A. J. Turner's manuscript prepared
with version 4.2 of the LaTeX class copernicus.cls.
Date: 29 August 2016

Network design for quantifying urban CO_2 emissions: Assessing trade-offs between precision and network density

Alexander J. Turner[1,2], Alexis A. Shusterman[3], Brian C. McDonald[4,*], Virginia Teige[3], Robert A. Harley[4], and Ronald C. Cohen[3,5]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA.
[2]Environmental Energy and Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
[3]Department of Chemistry, University of California at Berkeley, Berkeley, CA, USA.
[4]Department of Civil and Engineering, University of California at Berkeley, Berkeley, CA, USA.
[5]Department of Earth and Planetary Sciences, University of California at Berkeley, Berkeley, CA, USA.
[*]now at: Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado, USA.

Correspondence to: Ronald C. Cohen
(rccohen@berkeley.edu)

Abstract. The majority of anthropogenic CO_2 emissions are attributable to urban areas. While the emissions from urban electricity generation often occur in locations remote from consumption, many of the other emissions occur within the city limits. Evaluating the effectiveness of strategies for controlling these emissions depends on our ability to observe urban CO_2 emissions and attribute them to specific activities. Cost effective strategies for doing so have yet to be described. Here we characterize the ability of a prototype measurement network, modeled after the BEACO$_2$N network in California's Bay Area, in combination with an inverse model based on WRF-STILT to improve our understanding of urban emissions. The pseudo-measurement network includes 34 sites at roughly 2 km spacing covering an area of roughly 400 km^2. The model uses an hourly 1×1 km^2 emission inventory and 1×1 km^2 meteorological calculations. We perform an ensemble of Bayesian atmospheric inversions to sample the combined effects of uncertainties of the pseudo-measurements and the model. We vary the estimates of the combined uncertainty of the pseudo-observations and model over a range of 20 ppm to 0.005 ppm and vary the number of sites from 1 to 34. We use these inversions to develop statistical models that estimate the efficacy of the combined model-observing system at reducing uncertainty in CO_2 emissions. We examine uncertainty in estimated CO_2 fluxes at the urban scale, as well as for sources embedded within the city such as a line source (e.g., a highway) or a point source (e.g., emissions from the stacks of small industrial facilities). Using our inversion framework, we find that a dense network with moderate precision is the preferred setup for estimating area, line, and point sources from a combined uncertainty and cost perspective. The

dense network considered here (modeled after the BEACO$_2$N network) could estimate weekly CO_2 emissions from an urban region with less than 5% error, given our characterization of the combined observation and model uncertainty.

1 Introduction

Carbon dioxide (CO_2) is an atmospheric trace gas and the single largest anthropogenic radiative forcer, with a radiative forcing of 1.82 W m^{-2} in 2011 relative to preindustrial times (IPCC, 2013). CO_2 has increased from 280 ppm in preindustrial times to greater than 400 ppm in the present, largely due to changes in fossil fuel emissions. Over 70% of these fossil fuel CO_2 emissions in the United States (US) are attributable to urban areas (EIA, 2015; Hutyra et al., 2014), yet current bottom-up inventories still have large uncertainties. As such, quantifying and monitoring the emissions from urban areas is crucial to strategies for reducing future increases in CO_2.

Numerous studies have performed top-down estimations of CO_2 emissions using observations from urban surface monitoring networks of various sizes (e.g., Gratani and Varone, 2005; McKain et al., 2012; Newman et al., 2013; Lauvaux et al., 2013; Breon et al., 2015; Turnbull et al., 2015). However, it's not immediately clear how many sites are necessary to monitor the emissions from an urban area. Kort et al. (2013) found that a surface monitoring network would need at least 8 sites operating for 8 weeks to accurately estimate CO_2 emissions in Los Angeles. Yet most current urban monitoring networks have fewer than 8 sites but operate for much longer than 8 weeks. For example, Gratani and Varone (2005) used a single site in Rome, Newman et al. (2013) used a single site in Los Angeles, Lauvaux et al. (2013) used two sites in Davos, Switzerland, McKain et al. (2012) used a network of 5 sites in Salt Lake City, and Breon et al. (2015) used 5 sites in Paris. Recent work from Turnbull et al. (2015) employed a denser network of 12 sites in Indianapolis.

This issue is further complicated by bias and noise in both the measurements and the modeling framework. The combined model and measurement error is known as the model-data mismatch error (hereafter referred to as the "mismatch error"). Current monitoring networks use a mix of instruments and approaches to calibration with resulting variations of capital and operating costs, network precision, and potential instrument bias. Monitoring networks located in regions with complex orography are challenging for atmospheric transport calculations, making it more difficult to determine the dispersion from sources.

The tradeoff between measurement network density and mismatch error has yet to be characterized. Understanding these tradeoffs is crucial to reducing the uncertainty in emissions from urban regions and to developing cost-effective urban monitoring networks. Here we present a high-resolution inventory of CO_2 fluxes and a numerical model that relates atmospheric observations to high resolution surface fluxes. We then use this inventory and model in a series of observing system simulation experiments (OSSEs) to investigate the tradeoff between reductions in the mismatch error and increases in the measurement network density. We develop statistical models to characterize this relationship for different types of sources in the San Francisco Bay Area, identify limiting regimes, and recommend future observing strategies.

2 Constructing a high resolution regional CO_2 inventory

McDonald et al. (2014) demonstrated that 1×1 km^2 spatial resolution is necessary to resolve the gradients in urban CO_2 fluxes from highways. However, most of the existing CO_2 anthropogenic inventories are not available at this resolution. For example, EDGAR (European Commission, 2011) and VULCAN (Gurney et al., 2009) are only available at $0.1^\circ \times 0.1^\circ$ and 10×10 km^2, respectively. A notable exception is the Odiac fossil fuel CO_2 inventory (Oda and Maksyutov, 2011) which is based on satellite-observed nightlight data and available globally at 1×1 km^2 resolution. High resolution fossil fuel CO_2 emissions are available for select cities and sectors such as Paris through the AirParif inventory (Breon et al., 2015, http://www.airparif.asso.fr/en/index/index) and Indianapolis, Los Angeles, Salt Lake City, and Phoenix through the HESTIA project (Gurney et al., 2012, http://hestia.project.asu.edu/); three recent studies (Gately et al., 2013; McDonald et al., 2014; Gately et al., 2015) developed high resolution CO_2 emissions from vehicular traffic.

