Estimation of power plant SO_2 emissions using HYSPLIT dispersion model and airborne observations with plume rise ensemble runs

Tianfeng Chai, Xinrong Ren, Fong Ngan, Mark Cohen, and Alice Crawford

Response to the comments of Reviewer 1

July 31, 2023

The paper describes attempts to estimate SO2 emissions from power plants by use of a Lagrangian dispersion model and aircraft measurements. It emphasizes the uncertainty in plume rise due to stack heat input, which is treated as unknown. Two methods are used to find the optimal heat input. There is some good information here, but the presentation could be clearer, and the implications should be more clearly stated.

We thank the referee for thoroughly reading the manuscript and providing valuable comments. The manuscript has been revised for a better presentation of the objective, findings, and the implications.

Point-by-point responses to the referee's specific comments are given below.

General comments:

1) *The objective of the paper seems to be to find ways to determine the optimum simulation to produce the correct (known) emissions. Two methods are suggested, one based on correlation between the observed and simulated time series, and the other based on the RMS difference of that same time series. Unfortunately I have just explained the objective more clearly than the paper ever does. These are reasonable proposals for how to determine the optimum simulation, but they both have flaws, which are evident in the data. For example, both the correlation and the RMS are sensitive to misplacement of the plume, whereas the inversion may not be sensitive to that misplacement.*

   Thanks for pointing this shortcoming in the original manuscript. We have revised the paper to make the presentation clear. In particular, the abstract has been rewritten to better explain the objective of the study. Some of the details from the abstract has been removed to emphasize the main points of the paper as explained by the reviewer here.

2) *The heat input to the plume rise calculation is treated as a free parameter. There must be reasonable estimates of the real value available, based on the CEMS data and the characteristics of the plants, for example whether they have scrubbers or not. If the optimization process finds values that are well outside a reasonable range, that may indicate that the plume rise calculation is inadequate, which would be valuable information.*

   The following paragraph has been added to the Summary and discussion section.

   While the stack exit gas temperature data are not available for this study, a single constant stack exit temperature is provided for each facility in the 2020 National Emissions Inventory (NEI) (Personal communication with George Pouliot at the U.S. EPA). Using the average measured air temperature as the ambient temperature and the other CEMS data (United States Envi- ronmental Protection Agency (U.S. EPA), 2022), including hourly exit air flow rates, the morning/afternoon heat emissions are estimated as 52–59 MW/49–56 MW, 80–92 MW/76–87 MW, and 13–13 MW/12–13 MW, for Roxboro, Belews Creek, and CPI Roxboro, respectively. Note that the heat emission estimation is sensitive to the stack exit temperature which is expected to vary from hour to hour, similar to the exit air flow rates and the SO_2 emissions. Nonetheless, these estimated values indicate the reasonable ranges of the heat emission. When correlation-based and RMSE-based methods agree with each other in their "optimal" heat emission for Roxboro and Belews Creek morning segments, the "optimal" heat emissions are very close to the estimated stack heat emissions here. When the two methods disagree, the correlation-based "optimal" heat emissions of 90 MW/140 MW for CPI Roxboro morning/afternoon are unreasonably high, but the RMSE-based "optimal" emissions of 30 MW/50 MW could still be reasonable. This suggests that the RMSE-based results are probably more reliable.

3) *Only two of the many possible sources of uncertainty are explored here. That's fine if it is clearly stated. The two sources examined are the plume rise and the background specification. Errors in wind direction are present, and get some attention. Errors in vertical placement and mixing of the plume are probably also present, but are not discussed at all. As it stands, varying the heat input amounts to looking for the value that best compensates other errors in the model. That's not wrong as an empirical method, but again, it should be stated.*

   In the previous version, the uncertainty issues pointed out here were mentioned in the Summary and discussion section, but were not very clear. The statements have been modified and moved into a separate paragraph, as shown below.

While the uncertainty of the heat emission is focused here, there are a lot of other uncertainties associated with the emission estimates. For instance, uncertainties in many parameters, such as the assumed background SO_2 mixing ratios, the meteorological data input such as the wind direction and speed, and some of the HYSPLIT turbulence parameterizations related to the turbulent mixing, will all affect the final results. Even if the hourly exit temperature were available, the plume rise calculated using the Briggs algorithm may still misplace the plume. It is likely that the "optimal" heat emissions chosen here have compensated other errors in the model.

