Supplemental Material

Bayesian inverse estimation of urban CO_2 emissions: results from a synthetic data simulation over Salt Lake City, UT

Lewis Kunik[1]*, Derek V. Mallia[1], Kevin R. Gurney[2,3], Daniel L. Mendoza[1,4], Tomohiro Oda[5,6], and John C. Lin[1]

1. Department of Atmospheric Sciences, University of Utah, Salt Lake City, Utah, US
2. School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, US
3. School of Life Sciences, Arizona State University, Tempe, Arizona, US
4. Pulmonary Division, University of Utah School of Medicine, Salt Lake City, Utah, US
5. Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, US
6. Goddard Earth Sciences Technology and Research, Universities Space Research Association, Columbia, Maryland, USA

*Corresponding author: lewiskunik@gmail.com

List of Contents:

Text S-1. Overview of computational methods for efficiency

As noted by (Yadav and Michalak, 2013), large linear inverse problems, such as the one discussed in this study, often encounter computational bottlenecks when the discretized state space (in our case, the number of unknowns in the posterior emissions vector, $\hat{\mathbf{s}}$) becomes large. In addition, our inclusion of spatially- and temporally-varying error covariance structures requires full matrix multiplication (including off-diagonal terms which are often neglected) when considering the prior error covariance matrix, **Q**. The size of this matrix is large when considering the domain of interest in this study; 2386 cells are optimized over Salt Lake County at a 6-hourly time resolution for just over a 4-week window (114 time-steps). So, $m_s = 2386$ and $m_\tau = 114$, giving $m = m_s * m_\tau = 272{,}004$ total unknowns. Because **Q** is a covariance matrix describing the relationships between prior unknowns, it is a square matrix with dimensions equal to (272,004 x 272,004). Direct multiplication of a matrix of this size is infeasible even for most modern high-performance supercomputers, and as high-resolution inventories and models become available for larger megacity areas, it is increasingly important to utilize efficient computation methods to solve such problems.

Following (Yadav and Michalak, 2013), we compute **HQ** and $\mathbf{HQH}^T$ matrices at the time-step level, with the slight modification of using a non-constant prior standard error as described in methods for the CarbonTracker-Lagrange framework (www.esrl.noaa.gov/gmd/ccgg/carbontracker-lagrange/). These methods utilize the definition of **Q** as a Kronecker product of smaller matrices describing covariance in space and time (see Equation 5 in main text), and are applied to calculate **HQ** using the following equation:

$$\mathbf{HQ}=\left[\left(\left(\sum_{i=1}^{p}d_{i,1}\mathbf{h}_i\mathbf{I}_{\boldsymbol{\sigma}_i}\right)\mathbf{EI}_{\boldsymbol{\sigma}_1}\right)\left(\left(\sum_{i=1}^{p}d_{i,2}\mathbf{h}_i\mathbf{I}_{\boldsymbol{\sigma}_i}\right)\mathbf{EI}_{\boldsymbol{\sigma}_2}\right)\cdots\left(\left(\sum_{i=1}^{p}d_{i,q}\mathbf{h}_i\mathbf{I}_{\boldsymbol{\sigma}_i}\right)\mathbf{EI}_{\boldsymbol{\sigma}_q}\right)\right]\ (\mathbf{S}1)$$

where *p* and *q* are both equal to the number of timesteps (equivalent to m_τ). **H** is broken into *p* blocks, which are used to calculate **HQ** in a similar block-wise method as shown above. Similarly, $\mathbf{HQH}^T$ is calculated by the addition of time-step-scale summations using the equation:

$$\mathbf{HQH}^T=\left[\left(\left(\sum_{i=1}^{p}d_{i,1}\mathbf{h}_i\mathbf{I}_{\boldsymbol{\sigma}_i}\right)\mathbf{EI}_{\boldsymbol{\sigma}_1}\mathbf{h}_1^T\right)+\left(\left(\sum_{i=1}^{p}d_{i,2}\mathbf{h}_i\mathbf{I}_{\boldsymbol{\sigma}_i}\right)\mathbf{EI}_{\boldsymbol{\sigma}_2}\mathbf{h}_2^T\right)+\cdots+\left(\left(\sum_{i=1}^{p}d_{i,q}\mathbf{h}_i\mathbf{I}_{\boldsymbol{\sigma}_i}\right)\mathbf{EI}_{\boldsymbol{\sigma}_q}\mathbf{h}_q^T\right)\right](\mathbf{S2})$$

