Response to Reviewer Comments:

We thank the two Anonymous Reviewers for their thorough comments, particularly Reviewer #1 for catching an error in the presentation of the $\partial\eta/\partial n_s$ curve.

Reviewer #1 Comments:

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.

1.) Abstract could lead to false impressions: 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 CO_2 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.

We have updated the text in the abstract to include more qualifiers.

2.) Clarity of presentation: Additionally, error in describing the background condition, or the CO_2 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.

We have updated the text to include the reviewer's suggestions.

Minor Comments:

1.) 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 deployments, particularly 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 reviewer asks about the representativeness of observations made meters above the surface in a dense urban environment. However, as the reviewer notes, this is not a problem our paper aims (or needs) to address. This question is not relevant to a pseudo-data study like ours because our pseudo-data are generated using a model at 1-km resolution, so the pseudo-data are representative of a 1-km area. From a more experimental perspective, it is currently unknown whether sensitivity to local processes necessarily precludes the ability of surface-level sensors to represent domain-wide phenomena, as BEACO_2N is the first network with sufficient quantity and density of sensors to empirically investigate this trade-off. Future analyses of the real BEACO_2N dataset are positioned to answer these and related questions more quantitatively.

Regarding the potential additional source of bias error in 'cheap' network deployments, we have attempted to address this in the original manuscript through the "systematic bias" sensitivity test that was presented in Section 6.1 and Supplemental Section S6.2 where we added a systematic site-specific bias to each observational site in the network. This site-specific bias could be due to representation error, instrument error, etc. Additionally, the companion manuscript (Shusterman *et al.*, 2016) found the BEACO_2N sensors to detect weekly fluctuations in background concentrations to within ±2 ppm.

2.) 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.

Presenting technical details of the statistical models, as the reviewer proposes, would make the abstract overly cumbersome. A reader that is interested in the statistical models will need to consult the main text. However, we have updated the abstract to make it clear that the study area is the Bay Area.

Lines 6-7: "modeled after the BEACO_2N network in California's Bay Area"

3.) 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.

We have updated the text:

Lines 17-18: "Using our inversion framework, we find that…"

4.) 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).

Previous top-down studies have tended to focus on sparser networks and may have been site-limited. Kort *et al.* (2013) claim to constrain fluxes to 10% using a sparser network. Here is the final line of their abstract: "*We estimate that this network can distinguish fluxes on 8 week time scales and 10 km spatial scales to within ~12 g C* m^{-2} d^{-1} *(~10% of average peak fossil* CO_2 *flux in the LA domain).*" Additionally, we are discussing the constraints on the urban region (Area Source), which should be easier to constrain than 10 km grid cells due to it's large size. Given the differences in observational networks, our results are in line with previous estimates reported in the literature.

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

We have updated the text:

Lines 44-46: "Current monitoring networks use a variety of instruments and approaches to calibration with resulting variations in capital and operating costs, network precision, and potential instrument bias."

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

We have updated the text:

Lines 70-71: "The Bay Area Air Quality Management District (BAAQMD) provides detailed **annual** county-level CO_2 emissions information…"

7.) 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?

We have updated the text:

Lines 82-85: "The FIVE traffic CO_2 inventory provides a representative week of hourly CO_2 emissions for San Francisco and other nearby Bay Area cities at 10 km, 4 km, 1 km, and 500 m resolution. This representative week can be scaled to different years based on the state fuel sales (see McDonald et al. (2014) for additional details)."

8.) 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.

For the purpose of the OSSE, this is inconsequential because we are using pseudo-observations. We are attempting to generate an inventory that is a reasonable approximation of the true emissions. The agricultural emissions are only 1% of the anthropogenic emissions, including them will have no impact on the results presented here.

Also, see the response to Reviewer #2's minor comment #3.

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

We have added a detailed explanation to Supplemental Section S3.

10.) Line 102: The assumption of negligible diurnal cycle needs to be more thoroughly justified, especially since Nassar *et al.* (2013) emphasize the importance of diurnal variation.

Nassar *et al.* (2013) show that much of this diurnal variability is in the on-road mobile sector (see their Fig. 1). The other anthropogenic sectors (residential, industrial, electric vehicles, and commercial) show small diurnal variability (diurnal scale factor varies between 0.9 and 1.1). Our bottom up inventory includes diurnal variability from the on-road mobile sector.