The Bay Area Air Quality Management District (BAAQMD) provides detailed annual county-level CO_2 emissions information for San Francisco and California's Bay Area (Mangat et al., 2010). The BAAQMD found that the transportation sector accounted for 36% of the Bay Area anthropogenic emissions, industrial and commercial for 36%, electricity for 16%, residential fuel usage for 7%, off-road equipment for 3.0%, and agriculture for 1%. The BAAQMD also reports CO_2 emissions for 4,375 point sources in the Bay Area. We geocode these point sources based on the addresses provided by the BAAQMD. These point sources capture the emissions from the industrial, commercial, and electricity sectors. We map residential fuel usage to population using block level population data from the 2010 US Census and apply a temporal temperature scaling based on Deschłnes and Greenstone (2011); the resulting temporal scaling effect is small due to the temperate climate in the East Bay region of the SF Bay Area.

Here we use the traffic CO_2 emissions from the fuel-based inventory for vehicle emissions (FIVE) developed by McDonald et al. (2014). 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). The FIVE inventory is constructed by partitioning CO_2 emissions using state-level fuel data to individual roads with road-specific traffic count data and temporal patterns from weigh-in-motion data. In this manner, CO_2 emissions from the FIVE inventory will be consistent with state and national CO_2 budgets and can easily be scaled to different years.

Combining the industrial, commercial, electricity, residential, and traffic emissions account for 95.8% of the anthropogenic CO_2 emissions in the Bay Area. We do not have high resolution proxy data for the off-road equipment or agriculture sectors in the Bay Area and have chosen to assume their contributions are smaller than the uncertainty in the total budget; therefore we neglect these sectors in the construction of our inventory.

CarbonTracker CT2013B (http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/; Peters et al., 2007) provides 3 hourly fossil fuel, ocean, biogenic, and fire CO_2 fluxes at $1^\circ \times 1^\circ$ resolution. These fluxes are optimized to agree with atmospheric CO_2 observations. We regrid these fluxes to 1×1 km^2 spatial resolution (see Supplemental Section S3) and use the fire, ocean, and biogenic sectors to account for our natural fluxes.

Fig. 1 shows snapshots of the CO_2 fluxes from our inventory at 4 different times of day and the a-temporal fluxes from EDGAR v4.2 FT2010 (European Commission, 2011). From Fig. 1 we can see the inventory clearly resolves the large CO_2 gradients from highways, confirming that 1×1 km^2 spatial resolution is sufficient to resolve urban CO_2 fluxes from highways. The bottom panel of Fig. 1 shows a time series of Bay Area CO_2 fluxes broken down by source. The diurnal cycle in our inventory is largely driven by the traffic emissions with modest uptake from the biosphere during the middle of the day. Other anthropogenic sources were assumed to have a negligible diurnal cycle (Nassar et al., 2013). In what follows, we use EDGAR as the prior and the high spatio-temporal resolution inventory as the "truth".

[Fig. 1 about here.]

3 The Berkeley Atmospheric CO_2 Observation Network (BEACO$_2$N)

The Berkeley Atmospheric CO_2 Observation Network ("BEACO$_2$N", see http://beacon.berkeley.edu and Shusterman et al., 2016) was founded in 2012 as a web of approximately 25 carbon dioxide sensing "nodes" stationed atop schools and museums in the Oakland, CA metropolitan area (see Table 1). With sensors installed on an approximately 2 km square grid, BEACO$_2$N is the only surface-level (3 to 130 m a.g.l.) greenhouse gas monitoring system with roughly the same spatial resolution as the emissions inventories described above. Each node requires only a standard, 120V power source and is sited on pre-existing structures based on voluntary, no-cost partnerships. The BEACO$_2$N configuration therefore represents a reasonable expectation and is one model for future monitoring networks aimed at constraining CO_2 fluxes at neighborhood scales within an urban dome.

[Table 1 about here.]

BEACO$_2$N's unprecedented spatial density is achieved by exploiting lower cost instrumentation than has traditionally been utilized for ambient CO_2 detection. The non-dispersive infrared (NDIR) absorption sensor used in each BEACO$_2$N node (http://www.vaisala.com/en/products/carbondioxide/Pages/GMP343.aspx) has been seen to possess adequate sensitivity to resolve diurnal as well as seasonal phenomena relevant to urban environments (Rigby et al., 2008) and costs one to two orders of magnitude less than the commercial cavity ring-down instruments commonly used in other networks. However, the low-cost NDIR sensor is more susceptible to factors such as temporal drift and environmental instability that can negatively impact data quality. This trade-off between mismatch error and network density is explored below.

4 Observing system simulation experiments

CO_2 concentrations were simulated at 34 sites in the BEACO$_2$N network with the Stochastic Time-Inverted Lagrangian Transport (STILT) model (Lin et al., 2003), coupled to the Weather Research and Forecasting (WRF) meso-scale meteorological model run at 1×1 km^2 grid resolution (WRF-STILT; Nehrkorn et al., 2010). WRF-STILT computes footprints (Δ CO_2 per surface flux, or ppm per μmol$\cdot$m$^{-2}\cdot$s^{-1}; See Supplemental Section S1 and Lin et al. (2003) for additional details) for each observation that relate the hourly 1 km^2 CO_2 fluxes ($\mathbf{x}$; an $m\times1$ vector) to the observations ($\mathbf{y}$; an $n\times1$ vector):

$$\mathbf{y}=\mathbf{H}\mathbf{x} \tag{1}$$

Each row of the $n\times m$ Jacobian matrix ($\mathbf{H}=\partial\mathbf{y}/\partial\mathbf{x}$) is a reshaped footprint. Fig. 2 shows the location of the sites and the average network footprint for Sept 15 to 22.

[Fig. 2 about here.]

Our aim is to estimate hourly CO_2 fluxes at 1 km^2 over a one week period. As such, the model domain is 88 km $\times$ 101 km and we solve for 240 hours of fluxes (1 week plus 3 additional days of back trajectories). The resulting state vector has 2,133,120 elements ($m=m_t\cdot m_x\cdot m_y$ with $m_t=240$, $m_x=88$, $m_y=101$) 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.