These methods are then applied to calculate posterior emissions ($\hat{\mathbf{s}}$), as well as the time-averaged posterior uncertainty covariance ($\bar{\mathbf{V}}_{\hat{\mathbf{s}}}$), following the procedures defined in CarbonTracker-Lagrange documentation. Similar methods are used to calculate the reduced chi-square value of fit (see equation 9 in main text), using the following equation to represent the inverse of the **Q** matrix:

$$\mathbf{Q}^{-1}=\mathbf{I}_\sigma^{-1}\left(\mathbf{D}\otimes\mathbf{E}\right)^{-1}\mathbf{I}_\sigma^{-1}\quad \text{(S3)}$$

and further exploiting the definition of the Kronecker product to give:

$$\mathbf{Q}^{-1}=\mathbf{I}_\sigma^{-1}\left(\mathbf{D}^{-1}\otimes\mathbf{E}^{-1}\right)\mathbf{I}_\sigma^{-1}\quad \text{(S4)}$$

The reduced chi-squared value is thus calculated by the summation of *p* blocks, giving, in total, the product of $\mathbf{Q}^{-1}$ with the square of emissions residuals:

$$\chi_r^2=\frac{1}{\nu}\left[\left(\mathbf{z}-\mathbf{H}\hat{\mathbf{s}}\right)^T\mathbf{R}^{-1}\left(\mathbf{z}-\mathbf{H}\hat{\mathbf{s}}\right)+\left(\sum_{i}^{p}\left(\sum_{j}^{p}d_{j,i}^{-1}\mathbf{S}_j\mathbf{I}_{\sigma_j}^{-1}\right)\mathbf{E}^{-1}\mathbf{I}_{\sigma_i}^{-1}\mathbf{S}_i\right)\right]\ \text{(S5)}$$

where $\mathbf{S}_1$ is a vector of residuals ($\hat{\mathbf{s}}$ - $\mathbf{s}_p$) corresponding to time-step 1, $\mathbf{S}_2$ is a vector of residuals ($\hat{\mathbf{s}}$ - $\mathbf{s}_p$) corresponding to time-step 2, etc. The calculation of reduced chi-squared in this way results in similar computational savings as referenced in (Yadav and Michalak, 2013), with further reduction in complexity because of the inverse operation on smaller sub-matrices (**D**, **E**, and $\mathbf{I}_{\sigma}$) rather than the larger **Q** matrix. This implemented calculation is documented within the **chi_sq.r** script in the code referenced in the Data Accessibility Statement. Despite this improvement, the calculation of reduced Chi-squared remains computationally expensive and repeated calculations are a bottleneck given the computational resources used in this study.

Text S-2. Description of Monte-Carlo methods to determine grid-averaged emissions results

In this study, synthetic data are given random perturbations to represent prescribed model-data mismatch error. Perturbations are generated based on a random normal distribution with a standard error equal to the standard error described in the diagonal of the model-data mismatch matrix (**R**). Because of this feature, posterior adjusted emissions vary based on these random perturbations, and domain-averaged emissions can differ with a standard error of ~0.05 μmol m^{-2} s^{-1} for any given inversion run. In order to obtain a stable average value that reflects an expected set of posterior emissions, we run a Monte Carlo-style simulation where 10,000 unique sets of synthetic data are generated and used to solve for posterior emissions ($\hat{\mathbf{s}}$). Each set of synthetic data is given a unique seed to ensure randomness, and after all sets of synthetic data are generated, the mean value of domain-averaged emissions is taken. After the average posterior is calculated using this method, a random seed was selected by trial and error to determine a standard subset of iterations which produce results roughly equivalent to the large-ensemble method results after only 200 iterations. This seed was then used to produce posterior emissions for which the figures in this study were created.

Figure S-1a shows running averages of 50 ensembles of 200 iterations of random synthetic data generation, with the overall average value shown as a navy-colored line. Domain-averaged values from this study are reported in the main text of this study, and all figures displaying posterior emissions use a set of emissions (averaged over 200 random iterations) with a constant seed set to align closely with the domain average found from this Monte Carlo analysis.

Reduced Chi-squared values also vary with each set of randomly-generated synthetic data, as this statistic is determined from model residuals which vary from run to run. A mean value is calculated similarly to the method above for posterior emissions, but because the computational cost of running this Monte Carlo-style method, the number of ensembles (and iterations per ensemble) is reduced. 6 ensembles of 20 iterations are run, giving 120 random sets of synthetic data. Figure S-1b shows running averages of these ensembles along with the overall average of all Chi-squared iterations. Code for Monte Carlo simulations is given in the **monte_carlo.r** script in the R source code referenced in this study's Data Accessibility Statement.

