Atmospheric
Chemistry
and Physics
Discussions

Interactive comment on "Bayesian inverse modeling of the atmospheric transport and emissions of a controlled tracer release from a nuclear power plant" *by* Donald D. Lucas et al.

Donald D. Lucas et al.

ddlucas@alum.mit.edu

General response to Anonymous Referee #2

Anonymous referee #2 provided valuable feedback and comments about our ensemble analysis and inversion. The referee had six major comments, four minor comments, and recommended citing other studies using boosted regression trees (e.g. Sayegh et al). We greatly appreciate the reviewer's thoughtful comments, especially the suggestions to include additional information and analysis about the WRF variability using known source parameters. We address each of these comments below and

have revised the manuscript accordingly. The revised manuscript includes descriptions and analysis for three new figures and modifications to an existing figure and table. We believe that these revisions are sufficient to address the reviewer's comments for publication in ACP.

Replies to Major Comments

1. *Analysis using actual release parameters*
   The reviewer recommended additional analysis using the known source parameters to highlight the variability due to WRF alone. The submitted version did not include this analysis because we wanted to show that the source inversion algorithm works well without knowledge of the source. That is, we excluded known source simulations from the training data as a rigorous test of our method. However the reviewer's recommendation is important and valuable, so we have expanded Sect. 5.1 in the revised manuscript to describe the known source analysis. This section includes two new figures showing the variability due only to WRF when the source parameters are known. One of the new figures, which is similar to Fig. 7 in the submitted version, shows the ensemble spread caused by the differences among the 162 WRF runs in the time series of SF_6 at the four selected locations. The other new figure shows the mean squared error and correlation for the 162 WRF runs using the known source parameters. Fig. 9 in the submitted version has also been modified to show the WRF-only variations relative to the larger Latin hypercube ensemble variations. The results of the new analysis are consistent with the findings described in the original manuscript. The WRF-only variations are important, though not as dramatic as the full ensemble variations. Further, the categorical variables in WRF related to reanalysis data and initialization time explain most of the spread within the known source WRF-only ensemble.

2. *Table of actual and estimated parameters*
The reviewer requested a table that compares the source parameters estimated by the inversion algorithm to the actual values used for both the synthetic and real tracer releases. Table 3 in the submitted manuscript already shows the comparison for the real tracer release. This table is small and cross-referenced toward the end of the conclusions section, so it is not surprising that the reviewer missed it. In the revised manuscript, we expanded the table by adding the comparison for the synthetic tracer release and cross-reference it earlier in the discussion (in Sect. 5.5 and 5.6).

3. *Prior sampling ranges & insensitivity to release height*
The reviewer notes that some of the sampling ranges for the source parameters in the prior distribution, especially the release height, seem to be too narrow because the posterior histograms extend to the limits. This behavior is not necessarily undesirable because the source parameters can be estimated from likelihood maxima that occur within the limits. Moreover the widths of the posterior distributions provide information about the uncertainty in estimated source parameters. Wide posterior distributions can result from observations that are not constraining or a model that is insensitive to parameter variations.

For the release height, we used a narrow range to represent near-surface releases. We are not surprised that the simulations are not sensitive to the narrow variations because the reanalysis data sets and model simulations use a relatively coarse vertical resolution. We therefore expected to see similarities between the posterior and prior distributions for the release height (i.e., both are "flat"), which was confirmed by our results. It is often useful to include non-sensitive parameters in this type of Bayesian analysis to verify that the algorithm works properly.

As the reviewer suggests, it would be interesting to broaden the sampling range for the release height to also cover elevated releases. However, this would re-

quire running a large number of new ensemble simulations, which we are unable to perform at this time (note that narrowing the range would not require new simulations). We will revise the manuscript to better describe the relationship between the posterior widths and parameter sensitivities, with an emphasis on the release height. We will also mention our interest to add elevated releases in future work.

4. *WRF resolution*
The 300-meter resolution of the innermost WRF domain was selected to resolve some of the terrain features surrounding Diablo Canyon. For instance, a line of measurement sensors detected SF_6 in a small canyon stretching from location 413 to the northwest (see Fig. 3). The selected model resolution is a compromise between being able to simulate the transport of SF_6 through these features and having scale-dependent parameterizations that are still valid.

Not all of the sensor locations require 300-m resolution, however, so the reviewer's suggestion to incorporate WRF resolution into the analysis is very interesting and relevant. We could potentially speed up the inversion by initializing with coarser WRF resolutions and progressing to finer resolutions. We could also quantify the joint effects of resolution differences and parameter variations. Though we haven't conducted this analysis for WRF and FLEXPART, we have done something similar for a global climate model. In that work, we ran ensembles at different resolutions and included resolution as an input parameter in our machine learning models. We will include additional discussion about WRF resolution in the revised manuscript. We also welcome future collaborations with the reviewer and others on this interesting topic.

5. *Number of observations*
The Diablo Canyon field study was heavily instrumented because of the complexity of the surrounding terrain and due to close proximity to populated areas. We agree that this data would be valuable for optimizing the effects of measurement density and location. Although we have conducted optimal analysis of this type

for other studies and regions (e.g., see Lucas et al., 2015), we did not perform a similar optimization here. Instead, we roughly assessed the impacts of reduced data density by selecting fractions of measurement points at random, for fractions between 5% and 90%. We then re-computed the R^2 coefficients of determination for the gradient boosting regression models, as described in Sect. 5.4. As fewer measurement points are used in the fits, the quality of the GB models degrade, which, in turn, limits the ability to perform an inversion. From this analysis, using only 20% of the measurement points still results in R^2 values of about 0.82 and 0.96 for the *mse* and *corr*, respectively, which should be sufficient for estimating the most influential source parameters. Because some measurement locations are more important than others, we should be able to reduce the fraction of measurements below 20% without adversely affecting the inversion, if we optimize the locations. We will include a summary of this discussion in the revised manuscript and consider future observational optimization studies for Diablo Canyon.

6. *Optimized versus regular parameters*
   We agree with the reviewer that it is a good idea to include a new figure that compares good and bad model simulations to the measurements. We selected high and low likelihood cases from the 40,000 member Latin hypercube ensemble and compare their SF_6 concentrations to measurements. This figure clearly shows that the high likelihood case provides a much better match to the measurements than the low likelihood case. It also shows that simulations have a positive bias, which leads to the minimum *mse* of about 0.2 instead of 0. This figure will be included and described in Sect. 5.4. Along with the modified version of Fig. 9 described in item 1 above, these figures will show the improvements that can be gained by varying WRF parameters and source parameters.

Replies to Minor Comments

As suggested by the reviewer, the revised manuscript will provide

- more details in the abstract about the optimized WRF configuration and accuracy of the parameter estimates,

- an improved caption for Fig. 1,

- further discussion about feature scores and parameter sensitivity,

- and additional references and information about gradient boosting applications in atmospheric science.