Here we use our high resolution CO_2 inventory ($\mathbf{x}^\star$; an $m\times1$ vector) to generate synthetic observations ($\mathbf{y}^\star$; an $n\times1$ vector):

$$\mathbf{y}^\star=\mathbf{H}\mathbf{x}^\star+\boldsymbol{\varepsilon} \tag{2}$$

where $\boldsymbol{\varepsilon}$ is an $n\times1$ vector of normally distributed noise with mean $\boldsymbol{\epsilon}_b$ and diagonal covariance matrix $\mathbf{R}$: $\boldsymbol{\varepsilon}\sim\mathcal{N}(\boldsymbol{\epsilon}_b,\mathbf{R})$. Using a diagonal $\mathbf{R}$ matrix means that we have assumed our mismatch errors are uncorrelated. Our base case inversion assumes the mean bias is zero: $\boldsymbol{\epsilon}_b=\mathbf{0}$. We evaluate the sensitivity to this assumption in Section 6 and Supplemental Section S6.2. These synthetic observations can then be used in a Bayesian inference framework to estimate the optimal CO_2 fluxes (c.f.

Rodgers, 2000). Assuming the prior and likelihood distributions are Gaussian gives us a closed-form solution for the posterior CO_2 fluxes:

$$\hat{\mathbf{x}} = \mathbf{x}_p + (\mathbf{HB})^T \left(\mathbf{HBH}^T + \mathbf{R}\right)^{-1} (\mathbf{y}^\star - \mathbf{H}\mathbf{x}_p) \tag{3}$$

where $\mathbf{x}_p$ is an $m \times 1$ vector of prior CO_2 fluxes, comprised of a coarse ($10{\times}10$ km^2) a-temporal EDGAR v4.2 FT2010 anthropogenic CO_2 inventory and natural fluxes from CarbonTracker CT2013B, regridded to $1{\times}1$ km^2. $\mathbf{B}$ is the $m \times m$ prior error covariance matrix. The prior error covariance matrix can be expressed as a Kroenecker product (cf. Meirink et al., 2008; Singh et al., 2011; Yadav and Michalak, 2013) of temporal and spatial covariance matrices: $\mathbf{B} = \mathbf{D} \otimes \mathbf{E}$ where $\mathbf{D}$ is the temporal covariance matrix and $\mathbf{E}$ is the spatial covariance matrix. The $\mathbf{B}$ matrix has an uncertainty of 100% at the native resolution and the spatial and temporal covariance matrices are fully populated (see Supplemental Section S2 for more details).

We do not explicitly represent the individual error terms contributing to the $\mathbf{R}$ matrix (instrument error, model error, and representation error). Instead, we have assumed that the $\mathbf{R}$ matrix is diagonal and can be characterized by a single parameter: the total mismatch error (σ_m; $\mathbf{R} = \sigma_m^2 \mathbf{I}$), which represents the combined effects of the different error components.

Fig. 3 shows an example of the estimated CO_2 fluxes. We can see that the posterior fluxes capture more of the spatial variability in the CO_2 fluxes than the prior fluxes in the region where the network is deployed. We find substantial improvements in the diurnal cycle (see panel d). Previous work has used the posterior covariance matrix ($\mathbf{Q} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1}\right)^{-1}$), averaging kernel matrix ($\mathbf{A} = \mathbf{I} - \mathbf{Q}\mathbf{B}^{-1}$), and the degrees of freedom for signal ($\mathrm{DOFs} = \mathrm{tr}(\mathbf{A})$) as metrics to evaluate the information content of different observing systems (e.g., Kort et al., 2013; Wu et al., 2016). However, it is computationally infeasible to construct these $m \times m$ matrices for our application as $m > 10^6$ and storing them would require $\sim$36 Tb of memory (assuming double precision, dense matrices).

[Fig. 3 about here.]

Instead, we evaluate the efficacy of the posterior fluxes by taking the norm of the difference between the posterior fluxes and the true fluxes: $||\hat{\mathbf{x}} - \mathbf{x}^\star||_2$. We express this as a relative improvement by comparing the norm of the difference between the prior fluxes and the true fluxes:

$$\eta = 1 - \frac{||\hat{\mathbf{x}} - \mathbf{x}^\star||_2}{||\mathbf{x}_p - \mathbf{x}^\star||_2} \tag{4}$$

This error metric, η, was chosen as it has a similar form to the averaging kernel matrix but it also allows us to directly compare the posterior fluxes to the true fluxes. This relative error metric can be related to the flux error (see Supplemental Section S5). As such, we can use the error metric to evaluate the ability of the observing system to resolve three types of emission sources: (1) area, (2) line, and (3) point sources, by examining a subset of grid cells in the domain (see Section S3 for more details). The area source (AS) examined here is the East Bay urban dome (147 $\pm$ 55 tC hr^{-1};

uncertainty is the 1-σ range of hourly fluxes from the high resolution inventory), the line source (LS) is Interstate 880 and the Bay Bridge (45 $\pm$ 20 tC hr^{-1}), and the point sources (PS) are 4 large CO_2 sources in the East Bay (9 $\pm$ 4 tC hr^{-1}). For comparison, Salt Lake City emits $\sim$300 $\pm$ 50 tC hr^{-1} (McKain et al., 2012). The top panel of Fig. 2 shows these three source types.

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 ($n_s = [1,2,\ldots,34]$), mismatch error ($\sigma_m = [0.005,0.01,0.02,0.05,0.1,0.2,0.5,1,2,5,10,20]$ 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.

[Fig. 4 about here.]