Text S-3. Additional details regarding Error Covariance and Emissions calculations

In this study, particular attention is given to the **Q** and **R** uncertainty covariance matrices. Corrections to the prior are driven largely by a combination of the structures of these matrices along with individual footprints assigned to each receptor site. As described in the main text (section 2.5), a degree of these corrections are spread to neighboring cells in both space and time, depending on the spatial and temporal correlations of prior and observations errors. Determination of prior error covariance parameters are described in section 2.5, with variogram and auto-correlation analyses shown in Figure S-2. Sensitivity analyses for these values determined here (l_s and l_τ) are given in section 3.2 of the main text, and further analyses of error reduction based on these values are shown in Figure S-3.

Correlations within the **R** matrix are described in section 2.5 as well; however, the **R** matrix is eventually aggregated to represent daily afternoon average observations. Prior to aggregation, observational errors are first defined as described for hourly observations from 18-23 UTC at each site, forming the matrix **R'** which is an (*n' x n'*) matrix where *n'* is equal to the total original number of afternoon hourly observations. Observations and errors are then aggregated using an aggregation operator $\mathbf{W}_n$ in order to reflect a single afternoon-averaged observation per day at each site, given by the equation:

$$\mathbf{R}_{diagonal} = \mathbf{W}_n \mathbf{R}' \quad \text{(S6)}$$

where $\mathbf{R}_{diagonal}$ is a vector of length *n* forming the diagonal elements of the aggregated **R** matrix, and $\mathbf{W}_n$ is the observational aggregation operator weighting each observation proportional to the total number of measurements in each given afternoon. In our synthetic data case, *n'* is equal to 1008 original observations, which are reduced via the above aggregation technique to *n* = 168 observations. Given our use of synthetic data in this experiment, the original observation vector is free of missing data, making $\mathbf{W}_n$ a vector of length *n'* with values uniformly equal to 1/6 (corresponding to the 6 total enhancement values per afternoon per site).

The resulting **R** matrix considers the errors of observations averaged over each afternoon and includes the weighted sum of covariance between observational errors. Errors in the daily afternoon-averaged **R** matrix are uncorrelated between days and sites and are expressed as a diagonal matrix which is reduced in size by a factor of 6. Details of the components of the R matrix are summarized in the main text in Table 1.

Within the emissions vectors, similar aggregation methods are used. Posterior emissions estimates are compared in the main text with "true" emissions on an aggregated space- and time-domain scale. In order to determine domain-averaged emissions, we define the aggregation operator **W**, similarly to equation 8 from Gerbig et al. (2006), to derive a single value representing domain-averaged emissions and associated uncertainty, given by:

$$s_{tot} = \mathbf{W}\mathbf{s}_{grid} \quad \text{(S7)}$$

where $\mathbf{s}_{grid}$ is a vector of grid-scale emissions averaged over time. Using this equation, prior, true, and posterior emissions may be compared at the space- and time-aggregated scale in order to estimate how each emissions vector agrees on the domain average. Posterior emissions are represented on the domain-averaged scale as expected values calculated by a Monte Carlo-style simulation of randomly generated synthetic observations described in Text S-2 of the Supplemental Material.

We can also apply this method in a similar fashion to equation 9 of Gerbig et al. (2006), to derive corresponding values of domain-averaged prior and posterior uncertainty:

$$Q_{tot} = \mathbf{W}\mathbf{Q}_{sum}\mathbf{W}^T \quad \text{(S8)}$$

$$V_{\hat{s}\_tot} = \mathbf{W}\bar{\mathbf{V}}_{\hat{s}}\mathbf{W}^T \quad \text{(S9)}$$

Here, $\mathbf{Q}_{sum}$ is a matrix of grid-scale time-averaged prior uncertainty, and similarly, $\bar{\mathbf{V}}_{\hat{s}}$ is a matrix of grid-scale time-averaged posterior uncertainty. We can use these domain-aggregated values to get a sense of the overall prior and posterior uncertainty, expressing information gain from the inversion as a percent reduction in uncertainty, or error reduction (ER):

$$ER = \frac{Q_{tot} - V_{\hat{s}\_tot}}{Q_{tot}} * 100\,\% \quad \text{(S10)}$$

Through this, we obtain a measure of the amount of uncertainty that is reduced as a result of our constraints on emissions. Note that here, Q_{tot} and $V_{\hat{s}\_tot}$ can be averaged over any time or space domain in order to examine the information gain over specific time periods or grid regions within the inversion domain.

