Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-159-AC2, 2018

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Weak-constraint inverse modeling using HYSPLIT Lagrangian dispersion model and Cross Appalachian Tracer Experiment (CAPTEX) observations – Effect of including model uncertainties on source term estimation" *by* Tianfeng Chai et al.

Tianfeng Chai et al.

tianfeng.chai@noaa.gov

Received and published: 4 October 2018

General comments:
The manuscript addresses an inverse modelling study using CAPTEX data in which the effect of including model uncertainties is analyzed in the frame of source term estimation. The manuscript is well written, however, the potential contribution of the study is not clear. The study is focused on highlighting two major points: (i) advantage in us-

Printer-friendly version

Discussion paper

ing differences of logarithm of concentrations in source estimation and (ii) improvement in source estimation using a hypothetical form of observation and model uncertainty. This is not new and already established in the literature of parametric estimation problems and in the solution of inverse problems. Based on result and discussion, the study seems another application of source term estimation with sensitivity to their hypothetical model uncertainty formulation but there is no development in view of model, methodology or estimation.

We thank the referee for thoroughly reading the manuscript and providing valuable comments. For the two major points pointed out by the referee, we try to emphasize the second point, "the effect of including model uncertainties on source term estimation", as explicitly stated in the subtitle. Although it is not new to consider model uncertainty in the inverse problems, the model uncertainties in the literature of parametric estimation problems are mostly given as static terms and they will not vary with model source terms. n source estimates This has been clarified in abstract.

In abstract, "Before introducing model uncertainty terms" has been changed to "Before introducing model uncertainty terms that depend on source estimates".

Another uniqueness of this source term estimation experiment is that the exact source terms in CAPTEX field experiment are known. So, the source term estimation methodology can be thoroughly evaluated, including "the effect of including model uncertainties on source term estimation" emphasized in this study.

Major comments:

1. *A major question is regarding the hypothetical form of the error formulation?. It is not clarified why this particular form is chosen? Also, what is the evidence or guarantee that the same formulation with observed coefficients would work in*

other or similar source term estimation problems?

The hypothetical error formulation is used mainly for its simplicity. It is clarified when the formulation is first introduced (see below). The same formulation may or may not work in other or similar source term estimation problems, but we believe that does not affect the demonstration of "the effect of including model uncertainties on source term estimation".

The following text,

Firstly, the observational uncertainties are formulated to include a fractional component $f^o \times c^o$ and an additive part a^o. No model uncertainties are considered to contribute to ϵ.

has been replaced with,

Firstly, no model uncertainties are considered to contribute to ϵ. The observational uncertainties are formulated to include a fractional component $f^o \times c^o$ and an additive part a^o. Note that this general formulation is chosen for its simplicity. It should be replaced when more uncertainty information is available.

2. *The Authors did not explain well if their inverse problem is over-determined or under-determined. By noting their discretized grid and number of measurements, it seems an over-determined problem, If so, why do you need a smoothing constraint?*

   The smoothing constraint was not needed for the current over-determined problem. It was included to make the formulation more general. In the revised version, the smoothing term was removed from both Equations (1) and (5).

   The following statement has been added at the end of Section 2.3 to clarify that all problems are over-determined.

   "Note that the cases presented in this study are all formulated as over-determined problems".

3. *Authors did not explain why the using concentration difference and logarithmic concentration difference results so differently for the estimation of release rate. How could be the difference is so drastic between table 2 and table 3.*

   The drastic differences between Table 2 and Table 3 are caused by the distinct bias directions when using concentration and logarithmic concentration in comparing model predictions and observations before introducing model uncertainty terms.

   Text in the third paragraph of Section 3.1 shown below explains the cause of the significant underestimation when using concentration differences.

   The significant underestimation of the release strength is caused by the implicit assumption of a perfect model when ϵ does not include the model uncertainties. Figure 2 shows the comparison between the predicted and measured concentrations when the actual release rate of 67 kg/hr is applied. Large discrepancies still exist even when the exact release is known and used in the simulation. For the measured zero concentrations, most of the predicted values are non-zero and can be above 1000 pg/m^3. As $\epsilon_m = a^o$ for these zero concentrations, $\frac{(c_m^h - c_m^o)^2}{\epsilon_m^2}$ will dominate the cost function when a^o is not large enough. This explains that the underestimation is not as severe for a^o =50 pg/m^3 as that for a^o =10 pg/m^3. While ϵ do not change with f^o for the zero concentrations, smaller f^o values help increase the weighting of the terms $\frac{(c_m^h - c_m^o)^2}{\epsilon_m^2}$ associated with large measured concentrations. So, the estimated emission strength when $f^o = 10\%$ is better than when $f^o = 50\%$.

   The text in the same section after Equation 2 shown below explained the overestimation when using logarithmic concentration difference. It is probably not very clear and it has been revised.

   "The $\frac{a^o}{c_m^o}$ term in Equation 2 makes $\epsilon_m^{ln(c)}$ larger for measured low concentrations than those measured high concentrations, thus makes the measured zero con-

centrations have little effect in the final emission strength estimates".

In the current version, it is replaced with the following.

