Atmos. Chem. Phys. Discuss.,
doi:10.5194/acp-2016-355-RC1, 2016

Atmospheric Chemistry and Physics
Open Access
EGU
Discussions

Interactive comment on "Network design for quantifying urban CO_2 emissions: Assessing trade-offs between precision and network density" *by* Alexander J. Turner et al.

Anonymous Referee #1

Received and published: 27 June 2016

The paper addresses the relative effectiveness of different strategies for estimating urban CO2 emissions using in situ concentration measurements, focusing on the relationship between the number of measurement locations, the precision/accuracy of the measurements, and the uncertainty of the resulting flux estimate. As the authors note, this important relationship has not been fully characterized. The approach, fitting a statistical model to a sample of network designs, is novel but intuitive. While the detailed analysis could be more robust in some ways, the conclusions are a significant advance over the current state of knowledge and will have concrete value for the design of future observing systems, and I encourage publication once my concerns are address.

My primary concern the author's can address readily. As currently written, the abstract

could lead to false impressions that this analysis definitively concludes that moderate cost sensors in a denser network is the optimal configuration for any urban area and that weekly CO2 emissions with uncertainties of less than 5% can be achieved. These sections of the abstract in particular should be re-worked, as the authors actually are finding that with their specific modeling framework, higher density/moderate cost sensors provides an improved basis for flux estimation for the Bay Area. Further, given some of the assumption of diagonal error co-variance matrices, the representativeness of low altitude measurements in an urban region, and the gap presented by neglecting night-time data, the 5% monthly conclusion would appear to be an optimistic/idealized result and needs to be presented as such.

Additionally, error in describing the background condition, or the CO2 levels before impact of the urban region, have been found to be of high importance in other urban studies, and more discussion on the construction of this and assumptions used would be helpful. Otherwise my recommendations center around the clarity of presentation. In order to be of most use to a wider audience, including to researchers who may wish to perform similar analyses as they design networks in other cities, the methods need to be described more fully and precisely. Some justification should be provided for the choices and assumptions made in the analysis; I point out some examples below, but the authors should make a thorough review. The figures, especially figure 4, should be made more clear.

Detailed Comments How are the representativeness of observations made at just meters above the surface in a dense urban environment addressed? Depending on how these observation sites are setup, they could be biased in their sampling to see traffic, people, or biosphere in a courtyard. This paper does not need to solve this problem, but it should be discussed as a potential additional source of bias error in 'cheap' network deploymentsâĂŤparticularly as more sites are deployed (which can be very challenging to secure sites for deployment) and less ideal deployment locations are used. There is another component of this question, or way to frame it, which is a model designed to work at 1km will not be able to represent the sub-km variability sampled by a network not deployed to make observation representative of 1km areas, and thus potential biases might result.

The abstract should make clear that the statistical models estimate the uncertainty reduction as a function of the number of sites and the model- data mismatch. It should state that the study region is the Bay Area.

Line 17-19: Need to specify that with this particular WRF-STILT framework and assumptions on error co-variance the moderate precision array is preferred.

Line 19-21: This might technically be accurate, but is a bit misleading as some of the assumption made here likely will cause issues with absolute flux accuracy that are much larger than 5%. No top-down method has ever been demonstrated to this point to have fidelity greater than 10% (some would argue demonstration of 20% has yet to be actually achieved).

Line 43: Clarify whether instruments and calibration approaches are mixed within individual networks, between networks, or both.

Lines 69-70: Specify the temporal resolution of the BAAQMD inventory.

Lines 81-86: Make clear in this paragraph exactly what the FIVE product consists of. Is it a particular representative week of hourly emissions, which can be scaled by the user to fit other weeks? Or is scaled by McDonald et al. and provided for any week desired by the user?

Lines 88-91: As a simple approximation, could agricultural emissions be attributed uniformly to farmland? If this approximation is worse than omitting agricultural emissions entirely, state why.

Line 94-95: Please explain in more detail how you regrid from 1 degree to 1 km – this could be done in different manners.

Line 102: The assumption of negligible diurnal cycle needs to be more thoroughly justified, especially since Nassar et al. emphasize the importance of diurnal variation.

Lines 127-132: This description of STILT is confusing. It would be more clear to first explain how STILT is used to calculate influence footprints and only then to describe how the footprints are used to simulate CO2 concentrations at the site locations.

Lines 152-153: What does it mean that "the B matrix has an uncertainty of 100% at the native resolution?" One might take this to mean that the prior estimate is assigned a factor-of-two uncertainty. In supplemental section S2, it seems as though "100% uncertainty" means only that a multiplicative factor fσ is introduced and then set to one.

Lines 155-158: The impact on the result of the choice of a diagonal R matrix should be described.

Line 173: Why not just use the posterior flux error, which is more intuitive and which is shown in the key figure (figure 4)? Why is it an advantage to use a metric similar in form to the averaging kernel matrix?

Lines 184-185: The single sentence "We vary the number of sites (ns) and mismatch error (σm) and perform an ensemble of 20 inversions for each combination to ensure the results are robust." is not adequate to explain this key step in the analysis. For how many different combinations of ns and σm was the error calculated? Which combinations? How were site locations chosen for non-maximal ns? What differs between the 20 inversions performed for the same combination of parameters: the choice of site locations, the random errors, the STILT footprint calculation?

Line 193: The first two parameters are motivated by the assumption of Gaussian errors; what motivates the choice of the other five parameters?

Lines 208-218: This critical part of the procedure is not clear. The derivative of ηËĘ with respect to ns expresses the error reduction to be obtained by adding additional sites. In order to say whether a particular network configuration is noise-limited or site-limited, this reduction should be compared to the reduction to be obtained by reducing the mismatch error, expressed by the derivative of ηËĘ with respect to σm. But the latter is never calculated. Furthermore, it's not clear what is meant by "the $\partial\eta$ËĘ/∂ns curve," how such a curve can be plotted on axes neither of which corresponds to that derivative (as in Figure 4), or what it means for a particular system to be above or below the curve.

Lines 232-237: This is an important point and should be explained more clearly. Exactly what fluxes do you estimate in the averaged case, and what errors are you comparing? Precisely what does "10x better" mean? In the figure it looks as though the errors do not decrease as quickly as predicted by the CLT but seem to level off after about 96 hours; you might explain why this is to be expected. âĂÍ2

Line 273-274: How is this statement about the large systematic error consistent with the 5% uncertainty conclusion highlighted in the abstract?

Section 6.1: Since the text of the supplemental section S5 contains little additional information, consider integrating it into the main text, possibly combining figures S6-S8. Also, specify whether all the observing systems tested in Section 4, or only a subset, were included in the test of sensitivity to domain size.

Section 7: The conclusions should include at least some description of which system designs were found to be site-limited and which noise-limited, since that information is of immediate use to other researchers designing or evaluating their own networks.

Figure 3: The color scheme in the top row is not intuitive to perceive, especially at low resolution as in panel a. Is white used for no estimate as well as for zero flux?

Figure 4: This figure is crucially important, and the design is generally good. However, the shading needs to be reworked so that the gradient is more visible. Also, as mentioned above, it's not clear what defines the red line that separates noise- from site-limited regimes.

Supplement line 56: Why were the decay parameters chosen as they were?

Figure S2: Four judiciously chosen panels would probably be sufficient âĂÍand could be shown at a larger size.

Figure S3: Panels c-e are not as informative and could be omitted.

Figures S6-S8: As in Figure 4, the gradient is not visible enough. Also, the left column corresponds to main text Figure 4, not Figure 3.

Appendix A: In my opinion, this table is not necessary.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-355, 2016.