5 Simplified statistical models of error reduction

We develop statistical models to predict the error reduction and quantify the importance of the different factors governing the error reduction. We tested all combinations of models with the following 7 parameters (127 possible combinations): $\sqrt{\sigma_m}$, $\sqrt{n_s}$, $\ln(\sigma_m)$, $\ln(n_s)$, σ_m, n_s, and a constant. These statistical models were evaluated using Akaike information criterion (AIC) and Bayesian information criterion (BIC). The following statistical models were found to be best:

$$\hat{\eta}_{\mathrm{AS}} = \beta_6\sqrt{\sigma_m} + \beta_5\sqrt{n_s} + \beta_4\ln(\sigma_m) + \beta_3\ln(n_s) + \beta_2\sigma_m + \beta_0 \quad (5)$$

$$\hat{\eta}_{\mathrm{LS}} = \beta_6\sqrt{\sigma_m} + \beta_5\sqrt{n_s} + \beta_4\ln(\sigma_m) + \beta_3\ln(n_s) + \beta_2\sigma_m + \beta_1 n_s \quad (6)$$

$$\hat{\eta}_{\mathrm{PS}} = \beta_6\sqrt{\sigma_m} + \beta_5\sqrt{n_s} + \beta_4\ln(\sigma_m) + \beta_2\sigma_m + \beta_0 \quad (7)$$

All the regression coefficients (β_i) in the statistical models yielded statistically significant ($p < 0.001$) parameters based on F-tests (see the Supplemental Section S7 for the regression coefficients and model selection criterion).

We find the $\sqrt{\sigma_m}$, $\sqrt{n_s}$, $\ln(\sigma_m)$, and σ_m parameters in all three statistical models (Eq. 5–7). This dependence on $\sqrt{n_s}$ and $\sqrt{\sigma_m}$ logically follows from the assumption of Gaussian errors in the derivation of the posterior CO_2 fluxes (Eq. 3) and the basic properties of variance. These two parameters tend to be dominant and generally explain more than 50% of the variance. As such, we suspect that these two parameters are the most important and that other terms are capturing higher-order effects.

These statistical models can also be used to define the regimes where increasing the number of sites in the observing system is more important and those where reducing the mismatch error is more important. We estimate these regimes using the ridge line from the statistical models (Eq. 5–7). From Fig. 4 we can see two distinct regimes: *noise-limited* and *site-limited*. Observing systems that lie above the ridge line are in the the noise-limited regime where the error reduction is largely governed by the mismatch error in the observing system. Conversely, observing systems below the ridge line are in the the site-limited regime where the error reduction is largely governed by the number of sites in the observing system.

The mismatch error is controlled by the instrument, representation, and model error. In the noise-limited regime reducing these errors will provide the greatest benefit. Whereas, in the site-limited regime the greatest benefit will come from increasing the number of sites in the observing system and there will only be marginal benefit from reducing the instrument, representation, and model error.

6 Discussion

Three conclusions we can draw from Fig. 4 for California's East Bay are:

1. Achieving $\sigma_m = 1$ ppm adds value. There is relatively little additional benefit to reducing mismatch error to 0.1 ppm, particularly for estimating line or point source emissions.

2. At $\sigma_m = 1$ ppm there is a benefit to increasing the number of sites, but this benefit increases slower than $\sqrt{n_s}$.

3. At $\sigma_m = 5$ ppm there is little benefit from increasing the number of sites; reducing the noise would add more value.

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 $10\times$ better than our hourly emission estimate.

[Fig. 5 about here.]

6.1 Additional factors affecting observing system design

We considered three additional factors that could adversely impact an observing system: (1) inversion domain size, (2) site-specific systematic biases, and (3) using only daytime observations.

Our results are found to be largely insensitive to the inversion domain size (see Fig. S6). This is discerned through a set of sensitivity OSSEs with a reduced domain size. We find that inversions on the reduced domain were only marginally worse at reducing the error ($\sim 1\%$) than inversions on the full domain (see Supplemental Section S6.1). This is due to the strong local signal in the footprint of the measurements (see bottom panel of Fig. 2). As such, the non-local emission sources do not adversely impact our ability to estimate urban emissions.

Biases can adversely impact the observing system (see Fig. S7). To test the impacts of biases in the modeling-measurement framework, we repeated the OSSEs outlined in Section 4 but included a systematic bias. The bias was unique to each site and was drawn from a normal distribution ($\boldsymbol{\epsilon}_b \sim \mathcal{N}\left(\mathbf{0}, \sigma_b^2 \mathbf{I}\right)$; $\sigma_b = 1$ ppm). There are three major findings from the OSSEs with systematic biases:

1. Systematic biases become particularly problematic when the spread of the potential biases (defined here as σ_b) is larger than the mismatch error ($\sigma_b > \sigma_m$). This is because we have defined the observational error covariance matrix as: $\mathbf{R} = \sigma_m^2 \mathbf{I}$. However, if $\sigma_b > \sigma_m$ with a dense observing system then the site-specific biases will artificially inflate the observational error covariance matrix: $\mathbf{R} \approx \left(\sigma_m^2 + \sigma_b^2\right) \mathbf{I}$ and the errors will be incorrectly characterized in the observing system. As long as $\sigma_b < \sigma_m$ then $\mathbf{R} = \sigma_m^2 \mathbf{I}$ and the characterization of the errors will be appropriate.

2. Observing systems with more sites are generally less affected by site-specific systematic biases. This is because observing systems with a small number of sites rely heavily on those few sites. An observing system with many sites is less reliant on a single site and the site-specific systematic biases act more like additional noise in the observing system.

3. Systematic biases have a greater impact when estimating an area source than line and point sources. This is because an airmass sensitive to a line or point source will have a greater enhancement relative to the background compared to a diffuse area source, thus there is a larger signal-to-noise ratio for these sources and a systematic bias is less important.

During the day, model calculations of the PBL height are more reliable leading to a temptation to omit the nighttime data from the analysis. However, emissions at night can be as much as 30% of the total and ignoring them makes estimates of urban emissions strongly dependent on prior assumptions. Our observing system would be unable to correct the misrepresented nighttime emissions of our a-temporal prior without using nighttime observations. As a result, even our most optimistic observing system would have a systematic $\sim$50 tC hr^{-1} error ($\sim$30%) in the estimated area source emissions due to the misrepresented nighttime emissions.

6.2 Potential cost tradeoffs

We consider two potential observing systems:

1. "Network A" ($n_s = 25$, $\sigma_m = 1$ ppm): A dense network with moderate-precision instruments. This network is similar to the BEACO$_2$N network described in Section 3. We assume a cost

of \$5,000 per instrument giving a total cost of \$125,000. This network is shown as a purple star in the left column of Fig. 4.

2. "Network B" ($n_s = 3$, $\sigma_m = 0.1$ ppm): A sparse network with of high-precision instruments. This network uses cavity-ring down instruments. We assume a cost of \$50,000 per instrument giving a total cost of \$150,000. This network is shown as a green star in the left column of Fig. 4.

