text replaced.
For more information, please refer to the papers cited in Bae et al., 2006.

P12024, L10: This sentence appears to be incomplete, for instance, which "Potential Source contribution function" or were there more than one. Please also add a reference.
Text clarified: "Lee et al., 2006, used the Potential Source Contribution Function method based on back-trajectories to show that sulfate levels at the site were impacted by the Ohio River Valley, while nitrate levels were associated with transport from the west and northwest."

P12024, L15: Do the authors mean at the St. Louis Midwest Supersite, if so, this should be specified.
Text changed, thank you: "measured at the St. Louis Midwest Supersite."

P12024, L18-20: Please rephrase this sentence to make it clearer that EC is a passive tracer whereas OC is produced also in the atmosphere. The way it is written it is not clear where OC is "created" and it implies that EC is not emitted, which of course is not the case.
Text clarified: "As discussed above, EC is not formed in the atmosphere but rather emissions are transported until they are removed by deposition such that they can be simulated as passive tracers. In contrast, OC is both emitted and produced in the atmosphere."

P12026, L7: Fig. 2 is referred to before Fig. 1, suggest either referring to Fig. 1 beforehand or reversing the order of the figures.
Thank you for pointing this out. Reference added to Fig. 1 in Sec 2.1: "Fig. 1 shows the location of the measurement site."

P12026, L9: Table 1 has not yet been referred to, suggest reversing the order of Table 1 and 2.
Thanks, Table #1 moved to #3.

P12026, L20: Suggest that the authors start this paragraph by mentioning that an alternative emission dataset was prepared, to compare with the LADCO one, from the NEI data. Otherwise it is difficult to follow the text.
Thanks for the suggestion, text changed: "In order to have an additional comparison to the LADCO prior emissions and the inverse model results, the 2008 National Emissions Inventory (NEI) version 3 was obtained from the US Environmental Protection Agency."

P12028, L20: Is dry/wet deposition accounted for in the FLEXPART simulations? If yes, please indicate the scheme for this. If no, please comment on how this may impact your residence time analysis.
RTA considers only air mass movements which are identified from passive tracers without deposition, as discussed in Ashbaugh et al., 1985.
Text added: "The particles were treated as passive tracers with neither wet nor dry deposition."

P12029, L12: What is the impact of not using the aerosol module in CAMx, for instance, does this mean that dry/wet deposition of aerosols is not accounted for. If so, how will this impact your simulations?
We used wet and dry deposition in CAMx, and apologize for failing to mention this in the text.
Text added: "Dry deposition was calculated using the Zhang et al., 2003 scheme, and wet deposition using the standard scheme in CAMx."

Text added to clarify the limitations of our study:
"Both EC and OC are therefore simulated as passive tracers with wet and dry deposition. This is adequate for EC, and so the inverse model results can be straightforwardly compared to the emissions inventories. In contrast to EC, there is extensive formation of OC in the atmosphere which is not simulated in our model. This means that the inversion will not distinguish between primary and secondary OC, and that results are therefore best interpreted as impacts at the measurement site rather than as emissions at the source location. It also means that we are not able to evaluate the non-linear interactions of different plumes together."

P12029, L25: Please specify the "two-step" method. In Rigby et al. and Rödenbeck

et al., an Eulerian model is used to take into account the influence on the air masses which is not accounted for in the time frame of the back trajectories (in this study 4 days). It is not clear in this study, however, how the background influence is accounted for or how the Eulerian model simulations would provide the background influence information. Also, please note that Rigby et al. 2011 actually use a 1-step method.

Sorry for the mistake. It is not our intention to describe alternative methods at this point but rather to make sure that they are cited.

EC and OC levels are very low in St. Louis and the background can be assumed to be close to 0, see Fig. 4. We therefore do not worry about the influence of areas beyond those described in Sec. 2 - these are negligible, especially when you take deposition into account. Both Rigby and Roedenbeck were looking at global emissions of inert gases.

The paragraph was expanded as follows:

"Inverse models based on back-trajectories alone include Stohl et al., 2009, Brioude et al., 2011 and Brioude et al., 2013. This work combines back-trajectories with Eulerian simulations, and in this respect is similar to the methods presented in Rigby et al., 2011 and Roedenbeck et al., 2009. The purpose of combining the Lagrangian and Eulerian simulations for Rigby et al., 2011 and Roedenbeck et al., 2009 was to combine global transport of inert species with higher definition impacts from specific locations. In our case, the background levels of EC and OC are very low (see Fig. 4), and we expect minimal impacts from sources outside the study area. The purpose of combining Eulerian with Lagrangian simulations is therefore to estimate adjustments to known emission inventories with the Eulerian simulations, and to estimate impacts from unknown area sources in an overlapping domain with the Lagrangian simulations."