11.) 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 CO_2 concentrations at the site locations.

We have updated the text and our description in Supplemental Section S1. See also our response to Reviewer #2's major comment #1.

12.) 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 f_σ is introduced and then set to one.

It means the uncertainty is equal to the standard deviation of the hi-res inventory. We initially considered using other scale factors (f_σ) but ultimately settled on a multiplicative factor of one. However, we left this in the supplement for future studies to potentially modify.

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

Using a diagonal **R** matrix means that all of our model-data mismatch errors are uncorrelated. As such, using a non-diagonal **R** matrix will mean there is less information in the observations because the observations are not independent. We have used a diagonal **R** matrix because the real observations in the BEACO$_2$N network are made at 1 Hz. Here we use hourly observations. As such, it seems fair to assume the original 1 second observations are independent when aggregated to 1 hour. A study with real data could look at the autocorrelation of the model-data mismatch to infer a proper decorrelation length scale, which will almost certainly be much less than 1 hour.

We have updated the text:

Lines 152-153: "Using a diagonal **R** matrix means that we have assumed our mismatch errors are uncorrelated."

14.) 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?

The posterior flux error is an absolute metric and, as such, is less generalizable. The error metric chosen here is more similar to previous work (e.g., Kort *et al.*, 2013) and more generalizable to other studies.

15.) Lines 184-185: The single sentence "We vary the number of sites (n_s) and mismatch error (σ_m) 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 n_s and σ_m was the error calculated? Which combinations? How were site locations chosen for non-maximal n_s? 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?

We have updated the text:

Lines 195-202: "Fig. 4 shows the error in the estimated CO_2 fluxes using the observations over a wide range of observing system scenarios. We vary the number of sites (n_s = [1, 2, …, 34]), mismatch error (σ_m = [0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20] ppm), and perform an ensemble of 20 inversions for each combination to ensure the results are robust. Each ensemble member uses a unique observational network by randomly drawing n_s sites from the population of 34 possible sites. In total, we perform 8,160 inversions. Fig. 4 shows the mean error in the estimated CO_2 fluxes for the area source, line source, and point source as a function of σ_m and n_s. This figure represents the uncertainty in the estimated emissions at a given hour."

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

The x-axis is a linear scale and the y-axis is a log-scale and we can qualitatively see structure in the figure. As such, we assumed that linear and log relationships could yield a significant relationship.

17.) Lines 208-218: This critical part of the procedure is not clear. The derivative of η with respect to n_s 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 η with respect to σ_m. But the latter is never calculated. Furthermore, it's not clear what is meant by "the $\partial\eta/\partial n_s$ 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.

We appreciate the reviewer for catching this mistake. We no longer use the derivatives to estimate these regimes. We now estimate them as the ridgeline from the statistical models.

18.) 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 "10× 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.

The errors in emissions aggregated over 1 week are about a factor of 10 less than the error in hourly emissions. We have updated the text:

Lines 243-249: "Our work is primarily focused on estimating hourly fluxes, however we can further reduce the uncertainty in our estimates by considering temporally averaged fluxes (e.g., what are the weekly or monthly emissions?). Fig. 5 shows the error in our estimate of the area source emissions aggregated over various time-scales. We find the error in our estimate greatly decreases over the first 72 hours. The central limit theorem provides a lower bound on the error reduction we might expect and the error reductions follow this limit reasonably well over the first 72 hours. This implies that our weekly-averaged emission estimate would be 10× better than our hourly emission estimate."

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

The 5% uncertainty is referring to the case shown in the main text (case without the imposed systematic bias).

20.) 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.

We feel that this content is better suited to the supplement.

21.) 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.

We have refrained from putting that in the Conclusions section because it is, somewhat, dependent on the specification of transport error. We allude to this in the conclusion when we qualify one of our findings with the following statement: "*if the dominant source of error is instrument precision.*"

22.) 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?

It is unclear what "*low resolution*" the reviewer is referring to, all of the fluxes are plotted at the same resolution (1 km^2). As for the color scheme, we tested many different color schemes and settled on this one because it facilitated comparison between the different panels (other color schemes were much worse).

White indicates a flux of zero. There is no "*no estimate*", all panels have fluxes at all grid cells.

23.) 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.

We have updated the coloring and the method for defining noise- and site-limited regimes.