We note that the assumed mismatch error for these two potential observing systems is defined as the instrument error and assumes there is no contribution from model or transport errors.

The cost for these two networks is comparable. From Fig. 4, we find that the sparse "Network B" is site-limited in all cases whereas the dense "Network A" is in the noise-limited regime. Further, we find that the dense "Network A" has less error in the estimate of all source types in San Francisco's East Bay. Networks sitting on the ridge line are at the optimal balance between precision and number of sites.

6.3 The relationship between network density and transport error

In this work we have treated transport error and the number of measurement sites as independent. However, in practice, there would be a relationship between the transport error and measurement network density. This can be understood with a thought experiment using two different observing systems to estimate emissions: a sparse network with a single site and an infinitely dense network (sites at each grid cell in our domain). Estimating emissions with the sparse network would require us to simulate the atmospheric transport with high fidelity if we are to reliably say anything about emissions upwind of our site. This is especially true for point sources. Any errors in the simulated atmospheric transport would adversely impact the estimated emissions, whereas the infinitely dense network could potentially neglect atmospheric transport and use data from only the local grid cell to estimate emissions. This is because the differential signal at each site would be largely governed by the local emissions. Explicitly quantifying this relationship between transport error and measurement network density should be the focus of future work.

7 Conclusions

Understanding the factors that govern our ability to estimate urban greenhouse gas emissions are crucial to improving an observing system and reducing the uncertainty in emission estimates. Here we have quantitatively mapped the errors in CO_2 emission estimates from different observing systems for three different types of sources in California's Bay Area: area sources, line sources, and point sources. Our results show that different observing systems may fall into noise or site-limited regimes where reducing the uncertainty in the estimated emissions is governed by a single factor; these regimes differ for the source types. Identifying the regime an observing system is in will help inform future improvements to the observing system. A number of prior urban CO_2 experiments have defined as a goal, the understanding of emissions to less than 10% (e.g., Kort et al., 2013; Wu et al., 2016). We find that a BEACO_2N-like network could achieve this accuracy and precision with 1 week of observations, if the dominant source of error is instrument precision. This conclusion may motivate a re-examining of the conventional instrument quality-oriented design of CO_2 observing systems, according to the stated goal of a given network.

Acknowledgements. This work was supported by a Department of Energy (DOE) Computational Science Graduate Fellowship (CSGF) to AJT, a National Science Foundation (NSF) Grant 1035050 to RCC, and a Bay Area Air Quality Management District (BAAQMD) Grant 2013.145 to RCC. AAS was supported by a National Science Foundation Graduate Research Fellowship. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank M. Sulprizio (Harvard University) for gridding the US Census population data and the UC Berkeley Academic Computing center for access to computing resources.

References

Breon, F. M., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Remy, I., Ramonet, M., Dieudonn, E., Lopez, M., Schmidt, M., Perrussel, O., and Ciais, P.: An attempt at estimating Paris area CO_2 emissions from atmospheric concentration measurements, Atmospheric Chemistry and Physics, 15, 1707–1724, doi: 10.5194/acp-15-1707-2015, 2015.

Deschłnes, O. and Greenstone, M.: Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US, American Economic Journal: Applied Economics, 3, 152–185, doi: 10.1257/app.3.4.152, 2011.

EIA, U.: Emissions of Greenhouse Gases in the U.S., Tech. rep., U.S. Energy Information Administration, 2015.

European Commission: Emission Database for Global Atmospheric Research (EDGAR), release version 4.2, Tech. rep., Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL), 2011.

Gately, C. K., Hutyra, L. R., Wing, I. S., and Brondfield, M. N.: A bottom up approach to on-road CO_2 emissions estimates: improved spatial accuracy and applications for regional planning, Environ Sci Technol, 47, 2423–30, doi:10.1021/es304238v, 2013.

Gately, C. K., Hutyra, L. R., and Sue Wing, I.: Cities, traffic, and CO_2: A multidecadal assessment of trends, drivers, and scaling relationships, Proc Natl Acad Sci U S A, 112, 4999–5004, doi:10.1073/pnas.1421723112, 2015.

Gratani, L. and Varone, L.: Daily and seasonal variation of CO_2 in the city of Rome in relationship with the traffic volume, Atmospheric Environment, 39, 2619–2624, doi:10.1016/j.atmosenv.2005.01.013, 2005.

Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., and de la Rue du Can, S.: High resolution fossil fuel combustion CO_2 emission fluxes for the United States, Environ Sci Technol, 43, 5535–41, 2009.

Gurney, K. R., Razlivanov, I., Song, Y., Zhou, Y., Benes, B., and Abdul-Massih, M.: Quantification of fossil fuel CO_2 emissions on the building/street scale for a large U.S. city, Environ Sci Technol, 46, 12 194–202, doi:10.1021/es3011282, 2012.

Hutyra, L. R., Duren, R., Gurney, K. R., Grimm, N., Kort, E. A., Larson, E., and Shrestha, G.: Urbanization and the carbon cycle: Current capabilities and research outlook from the natural sciences perspective, Earth's Future, 2, 473–495, doi:10.1002/2014ef000255, 2014.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Tech. rep., 2013.

Kort, E. A., Angevine, W. M., Duren, R., and Miller, C. E.: Surface observations for monitoring urban fossil fuel CO_2 emissions: Minimum site location requirements for the Los Angeles megacity, Journal of Geophysical Research: Atmospheres, 118, 1577–1584, doi:10.1002/jgrd.50135, 2013.

Lauvaux, T., Miles, N. L., Richardson, S. J., Deng, A., Stauffer, D. R., Davis, K. J., Jacobson, G., Rella, C., Calonder, G.-P., and DeCola, P. L.: Urban Emissions of CO_2 from Davos, Switzerland: The First Real-Time Monitoring System Using an Atmospheric Inversion Technique, Journal of Applied Meteorology and Climatology, 52, 2654–2668, doi:10.1175/jamc-d-13-038.1, 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, ACH 2–1–ACH 2–17, doi:10.1029/2002jd003161, 2003.

Mangat, T. S., Claire, S. J., Dinh, T. M., Fanai, A. K., Nguyen, M. H., and Schultz, S. A.: Source inventory of Bay Area greenhouse gas emissions, Tech. rep., Bay Area Air Quality Management District, 2010.