Model performance is also assessed by comparing modeled posterior observations with synthetic data generated at each of the measurement network sites. By convolving footprints described in the **H** matrix with prior and posterior emissions ($\mathbf{s}_{\boldsymbol{p}}$, $\hat{\mathbf{s}}$), we obtain modeled enhancements ($\mathbf{Hs}_{\boldsymbol{p}}$, $\mathbf{H}\hat{\mathbf{s}}$) at each receptor location using the same methods as are used to create synthetic data (without the addition of random errors). Looking at modeled vs. “true” observations, we can observe the standard error (RMSE) and coefficient of determination (r^2) between these observations. Results of these comparisons are shown in Table S-1.

Text S-4. Results using alternative methods for synthetic data generation

An additional method which adopts conventions from real-data analyses is the creation of synthetic data using fine-scale emissions and footprints at the hourly level. Using this method, emissions at the hourly resolution are convolved with un-aggregated hourly footprints in order to obtain individual hourly enhancements from 12-17 MDT. A total of 1008 enhancements are generated from both prior and true hourly emissions and are then aggregated across 12-17 MDT to be expressed as single afternoon-aggregated observations for each site. From this point on, random error is applied to true signals in order derive synthetic data (Equation 4, main text) and methods following this are identical to those from the baseline case.

Additional model runs were carried out using this method in order to examine the impact of hourly emissions and footprints on synthetic data results. This method results in an average individual signal difference of 0.10 ppm in true signals ($\mathbf{Hs}_{\text{truth}}$) and 0.01 ppm in prior signals ($\mathbf{Hs}_{\text{p}}$) between the hourly-derived and 6-hourly-derived (baseline) methods. Differences in average posterior emissions are small between this method and baseline when averaged over the time and space domains - the resulting posterior emissions for this method are 4.60 ± 0.03 μmol m^{-2} s^{-1} (compared to 4.63 μmol m^{-2} s^{-1} from baseline). While these differences are small, this analysis shows that changes in resolution of the fluxes and footprints which derive the inversion's enhancements do have some degree of power to influence posterior emissions.

References for Supplemental Material

Gerbig, C, Lin, JC, Munger, JW, and Wofsy, SC. 2006. What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes? *Atmos Chem Phys* **6**(2): 539–554. DOI: https://doi.org/10.5194/acp-6-539-2006

Yadav, V and Michalak, AM. 2013. Improving computational efficiency in large linear inverse problems: an example from carbon dioxide flux estimation. *Geosci Model Dev* **6** 583-590. DOI: https:// doi:10.5194/gmd-6-583-2013