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

The decay parameters were judiciously chosen through discussion with the co-authors who have experience creating the bottom-up inventory (Brian McDonald and Robert Harley). Future work could include these parameters in the inversion by defining them as hyperparameters. However, we would no longer have a conjugate prior and would need to move to a sampling approach to obtain the posterior. Given the large number of inversions performed in this study (32,640 inversions), this would be computationally infeasible.

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

This is a vector graphics image in the Supplement. Readers should be able to zoom in on the panels without losing quality.

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

We disagree with the reviewer on this point. The usefulness of these panels varied quite a bit depending on the audience. We have found these panels to be useful for explaining the methodology to scientists who do not typically construct state vectors themselves. However, the panels are not crucial to the manuscript, which is why we put them in the Supplement.

27.) 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.

We thank the reviewer for pointing out the incorrect labeling.

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

We feel that his table could be useful to some readers interested in the robustness of the different statistical models, which is why we have included it in the Supplement.

Reviewer #2 Comments:

It is certainly a novel piece of work and is beneficial to other urban measurement network designs and associated studies. However, some parts of the manuscript need improvements or additional details to better understand the results and their interpretations. Also some clarifications are necessary to improve the manuscript (see the comments below). Hence I would recommend this manuscript for publication after addressing my concerns and comments listed below.

1.) Footprint calculations: My major concern is about the footprint calculations presented in the manuscript. As far as I understand, what it is shown in Fig. 2 is the averaged footprints for all sites in the network in which the footprints are calculated separately for each sites. In that case, I am surprised with such a low value for the averaged footprints on the western side of the model domain even if there are many sites (especially the line source/high-way is on that side, Fig. 1). Although a part of this can be explained with the prevailed wind direction, I don't find enough reasons to justify the shown structure. i.e., it is difficult to believe that those sites don't give much information on surface fluxes for this period. Please clarify and also give additional details (e.g. set up of STILT receptor locations, how strong is the advection, details of vertical mixing etc.).

The reviewer thinks the observations should have more sensitivity to the "*western side of the model domain*". This may stem from how other work has presented the footprints. Much of the previous work showing footprints has plotted them on a log-scale (e.g., Lin *et al.*, 2003, 2004, 2007; Kort *et al.*, 2008) or as percentiles (e.g., Miller *et al.*, 2012, 2013; McKain *et al.*, 2015). We have added a Supplemental Figure (Fig. S3) that shows the footprint on a linear-scale and log-scale. The largest footprint values are in locations where we have an observation site, however there are diffuse signals that cover much of the domain and have a non-trivial contribution to the total signal. Further, the spatial footprints found here are broadly consistent with previous work by Bastien *et al.* (2015; their Fig. 2a-e) who used an adjoint model to determine the sources influencing air pollution in California's East Bay.

As for the request for additional details, it seems that the reviewer may have missed or overlooked some of the model description. Most of what the reviewer requested is already included in the manuscript or supplement. Specifically, the STILT receptor locations (latitude, longitude, and height above ground level) are listed in Table 1 and the details of the WRF simulations (including PBL and LSM schemes) are included in Supplemental Section 1. The advection and vertical mixing are, largely, determined by the WRF model. The cited literature (Lin et al., 2003; Nehrkorn et al., 2010) describe the advection, vertical mixing, and coupling of STILT to WRF in exhaustive detail.

2.) Inverse framework is not well explained: Another criticism is that the inverse framework, although it is a critical component of this study, is not well explained (Sect.4). For example, it is not very clear to me how the state vector is defined for this experiment. What is the spatial and temporal resolutions of the posterior fluxes? This is important to follow the inversion results. This section needs major improvement w.r.t giving additional details.

We have updated the text in Section 4. It now includes a more detailed explanation of the inverse framework including a paragraph describing the state vector.

3.) Reported error estimate: The reported error estimate of the posterior fluxes (5%) is for the best case OSSE and the inversion experiment (rather I would say that it is for "the most idealized case") in which the total model-data mismatch error is assumed to be 0.005 ppm. Since this mismatch error is totally unrealistic in the current scenario, it is not fair to include this "best case" result in the abstract unless the model-data mismatch error (+ other assumptions) is explicitly specified here. Since it is misleading, I would recommend authors to either remove this sentence or provide an error estimate for more reasonable scenario.