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., Wofsy, S. C., Nehrkorn, T., Eluszkiewicz, J., Ehleringer, J. R., and Stephens, B. B.: Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region, Proc Natl Acad Sci U S A, 109, 8423–8, doi:10.1073/pnas.1116645109, 2012.

Meirink, J. F., Bergamaschi, P., and Krol, M. C.: Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion, Atmospheric Chemistry and Physics, 8, 6341–6353, doi:10.5194/acp-8-6341-2008, 2008.

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 CO_2 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 forecastingstochastic time-inverted lagrangian transport (WRF–STILT) model, Meteorology and Atmospheric Physics, 107, 51–64, doi:10.1007/s00703-010-0068-x, 2010.

Newman, S., Jeong, S., Fischer, M. L., Xu, X., Haman, C. L., Lefer, B., Alvarez, S., Rappenglueck, B., Kort, E. A., Andrews, A. E., Peischl, J., Gurney, K. R., Miller, C. E., and Yung, Y. L.: Diurnal tracking of anthropogenic CO_2 emissions in the Los Angeles basin megacity during spring 2010, Atmospheric Chemistry and Physics, 13, 4359–4372, doi:10.5194/acp-13-4359-2013, 2013.

Oda, T. and Maksyutov, S.: A very high-resolution (1 km$\times$1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights, Atmospheric Chemistry and Physics, 11, 543–556, doi:10.5194/acp-11-543-2011, 2011.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M., Petron, G., Hirsch, A. I., Worthy, D. E., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proc Natl Acad Sci U S A, 104, 18 925–30, doi:10.1073/pnas.0708986104, 2007.

Rigby, M., Toumi, R., Fisher, R., Lowry, D., and Nisbet, E. G.: First continuous measurements of CO_2 mixing ratio in central London using a compact diffusion probe, Atmospheric Environment, 42, 8943–8953, doi:10.1016/j.atmosenv.2008.06.040, 2008.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, World Scientific, Singapore, 2000.

Shusterman, A. A., Teige, V., Turner, A. J., Newman, C., Kim, J., and Cohen, R. C.: The BErkeley Atmospheric CO_2 Observation Network: initial evaluation, Atmospheric Chemistry and Physics Discussions, pp. 1–23, doi:10.5194/acp-2016-530, 2016.

Singh, K., Jardak, M., Sandu, A., Bowman, K., Lee, M., and Jones, D.: Construction of non-diagonal background error covariance matrices for global chemical data assimilation, Geoscientific Model Development, 4, 299–316, doi:10.5194/gmd-4-299-2011, 2011.

Turnbull, J. C., Sweeney, C., Karion, A., Newberger, T., Lehman, S. J., Tans, P. P., Davis, K. J., Lauvaux, T., Miles, N. L., Richardson, S. J., Cambaliza, M. O., Shepson, P. B., Gurney, K., Patarasuk, R., and Razlivanov, I.: Toward quantification and source sector identification of fossil fuel CO_2 emissions from an urban area: Results from the INFLUX experiment, Journal of Geophysical Research-Atmospheres, 120, 292–312, doi: 10.1002/2014JD022555, 2015.

Wu, L., Broquet, G., Ciais, P., Bellassen, V., Vogel, F., Chevallier, F., Xueref-Remy, I., and Wang, Y.: What would dense atmospheric observation networks bring to the quantification of city CO_2 emissions?, Atmospheric Chemistry and Physics, 16, 7743–7771, doi:10.5194/acp-16-7743-2016, 2016.

Yadav, V. and Michalak, A. M.: Improving computational efficiency in large linear inverse problems: an example from carbon dioxide flux estimation, Geoscientific Model Development, 6, 583–590, doi: 10.5194/gmd-6-583-2013, 2013.

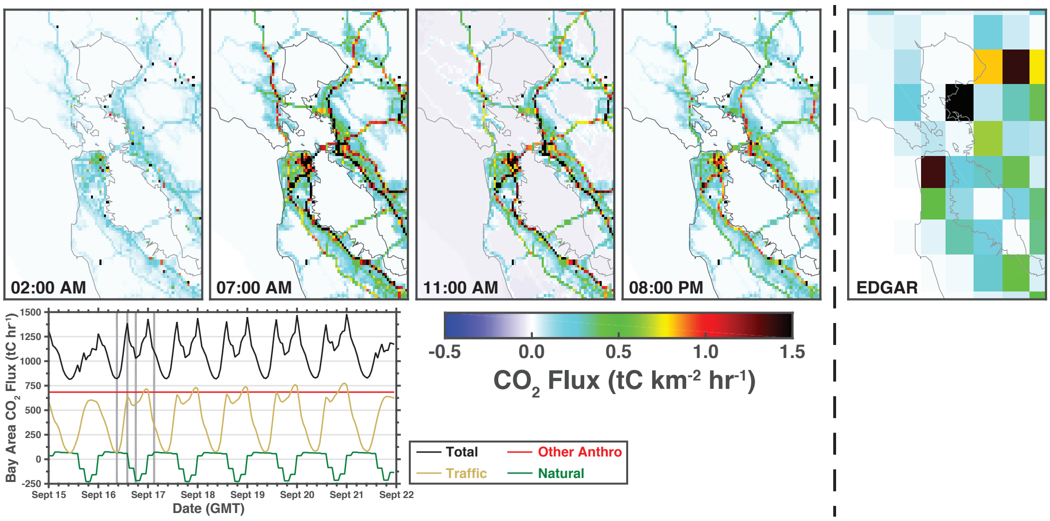

Fig. 1. September 2013 CO_2 fluxes from bottom-up inventories. Top row shows the fluxes in the Bay Area (122.0357° – 122.7683°W, 37.3771° – 38.2218°N) at four representative hours (hour in local time). Right panel shows the a-temporal EDGAR v4.2 FT2010 CO_2 flux in the Bay Area. Bottom panel shows the total Bay Area CO_2 flux (black), traffic (orange), other anthropogenic (red), and natural (green) sources. Vertical gray shading indicates the time slices plotted in the top and middle panels.