4) *The vertical structure of the simulated plumes should get more emphasis. Some of the figures in the Appendix should be promoted to the main text. It looks like the flights were rather close to the plants, that is, in the region where the plume is not well-mixed in the vertical. This is arguably a mistake in the flight planning, unless it is an error in the model (too slow mixing). In theory, the inversion should recover the correct emissions as long as the observation samples a reasonable amount of the plume, but this is a very strong constraint on the precision of the simulation.*

   We have moved the previous Figure A7 to the main text (as Figure 10 in the revision). While it is presumably true that the inversion should be able to estimate the emissions as long as the observation samples a reasonable amount of the plume, it is difficult for the inverse modeling to accurately estimate the emissions due to many other uncertainties in the model simulation, as pointed out by the reviewer. The statement, "The results here indicate the need to have more observations at different altitudes in the future flight planning", has been added after presenting all the results.

5) *It seems that we are to take the set of differences between the inverted and known emissions as a measure of the uncertainty of the method, but this is never stated. Although a formal uncertainty analysis is not really possible with such a small number of samples, some statement should be made. Clearly the differences are not Gaussian, and the large differences (which may or may not be "outliers"), are of concern.*

   We have added the following statements when presenting the results in Figure 10 (now Figure 11 in the revision).

   Note that the ranges of the inverted emissions with 10 MW above and below the "optimal" heat emissions are used as to indicate the sensitivities of the results to the heat emissions. While the differences between the emission estimates and the known CEMS data provide some confidence to the results, quantification of the uncertainties associated with the method probably requires further investigation in the future.

6) *More detail is needed on the WRF runs. WRF has many options. The chosen options, initial and boundary data, etc. must be stated in enough detail that WRF experts can judge whether they are reasonable, and others can plausibly replicate the results. In particular, whether using the mixed layer depth (PBLH?) out of WRF directly is reasonable depends on the physics options chosen.*

The following has been added to address the missing detail.

The WRF model was configured for three-nested domains with horizontal grid spacing of 27 km (D01), 9 km (D02), and 3 km (Figure 2). A total of 33 vertical layers were defined with a higher resolution near the surface and 100 hPa for model 110 top. There were 20 layers below 850 hPa with the first mid-layer height of the model at around 8 m. The simulations for the D01 were initialized by using the North American Regional Reanalysis (Mesinger et al., 2006) with 32-km grid spacing and available every 3 h. Then, the WRF results from the coarser domains provided the initial and boundary conditions for the inner domains. The daily WRF runs had a 30-hr duration including 6-hr a spin-up period (i.e., starting at 18 UTC on the previous day). The physics options for the WRF simulations were - the rapid radiative transfer model for radiation parameterization 115 (Iacono et al., 2008), WSM6 for microphysics i(Lim and Hong, 2010), the Grell 3D Ensemble for the sub-grid cloud scheme (Grell and Devenyi, 2002), Noah land-surface model (Chen and Dudhia, 2001), and Mellor-Yamada-Nakanishi-Niino 2.5 level TKE scheme for the planetary boundary layer (PBL) parameterization and its corresponding surface layer scheme (Nakanishi and Niino, 2006).

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E., Ek, M., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American regional reanalysis, Bull. Amer. Meteorol. Soc., 87, 343–360, https://doi.org/10.1175/BAMS-87-3-343, 2006.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, J. of Geophys. Res., 113, https://doi.org/10.1029/2008JD009944, 2008.

Lim, K.-S. S. and Hong, S.-Y.: Development of an Effective Double-Moment Cloud Microphysics Scheme with Prog- nostic Cloud Condensation Nuclei (CCN) for Weather and Climate Models, Monthly Weather Review, 138, 1587–1612, https://doi.org/10.1175/2009MWR2 2010.

Chen, F. and Dudhia, J.: Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity, Monthly Weather Review, 129, 569–585, 2001.

Grell, G. and Devenyi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophys. Res. Lett., 29, https://doi.org/10.1029/2002GL015311, 2002.

Nakanishi, M. and Niino, H.: An improved Mellor-Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog, Bound.-Layer Meteorol., 119, 397–407, https://doi.org/10.1007/s10546-005-9030-8, 2006.

7) *The SO2 background is clearly important in this region. I recommend simplifying the presentation by removing the parts where background is not used. Furthermore, I am concerned that the method used to derive the background may be chosen primarily because it gives the best results (given compensating errors). The explanation is not perfectly clear, but the 25th percentile within the plume seems like it should yield a value considerably higher than background, which is usually taken to be outside the plume.*

   We like to keep the results where background is not used in order to emphasize the importance of the including the background in the analyses. Since the choice of the background value is not trivial, we presented several options to estimate the background. The choice to take the segment-specific minimum as the background is actually very close to take the measurement outside the plume. In this study, the 25th percentile choice yields the best inversion results as shown in Table 5. It is possible that is partially due to some compensating errors. We do not imply that the 25th percentile choice would be universally suitable to other cases.