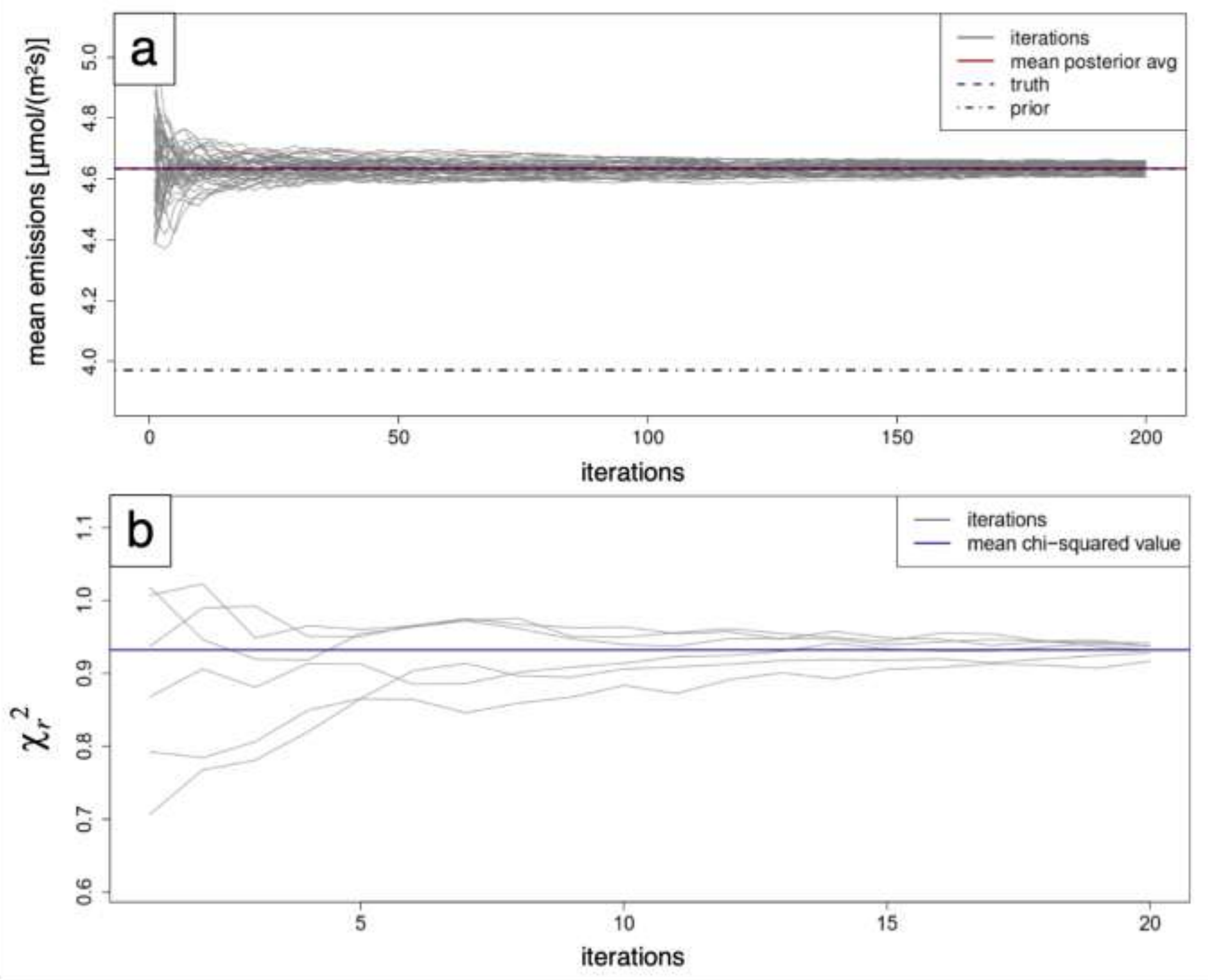

Figure S-1. Monte-Carlo simulation of domain-averaged emissions and reduced chi-square statistic. Ensembles of model simulations are run and **(a)** afternoon-only (18-24 UTC) domain-averaged emissions and **(b)** reduced chi-squared (χ_r^2) values are recorded as running averages for each ensemble. In **(a)**, 50 ensembles of 200-iteration model runs are averaged, and in **(b)**, 6 ensembles of 20-iteration model runs are averaged (due to high computational costs). In **(a)** and **(b)**, overall averages are shown by horizontal blue and red lines, respectively. Running average values for each ensemble are plotted in gray for both plots. Prior and True averages are shown in dotted lines in **(a)**.