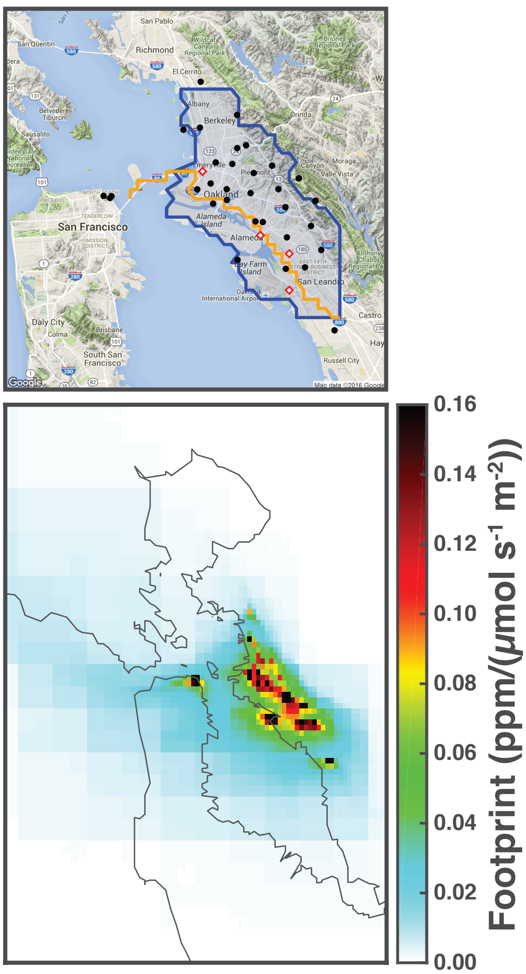

Fig. 2. Top panel shows the location of the sites (black circles), the area source (blue region), the line source (orange line), and point sources (red diamonds). Bottom panel shows the September 15 to 22 average footprint for the 34 sites in the network, see Table 1 for a list of the sites. The bottom panel is the full domain used for the inversion. Supplemental Fig. S3 shows the footprint on a log-scale.

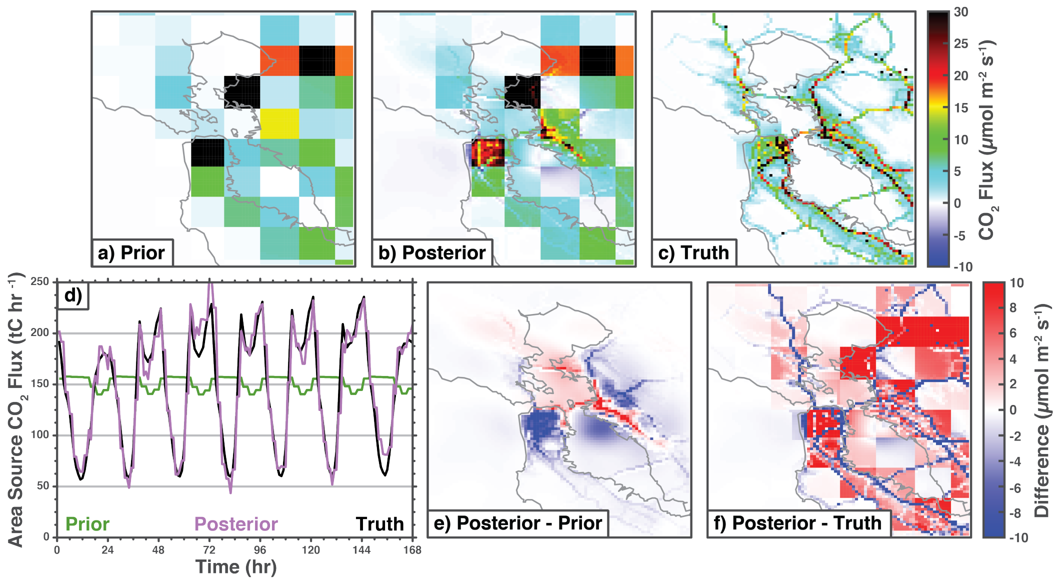

Fig. 3. Example of estimated CO_2 fluxes. Top row shows the average emissions from **(a)** the prior, **(b)** the posterior, and **(c)** the true emissions. Panel **(d)** shows a time series of the emissions from the area source with the prior (green), posterior (pink), and true emissions (black). Panel **(e)** shows the difference between the posterior and the prior. Panel **(f)** shows the difference between posterior and the truth. Posterior output is from the best case scenario ($n_S = 34$ and $\sigma_m = 0.005$ ppm).