8) *The conclusions should state the authors' recommendations for future studies. This should include a recommendation for which optimization method to use, or that another method is needed. Guidelines for deciding whether a given set of observations is useful or should be discarded would be helpful. Does a large RMS relative to the mean imply that a flight should be discarded? Implications for flight planning should be included. Do the authors recommend using single deterministic meteorology, or should ensembles be used?*

   We have revised the conclusion. In particular, the following two paragraphs have been added. The first one gives more evidence to show that the RMSE-based results are more reliable. The last paragraph implemented the reviewer's recommendation on the needed content. The third from last paragraph in the revision (not copied here) partially addressed "the guidelines for Guidelines for deciding whether a given set of observations is useful or should be discarded" and the question of "a large RMS relative to the mean imply that a flight should be discarded". We have added the statement, "However, special care is needed for such situations where large RMSEs also indicate the model deficiencies", at the end of the paragraph.

While the stack exit gas temperature data are not available for this study, a single constant stack exit temperature is provided for each facility in the 2020 National Emissions Inventory (NEI) (Personal communication with George Pouliot at the U.S. EPA). Using the average measured air temperature as the ambient temperature and the other CEMS data (United States Environmental Protection Agency (U.S. EPA), 2022), including hourly exit air flow rates, the morning/afternoon heat emissions are estimated as 52–59 MW/49–56 MW, 80–92 MW/76–87 MW, and 13–13 MW/12–13 MW, for Roxboro, Belews Creek, and CPI Roxboro, respectively. Note that the heat emission estimation is sensitive to the stack exit temperature which is expected to vary from hour to hour, similar to the exit air flow rates, and the SO_2 emissions. Nonetheless, these estimated values indicate the reasonable ranges of the heat emission. When correlation-based and RMSE-based methods agree with each other in their "optimal" heat emission for Roxboro and Belews Creek morning segments, the "optimal" heat emissions are very close to the estimated stack heat emissions here. When the two methods disagree, the correlation-based "optimal" heat emissions of 90 MW/140 MW for CPI Roxboro morning/afternoon are unreasonably high, but the RMSE-based "optimal" emissions of 30 MW/50 MW could still be reasonable. This suggests that the RMSE-based results are probably more reliable.

This study shows that RMSE is a better metric than correlation coefficient in choosing the best ensemble member for the SO_2 emission inversion. While the RMSE-based "optimal" plume rise runs appear to agree better with the observations than the correlation-based "optimal" runs, observations are often missing when and where the "optimal" runs are significantly different. Additional measurements at mulitiple altitudes would have been really helpful. In the future flight planning of similar top-down emission estimation studies more vertical profiles of the target pollutant should be measured. In addition, more upwind measurements are also recommended in order to better quantify the background concentrations caused by many other emission sources. It is also wise to choose relative steady meteorological conditions for the flight campaign since the unsteady conditions such as frequent wind direction changes pose great challenges not only to the inverse modeling but also to the meteorological simulation and the dispersion modeling. The current study shows the value of the ensemble simulations when certain model parameters are difficult to determine, such as stack heat emission here.

Specific comments:

1) *The abstract is long and detailed, but does not clearly state the objectives or method. It is more of an introduction than an abstract.*

   The abstract has been modified. The objective of the study is now explicitly stated at the beginning of the abstract. Some details are also removed to better reflect on the findings of the study.

2) *Line 273: The standard deviation of the 1-s observations is not a reasonable estimate of the observational uncertainty. It does not take into account sampling error (probably dominant here). How is the process affected by a larger observation uncertainty estimate?*

   In our previous study (Chai et al. 2018), it was found that the emission estimates were not very sensitive to the observation uncertainty estimates (Table 3 in Chai et al. 2018). This has been clarified in the revision.

   Chai, T., Stein, A., and Ngan, F.: Weak-constraint inverse modeling using HYSPLIT-4 Lagrangian dispersion model and Cross-Appalachian Tracer Experiment (CAPTEX) observations – effect of including model uncertainties on source term estimation, Geosci. Model Dev., 11, 5135–5148, https://doi.org/10.5194/gmd-11-5135-2018, 2018