It seems that this comment stems from a misunderstanding in the reported errors; our abstract does not report errors using the 0.005 ppm case. The reported error estimate of the posterior fluxes (5%) is for "Network A" (dense network with moderate-precision instruments) from Section 6.2, not the most optimistic case with 0.005 ppm. Fig. 3 shows the most optimistic case because there is only one combination of sites for that case, whereas networks with less sites have multiple configurations that could be shown.

We have updated the abstract to clarify this:

Line 20: "The dense network considered here (modeled after the BEACO_2N network)…"

3.) Fig. 3 and associated statements: I can't see a remarkable performance of inversion in retrieving posterior fluxes as one would expect here, given that the inversion uses a loose prior (100% uncertainty), used all 34 sites, and "unrealistically" low mismatch error (=0.005 ppm which includes model error, representation error, and instrument error). The spatial structure in the CO_2 fluxes is captured only for a few parts of the domain. Unfortunately, this says to me that the most of other sites are not much useful in this case, which is hard to believe. This again points back to my concern regarding the footprint calculation. Need to clarify.

It is unclear which "*associated statements*" the reviewer is referring to. Fig. 3 is merely presented as an example of the posterior fluxes. One of the few statements we make about Fig. 3 is in reference to the diurnal cycle in Fig. 3: "*We find substantial improvements in the diurnal cycle (see panel d).*"

As for the comment about the unremarkable performance of the inversion, we are surprised by this comment. We find the improvement, relative to the prior, quite remarkable. Panel d in Fig. 3 highlights this. The diurnal cycle in the prior is completely incorrect (wrong magnitude and out of phase) but the posterior is able to recover the true diurnal cycle. The spatial structure is recovered reasonably well in the region where we have measurements sites (California's East Bay). Not surprisingly, we see little improvement in distant regions because we do not have sites located there (so the observations are unable to constrain those regions). The results follow pretty much exactly with what one might intuitively expect.

Minor Comments:

1.) L24: Radiative forcing is variable over the years. Please give the value w.r.t year. 1.82 W m^{-2} looks more like the 2011 year values.

We have updated the text.

Line 25: "with a radiative forcing of 1.82 W m^{-2} in 2011 relative to preindustrial times (IPCC 2013)."

2.) L60: The issue is not only with the spatial resolution, but also with the large uncertainty ranges (reported or expected). This issue needs to be addressed clearly in the manuscript to draw the importance of the high resolution inversion modeling, which is to reduce the uncertainty of the emission fluxes. Also mention about the temporal resolution. This is also important especially when cities have peak traffic, industrial, or commercial hours. Need to be mentioned/addressed in the manuscript.

We have updated the text.

Lines 28-29: "…yet current bottom-up inventories still have large uncertainties."

3.) L92-95: From Fig.1 (bottom panel), I see that the natural sources accounts for about 17% (peak to peak, according to CT2013B) of the total fluxes and are varying as expected. This is considerable in comparison with the Bay area traffic sources which accounts for ~50% of the total fluxes. Hence I would expect that using the natural fluxes at coarse resolution (1×1) can generate additional uncertainty and may not be appropriate in this high resolution modeling scenario. Please comment on this.

While a good point, it's not really relevant here because we are performing an OSSE. As such, we have two main goals in constructing the bottom-up inventory: (1) create a bottom-up inventory that is a reasonable approximation of the true emissions and (2) create a bottom-up inventory that is fundamentally different from the prior inventory. For the former goal, the CarbonTracker natural fluxes should provide a reasonable approximation to the true diurnal cycle, albeit with coarse spatial resolution, while the anthropogenic inventory provides high spatio-temporal information about the urban region. Therefore, our bottom-up information should be a decent approximation of the true emissions. As for the latter goal, we are interested in learning what an observational network could tell us about the emissions. So we are using fundamentally different bottom-up inventories to generate the pseudo-observations and serve as the prior for the inversion.

4.) Fig.1: What is “other Anthro” (red line) based on?

We have updated the caption:

Fig. 1 Caption: “Other anthropogenic sources in the BAAQMD inventory (red).”

5.) Section 4: This section needs further improvements to better explain the inversion technique used in this study. Please modify. Also indicate the dimension of "*m*" and "*n*".