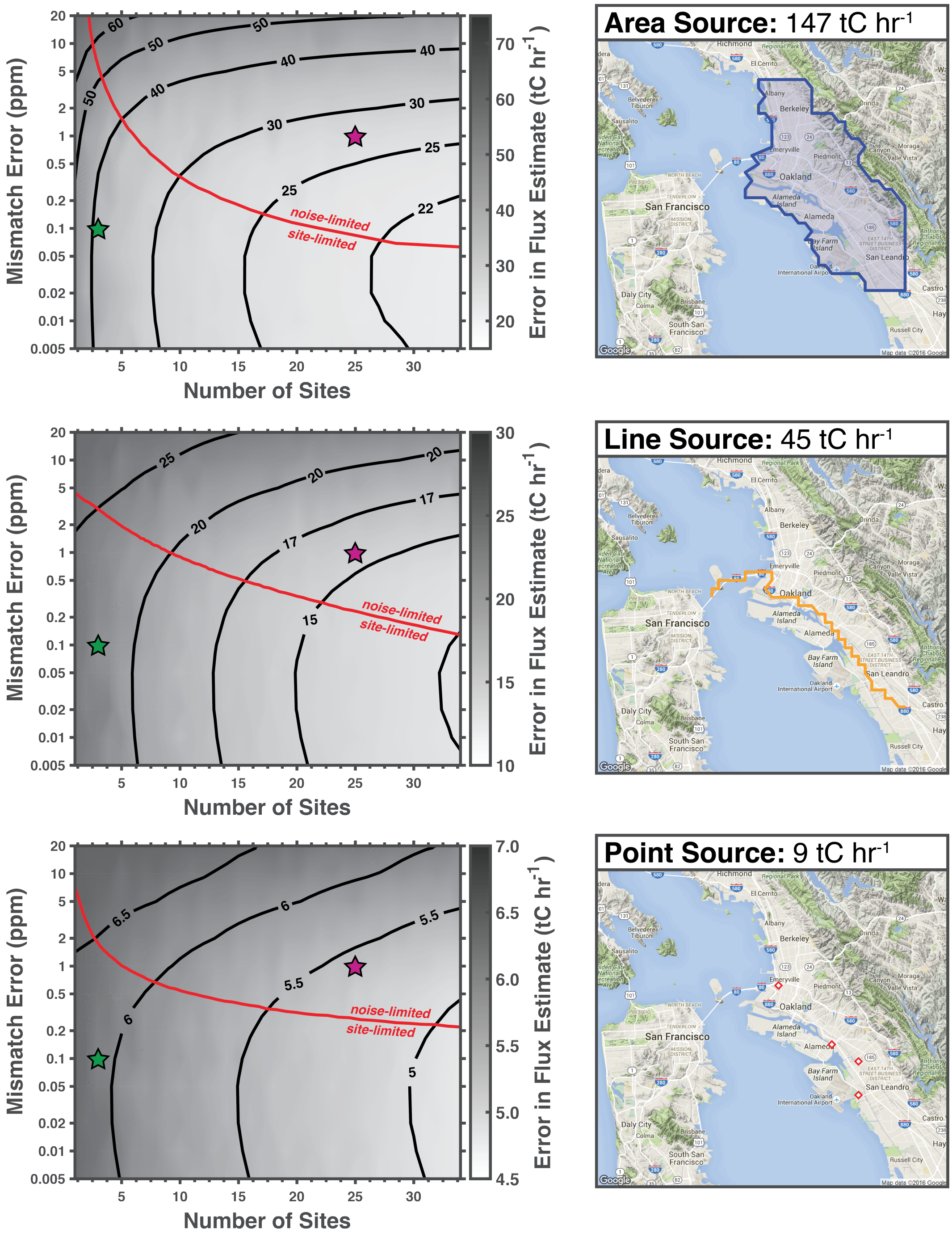

Fig. 4. Left column shows the error in the posterior CO_2 fluxes. Right column shows the fluxes being estimated. Top row is the area source, middle row is the line source, and bottom row is the point source. Inversions were performed using $n_s = [1, 2, \ldots, 34]$ sites and $\sigma_m = [0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 10, 20]$ ppm mismatch error. Results shown are the mean of a monte carlo analysis using 20 different combinations of sites for each n_s, σ_m pair. Contours are from the statistical models $\hat{\eta}$ (see Eq. 5–7) converted to flux errors and the red lines are the ridge lines that define the cutoff between the noise-limited and site-limited regimes. Purple star shows an observing system with 25 sites and 1 ppm noise. Green star shows an observing system with 3 sites and 0.1 ppm noise. Note the log-scale on the y-axis.

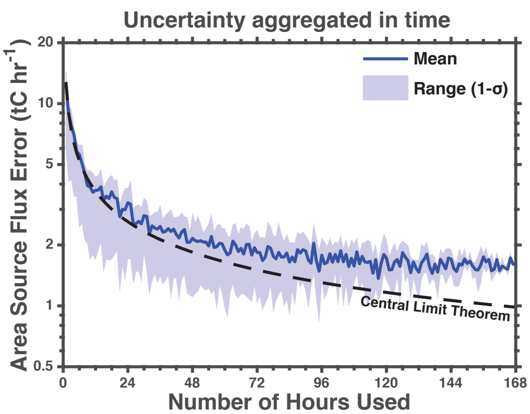

Fig. 5. Uncertainty aggregated in time for the best case inversion (see Fig. 3). The CO_2 flux estimate in this study has an hourly temporal resolution. The uncertainty in the emissions estimate declines as the estimate is averaged to longer temporal scales. Solid blue line is the mean uncertainty, shading is the 1-σ range, and the dashed black line is the uncertainty predicted by the central limit theorem. Note the log scale on the y-axis.

Table 1. 34 sites in the network[a] used in this study.

Site Code	Site name	Latitude (°N)	Longitude (°W)	Height (m a.g.l.)
AHS	Arroyo High School	37.680	122.139	3
BEL	Burckhalter Elementary School	37.775	122.167	5
BFE	Bayfarm Elementary School	37.744	122.251	3
BOD	Bishop O'Dowd High School	37.753	122.155	3
CES	Claremont Elementary School	37.846	122.252	3
CHA	Chabot Space & Science Center (low)	37.819	122.181	3
CHB	Chabot Space & Science Center (high)	37.819	122.181	9
COI	Coit Tower	37.8030	122.406	5
CPS	College Preparatory School	37.849	122.242	24
EBM	W. Oakland EBMUD Monitoring Station	37.814	122.282	3
ELC	El Cerrito High School	37.907	122.294	8
EXB	Exploratorium (Bay)	37.803	122.397	6
EXE	Exploratorium (Embarcadero)	37.801	122.399	3
FTK	Fred T. Korematsu Discovery Academy	37.738	122.174	3
GLE	Greenleaf Elementary School	37.765	122.194	3
HRS	Head Royce School	37.809	122.204	7
ICS	International Community School	37.779	122.231	3
KAI	Kaiser Center	37.809	122.264	127
LAU	Laurel Elementary School	37.792	122.197	12
LBL	Lawrence Berkeley National Lab, Bldg. 70	37.876	122.252	3
LCC	Lighthouse Community Charter School	37.736	122.196	3
MAR	Berkeley Marina	37.863	122.314	3
MON	Montclair Elementary School	37.830	122.212	3
NOC	N. Oakland Community Charter School	37.833	122.277	3
OMC	Oakland Museum of California	37.799	122.264	3
PAP	PLACE at Prescott Elementary	37.809	122.298	3
PDS	Park Day School	37.832	122.257	3
PHS	Piedmont Middle & High School	37.824	122.233	3
POR	Port of Oakland Headquarters	37.796	122.280	3
OHS	Oakland High School	37.805	122.236	3
ROS	Rosa Parks Elementary School	37.865	122.295	3
SHA	Skyline High School (low)	37.798	122.162	3
SHB	Skyline High School (high)	37.798	122.162	13
STL	St. Elizabeth High School	37.779	122.222	3

[a] This study uses both operational and proposed sites. See Shusterman et al. (2016) and "http://beacon.berkeley.edu/" for more information on the network.