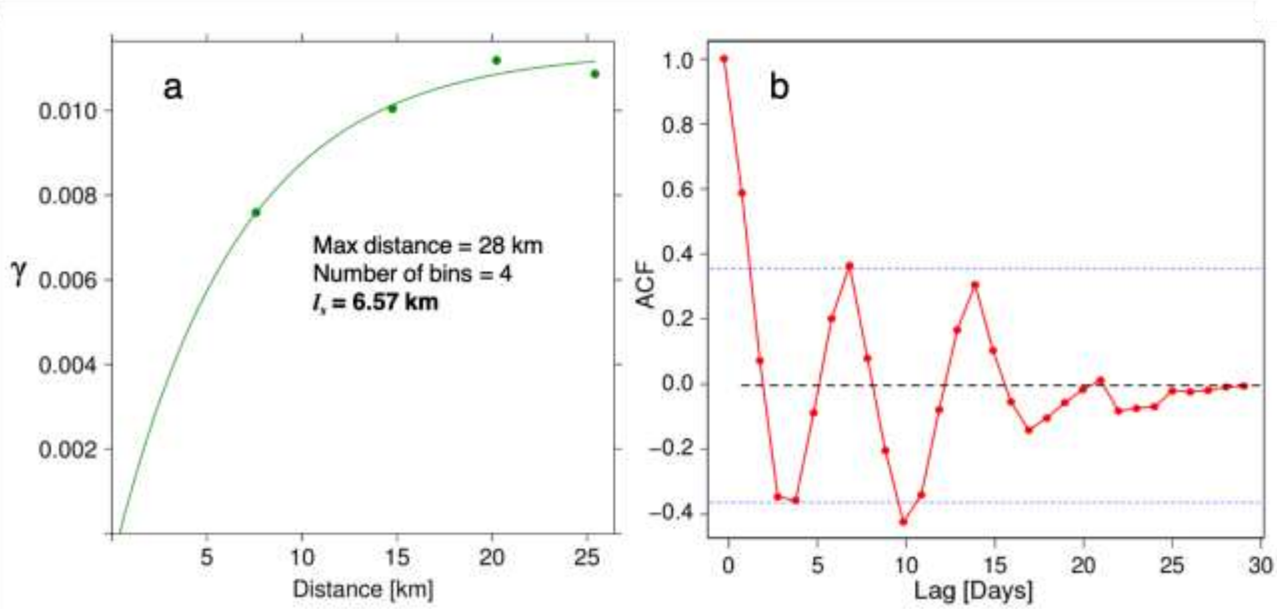

Figure S-2. Covariance length scale parameter determination
(a) Variogram fit using ODIAC-Hestia differences over SLV domain, averaged over the month of September 2015 and modified to omit large outliers. γ (y axis) indicates variance of the spatial distribution and distance (x axis) represents distance between grid cell pairs. Maximum distance for variogram fit is set to 28 kilometers (approximate width of SLV) and variogram parameters are chosen for adequate visual fit. Resulting length scale l_s = 6.57km is rounded down to 6km for this study.
(b) Autocorrelation function (ACF) value vs. lag, in days, calculated using daily afternoon ODIAC-Hestia differences (averaged over 18-24 UTC) over SLV domain for September 2015. Autocorrelations are considered uncorrelated (i.e. no longer statistically different from zero) after first crossing the upper threshold blue dashed-line, which we approximate to l_τ = 2 days.

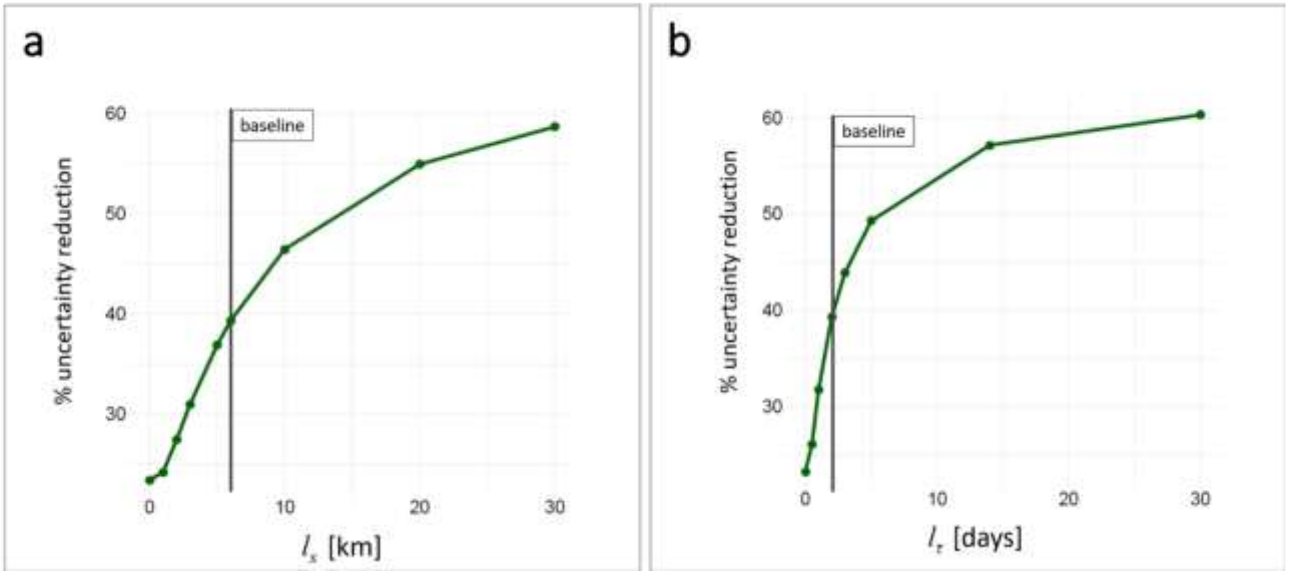

Figure S-3. Uncertainty reduction vs. spatial and temporal correlation length scales (l_s and l_τ). Flux error reduction percentage is shown at **(a)** increasing spatial length scales and **(b)** increasing temporal length scales. Results from baseline configuration are marked by vertical lines. At small length scales for both parameters, uncertainty reduction is minimized.