P12030, L6-8: If I have understood correctly, the simulations are not made using actual meteorology of each hour/day. Are these then average hours and days for the given year (i.e. 2002) or other? Please specify. Also, please specify if this was using CAMx or FLEXPART.

We apologize for the lack of clarity in our explanation. In fact, we perform hourly CAMx simulations for each source group for each time chunk. Each of the 600 input times series is a year-long hourly timeseries. We hope that the following paragraph is clearer:

"Hourly Eulerian simulations with CAMx were performed for the five different source groups in the LADCO inventory: On-Road, Non-Road, MAR, Other and Point Sources. Because we are interested in evaluating the temporal profiles of the sources, we carry out separate simulations for emissions during different times of the day and different days of the week. The time slots were selected based on the diurnal profile used in the emissions inventory: 11:00 p.m. to 05:00 a.m., 05:00 a.m. to 08:00 a.m., 08:00 a.m. to 02:00 p.m., 02:00 p.m. to 06:00 p.m., and 06:00 p.m. to 11:00 p.m. Days of the week were split into a weekday group and a group containing Saturdays, Sundays and Holidays. As an example, an hourly time series of concentrations was obtained from a CAMx simulation with On-Road emissions only between 05:00 a.m. to 08:00 a.m. on weekdays. With 5 source groups, 5 time slots and 2 day types, this means that there were 50 CAMx simulations. We are also interested in the annual profile of the emissions, and so we divide the 50 resulting concentration time series into 12 months for a total of 600 input time series into the inverse model."

We also clarified that the open burning time series are obtained with CAMx:

"The open burning emissions are included in the inversion as 6 time series simulated by CAMx for the entire year for the 6 geographic sectors shown in Fig. 3. We also include a CAMx time series representing impacts from biogenic emissions, as discussed in Sec. 3.2."

P12030, L19-22: Please specify that these were the averaged timeseries since they

are the averages of different the weekdays/weekends and 4 timeslots throughout 2002.
Hopefully the new paragraph clarifies this point. There are no averages, just lots of year-long hourly time series.

P12031, L1-2: This sentence does not follow from the preceding one. What does
"these" refer to in this sentence, it is not clear.
Thank you for pointing out the ambiguities in the explanation, we have changed the paragraph as follows:
"The inverse model derives a posterior estimate of emissions based on the Eulerian simulations that used the emissions inventory as a prior. In addition, the inverse model uses the Lagrangian simulations to derive an estimate of sources that may be missing from the inventory. This is done by using the polar grids of Residence Time Analysis that represent the impact that an emission in a given grid cell would have at the measurement site. As all the known sources were already included in the CAMx simulations with the emissions inventory, we use a field of zero prior emissions for the polar grids from the Lagrangian simulations."

P12031, L10: the elements of x' are not "factors" as there is no multiplication involved.
Please use rather "vector of emission corrections" or other.
Changed, thank you.

P12031, L11: From the preceding paragraph I understood that you use the emissions
estimates themselves and that these are not parameterized. Please clarify, are the
elements of x the prior emissions or prior parameters?
The term parameter was meant in a very general way. A large range of different metrics of emissions can be used in x, so long as they yield concentrations when they are multiplied by the corresponding column in H. In this paragraph, we would prefer to keep the description general. In the following paragraph, we explain what types of parameters are actually used in the study (bearing in mind that in a broad usage, a parameter could be an emission in tonne/year, or a scaling factor of some kind).
Sentence added:
"The individual entries in x can take different forms: they can be actual emissions in units of mass per time, or they can be non-dimensional scaling factors."

P12031, L11 and L20: These sentences are inconsistent. In L11 the authors state that
H is the operator to convert emissions to concentrations, while in L20 the authors state
that the columns of H contain the timeseries' from the CAMx and FLEXPART simulations
(in units of concentration for CAMx and emission sensitivity for FLEXPART).
We hope that having clarified that x can have different types of elements the inconsistency has been resolved. Please see changes to the text in the previous comment. What matters is that H times x yields concentration values that are consistent with y.