The dimensions of m was presented in Supplemental Section S2: “m = 2,133,120, m_t = 240, m_x = 88, and m_y = 101.” We now also included this in the main text Section 4:

Lines 144-147: “The resulting state vector has 2,133,120 elements ($m = m_t \bullet m_x \bullet m_y$ with m_t = 240, m_x = 88, and m_y = 101) and the posterior fluxes will have hourly temporal resolution and 1 km^2 spatial resolution. The dimension of n will depend on the number of sites in the observational network.”

6.) Mathematical formulas (e.g. Sect. 4): Please use standard formatting as followed by the most of the authors/textbooks. For e.g. prior fluxes, x_b in which “*b*” is subscript.

We have updated the notation.

References:

Bastien, L. A., McDonald, B. C., Brown, N. J., & Harley, R. A. (2015). High-resolution mapping of sources contributing to urban air pollution using adjoint sensitivity analysis: benzene and diesel black carbon. Environ Sci Technol, 49(12), 7276-7284. doi: 10.1021/acs.est.5b00686

Kort, E. A., Eluszkiewicz, J., Stephens, B. B., Miller, J. B., Gerbig, C., Nehrkorn, T., Wofsy, S. C. (2008). Emissions of CH4 and N2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations. Geophysical Research Letters, 35(18). doi: 10.1029/2008gl034031

Kort, E. A., Angevine, W. M., Duren, R., and Miller, C. E.: Surface observations for monitoring urban fossil fuel CO2 emissions: Minimum site location requirements for the Los Angeles megacity, Journal of Geophysical Research: Atmospheres, 118, 1577–1584, doi:10.1002/jgrd.50135, 2013.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near- field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, Journal of Geophysical Research-Atmospheres, 108, doi:10.1029/2002jd003161, 2003.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Grainger, C. A., Hollinger, D. Y. (2004). Measuring fluxes of trace gases at regional scales by Lagrangian observations: Application to the CO2 Budget and Rectification Airborne (COBRA) study. Journal of Geophysical Research-Atmospheres, 109(D15). doi: 10.1029/2004jd004754

Lin, J. C., Gerbig, C., Wofsy, S. C., Chow, V. Y., Gottlieb, E., Daube, B. C., & Matross, D. M. (2007). "Designing Lagrangian experiments to measure regional-scale trace gas fluxes". Journal of Geophysical Research, 112(D13). doi: 10.1029/2006jd008077

McDonald, B. C., McBride, Z. C., Martin, E. W., and Harley, R. A.: High-resolution mapping of motor vehicle carbon dioxide emissions, Journal of Geophysical Research-Atmospheres, 119, 5283–5298, doi:10.1002/2013jd021219, 2014.

McKain, K., Down, A., Raciti, S. M., Budney, J., Hutyra, L. R., Floerchinger, C., Wofsy, S. C. (2015). Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts. Proc Natl Acad Sci U S A, 112(7), 1941-1946. doi: 10.1073/pnas.1416261112

Miller, S. M., Kort, E. A., Hirsch, A. I., Dlugokencky, E. J., Andrews, A. E., Xu, X., Wofsy, S. C. (2012). Regional sources of nitrous oxide over the United States: Seasonal variation and spatial distribution. Journal of Geophysical Research, 117(D6). doi: 10.1029/2011jd016951

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., Sweeney, C. (2013). Anthropogenic emissions of methane in the United States. Proc Natl Acad Sci U S A, 110(50), 20018-20022. doi: 10.1073/pnas.1314392110

Nassar, R., Napier-Linton, L., Gurney, K. R., Andres, R. J., Oda, T., Vogel, F. R., and Deng, F.: Improving the temporal and spatial distribution of CO2 emissions from global fossil fuel emission data sets, Journal of Geophysical Research: Atmospheres, 118, 917–933, doi:10.1029/2012jd018196, 2013.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled weather research and forecasting stochastic time-inverted lagrangian transport (WRF–STILT) model, Meteorology and Atmospheric Physics, 107, 51–64, doi:10.1007/s00703-010-0068-x, 2010.

Shusterman, A. A., Teige, V., Turner, A. J., Newman, C., Kim, J., and Cohen, R. C.: The BErkeley Atmospheric CO2 Observation Network: initial evaluation, Atmospheric Chemistry and Physics Discussions, pp. 1–23, doi:10.5194/acp-2016-530, 2016.