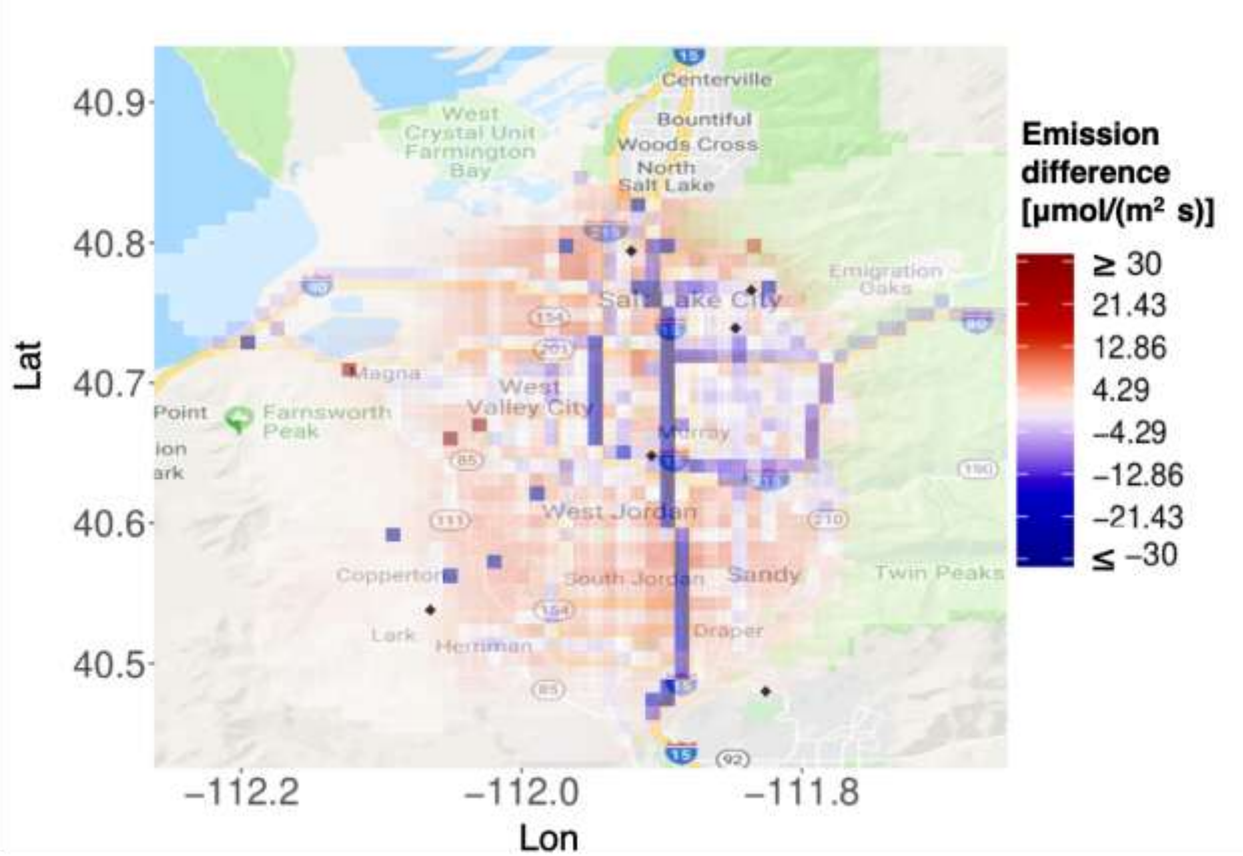

Figure S-4. Posterior minus true emissions averaged over September 2015. True emissions are subtracted from posterior emissions from the baseline case and averaged over the month of September 2015. Values are capped at ± 30 μmol m^{-2} s^{-1} for visualization.