P12031, L20-25: Having read this section, I understand that the authors have made
606 + 2880 simulations hourly for all of 2002. Is this correct? I suggest that they
re-write the section P12030, L17-27 to make this clearer as otherwise, the reader will
think that these are averaged timeseries, in which case, the description of H does not
follow.
Yes, your understanding is correct. Hopefully your suggestions for (P12030, L6-8) helped to clarify this up front. We have added the following to clarify things as you suggest, thank you:
" all of which are hourly time series for the whole of 2002."

P12031, L27: There is an inconsistency here with L10-11 of the same page. In L10-11 the authors state that x0 are the prior emissions and x are emissions. However, in L27 the authors state that x are scaling factors such that posterior emissions are equal to x*x0. Furthermore, given the definition of H, Hx (for the 2880 emission sensitivities) cannot be equivalent to y if they are scaling factors. Lastly, again given the definition of H, the elements of x cannot all have the same units. Please clarify.

The sentence was replaced with the following one, which is in line with the changes made in response to the comments above:

"For the CAMx time series, the entries in x are scaling factors on the LADCO emissions that went into the CAMx simulations. For the FLEXPART polar grids, the entries in x represent emissions."

P12031, L25-26: the vector x includes entries for the "gridded area sources" as well as the "group sources". Please be clearer about which sources are gridded and which are grouped as to be sure that the source types are not being double counted.

The polar grids cover the same area but have a zero prior. They are meant to provide a way for the inverse model to identify areas with emissions that are not represented in the LADCO prior.

The following text was modified to clarify this point:

"As all the known sources were already included in the CAMx simulations with the emissions inventory, we use a field of zero prior emissions for the polar grids from the Lagrangian simulations."

P12037, L12-13: I do not understand this sentence. What is meant by "yielded the most consistent estimate of impacts in the inversion"?

We tested CAMx simulations using different biogenic species for different model configurations, and we selected the one that gave the best improvement to the overall match of the inverse concentration time series with the measurement time series. This was the CG5 category.

The sentence was changed as follows:

"For OC, we tested different biogenic components and found that condensable gases category 5 "CG5" yielded the best inverse time series of OC compared to the measured time series."

P12037, L19: Is the 19% for the "Other" category refer to the posterior emissions. This should be stated and possibly also mentioned at L4-5.

We have added an explanatory sentence at the top of the section to help clarify this:

"By impacts, we mean the surface concentration of EC or OC at the measurement site that are due to transport of particular emissions to the site."

In L4-5, we clarify that we are talking about contributions to the simulated impacts by either the prior or the posterior: "simulated" added in three places in the paragraph.

In L19, likewise, we mean that 19% of the posterior simulated concentrations are due to emissions from the "Other" category.

P12037, L23-29: Bootstrapping will provide an estimate of the uncertainty that comes from sub-sampling the data. However, there is also data selection in that outliers of more than 3 SD are removed. Have the authors investigated the sensitivity of the results to outliers and the selection criteria?

We have expanded the description of the bootstrapping and included a new figure to show the uncertainties in the results, please see the response to Reviewer #3, General Comment #6.

We have performed various tests with different selection criteria and found that while the least squares method is sensitive to outliers, the results of our inversion are robust relative to the different ways of

handling those outliers. Since the method we employ is widely accepted and basic "textbook" material, we feel that it is not necessary to further justify it here.

Section 3.3: Suggest adding subheadings to this section to make it clear what type of emission is being discussed e.g. On-road emissions, Non-road emissions, etc.
Done, thank you for the suggestion.

P12044, L12: What is meant by "annualized" does this simply mean the emission for each period given as the emission per year? Please clarify.
This means that we take the emissions for 4 months and multiply by 3 to obtain an emission rate for a whole year that would be equivalent to having the same emissions as the 4 month chunk for the rest of the year.
The following has been added:
"The emissions rates are annualized by multiplying the emissions in tonnes per 4 months by 3 in order to have emissions in tonnes per year. This yields the annual emission rate that would be obtained if the emissions of the 4 months continue for an entire year."

Technical comments
P12021, L2: "a year-long" (since it is only one year). Changed "based on one year of hourly measurements" (There were 2 time series, one for EC and one for OC).
P12022, L20: replace "under-prediction" with "underestimation" as it is something can only be predicted or not and not "under" or "over" predicted. Changed.
P12025, L6: "mixed-use neighbourhood". Changed.
P12025, L17-18: Please use SI units, i.e. metric units throughout. Changed.
P12029, L20: Please correct: Rödenbeck et al. 2009. Changed.
P12040, L2: "during daylight hours" (remove "the"). Changed.
P12044, L16: missing full-stop after "December". Changed.