"The $\frac{a^o}{c_m^o}$ term in Equation 2 makes $\epsilon_m^{ln(c)}$ larger for measured low concentrations than those measured high concentrations. It causes more weighting towards measured high concentrations and results in overestimation shown in Table 3. The measured zero concentrations have little effect in the final emission strength estimates".

4. *It is not clear how cost function normalization can avoid spurious solutions in logarithm concentration difference? Does this spuriousness appear while using only concentration differences?*

   We only see the spuriousness problems when using logarithmic concentration differences. How cost function normalization can avoid spurious solutions is illustrated in Figure 3. As explained in text, the "smaller cost function when release strength approaches zero is due to the increasing $(\epsilon_m^{ln(c)})^2$ in Equation 4 as c_m^h gets smaller". The normalization in Equation 5 makes the sum of the mismatch weighting terms constant. To make it clear, we moved Equation 5 and the rewritten preceding text shown below to the beginning of the section before introducing Figure 3.

   To avoid having zero source as a global minimizer in such situations, the total weighted mismatch between model simulation and observations is normalized by the total weights when $q_{ij} = q_{ij}^b$, as shown in Equation 5.

   This has been replaced with the following.

   To avoid having zero source as a global minimizer in such situations, the sum of the weights of the mismatch between model simulation and observations is kept unchanged for varying q_{ij} by normalizing it with the weight sum when $q_{ij} = q_{ij}^b$, as shown in Equation 5.

5. *What is the utility of introducing the third part of the equation 1 if coefficient c_sm is always put to zero?.*

   This third part has been removed from both Equations 1 and 5.

6. *The coefficients a0, f0 or ah, fh are chosen arbitrary, there is no justification why a particular set has been chosen ?*

   The observation uncertainty parameters $f^o = 20\%$, a^o =20 pg/m^3 are chosen after introducing the model uncertainty. Other f^o and a^o values have been tested and they show quite similar results.

   This has been clarified in Section 3.2 by adding the following in the third paragraph.

   Additional tests with other chosen f^o and a^o values show similar but slightly different results. For brevity, they are not presented here.

   A specific combination of f^h and a^h are not chosen until Table 12, which chooses $f^h = 20\%$, a^h =20 pg/m^3. Tables 7, 9, and 11 have shown the results are not sensitive to the f^h and a^h choices. To clarify this, the following sentence has been added to the second paragraph of Section 3.5.

   When logrithm concentration is taken as the metric variable, the emission estimates are not sensitive to f^h and a^h choices, as indicated by the results in Tables 7, 9, and 11,

7. *In Figure 2, the scale of obervations and model concentration is not clear? Is it correct? What are the range of observed and modelled concentrations?*

   The largest observed and modelled concentrations are 24600 pg/m^3 and 25661 pg/m^3, respectively. For release 2, there are four observed 3-hr average concentrations above 10000 pg/m^3. To allow logarithm calculation, a constant 1 pg/m^3 is added to both observed and modelled concentrations. So zero observed and modelled concentrations appear as 1 pg/m^3 in the figure.

8. *Adding the model uncertainties to epsilon_m are simply an increase of magnitude in the previously chosen quadratic function based on observed concentration. Did you try by increasing only the values of f0 or a0 to analyze the same kind of effect?*

   The constant part of the model uncertainties, a^h, has the same kind of effect as its counterpart of the observational uncertainties, a^o. This has been pointed out in section 3.2 when presenting results in Table 4, as shown below.

   When $f^h = 0\%$, $a^h = 10$, 20, and 50 pg/m^3 while a^o=20 pg/m^3, the q estimates, 7.7, 9.1, and 13.6 kg/hr, are inline with the results shown in Table 2, where $q = 7.1$ kg/hr for a^o=20 pg/m^3 and $q = 12.6$ kg/hr for a^o=50 pg/m^3.

   However, f^h is applied to the predicted concentrations which vary with different source term estimation. Such "dynamic" effect of the "model uncertainty terms that will depend on source estimates" cannot be replicated with f^o which applies to the "static" measurements. In fact, this is the new aspect we want to emphasize in this paper. As stated earlier, the abstract has been changed to make it clear.

9. *Page 9, line 30, Another aspect ... as the metric variable. What does it mean that range of release estimates are not as large as those using concentration variable ?*

   It refers to the results in Tables 4 and 5. This has been clarified with the rewritten sentence as below.

   Another aspect of using logarithm concentration as the metric variable is that the range of the release estimates listed in Table 5 are not as large as those in Table 4 resulted from using concentration as the metric variable for the same 12 combinations of a^h and f^h.

10. *What about uncertainties in the source parameters due to varying nature of model or observation uncertainties. Is it possible to compute it with given procedure?*

The expected error $\epsilon_{q'}$ of the estimated release rate when assuming the actual release location is known has been calculated for each release. They are listed besides q' as the last column in Table 12. The following text has been added to the fourth paragraph in Section 3.5.

The posterior uncertainties of the release rate estimates $\epsilon_{q'}$ are also calculated and listed. They range from 1.8 kg/hr for release 2 to 6.2 kg/hr for release 1. The apparent underestimation is likely due to the model uncertainty assumption, including its simplified formulation as well as the chosen parameter values.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-159, 2018.

Printer-friendly version

Discussion paper