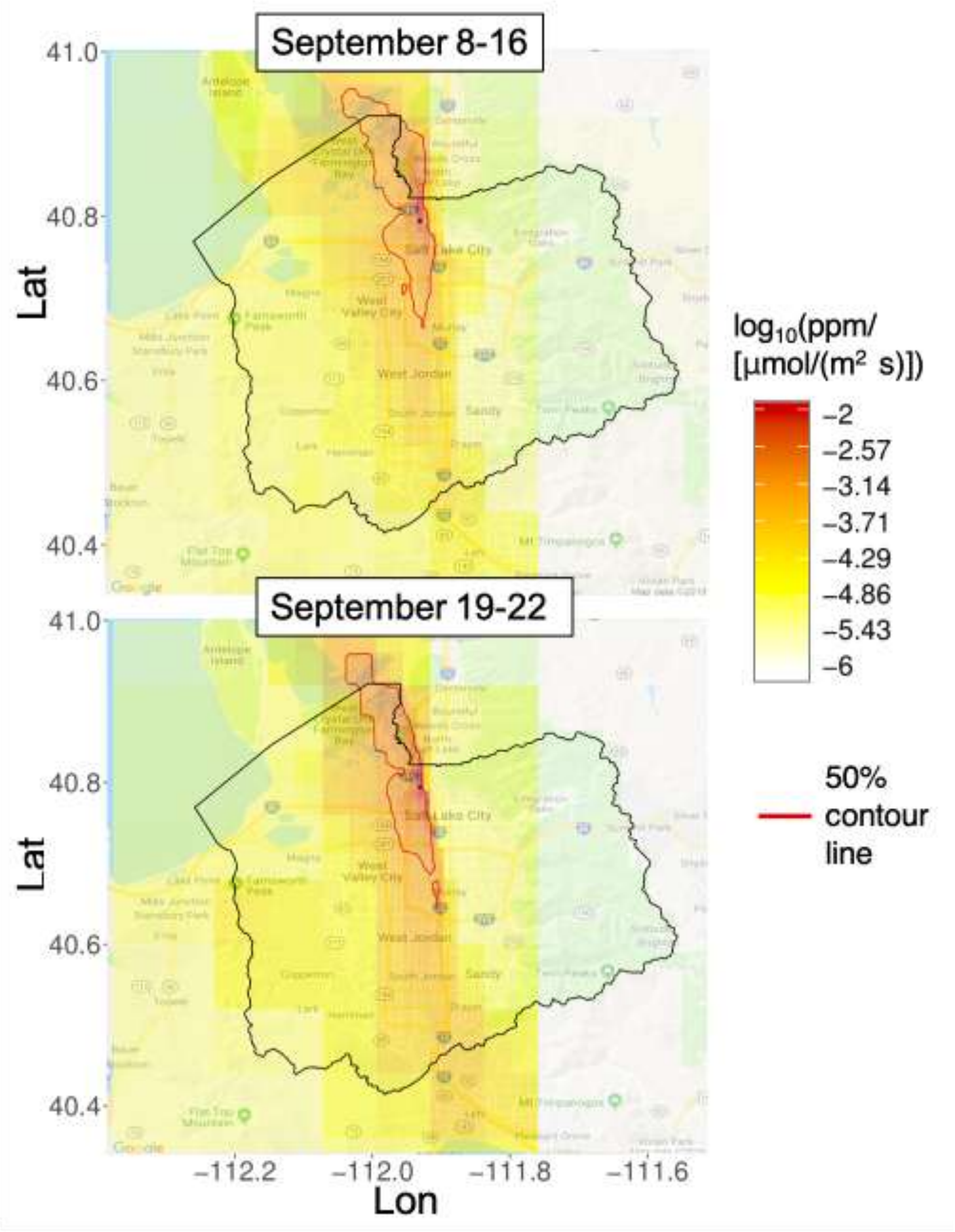

Figure S-5. Average footprint areas for September 8-16 and 19-22. Footprint influence maps are shown averaged over two discrete periods within the inversion domain month: September 8-16 (top) and 19-22 (bottom). 50% contour areas are drawn to show significant near-field areas, and are largely similar between the two time periods.

Site	RMSE [ppm]	r^2	bias (ppm)
UoU	1.11	0.80	-0.15
DBK	1.5	0.69	-0.10
RPK	0.9	0.94	0.31
SUG	1.32	0.79	-0.45
MUR	1.58	0.64	-0.08
SUN	1.4	0.58	0.08
All sites	1.32	0.80	-0.06

Table S-1. Comparison of posterior modeled enhancements vs. synthetic data at different sites across the Salt Lake Valley. Root mean squared error (RMSE) is given as the standard error (in ppm) between posterior enhancements and synthetic data enhancements. R^2 is the coefficient of determination between posterior enhancements and synthetic data enhancements, and bias is the mean posterior enhancement minus the mean synthetic data enhancement. Positive bias values indicate higher posterior signal, while negative bias indicates higher signal in the synthetic data.