

Response to Reviewer Comments:

We thank the Anonymous Reviewer #2 for their additional comments.

Reviewer #2 Comments:

1.) Footprints: My concern about the footprints is definitely not anything related to the plotting scales, but about the low sensitivity of measurements to the changes in emission fluxes on the western side of the model domain. Authors do agree that there are large footprint values for locations where there is an observation site. But the next statement in the authors' response is not very clear in its present form. i.e. how is it (more specifically what is plotted in Fig. 2) related to the diffuse signals? Further examination of the transport mechanism for this case would have been beneficial to explain these spatial patterns, as I stated in the former reviewer comments. Since it points to the representativeness of the observations (i.e. what is "seen" by these sites), I recommend authors to add necessary statements in the manuscript to support/justify the spatial patterns of the footprints presented here. I would think that it is rather easy to include it in the manuscript as it does not alter any conclusions made.

We have updated the text to include a reference to another paper that found similar spatial sensitivities in California's Bay Area.

Lines 141-143: "The spatial extent of the footprints found here are similar to those found in Bastien *et al.* (2015), who performed an adjoint-based sensitivity analysis of urban air pollution in the San Francisco Bay area (see their Fig. 2)."

2.) Reported error estimate in the abstract: I do still think that it is important to be specific in terms of uncertainties. Authors know that there are other important uncertainties that affect the retrieved flux accuracy and these uncertainties are ignored in the current estimation. Hence I suggest authors to state the value(s) of combined observation and model uncertainty which resulted in the 5% posterior error.

We have updated the abstract to include the reviewer's suggestion.

Lines 19-22: "The dense network considered here (modeled after the BEACO₂N network **with an assumed mismatch error of 1 ppm at hourly temporal resolution**) could estimate weekly CO₂ emissions from an urban region with less than 5% error"

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

A. J. Turner's manuscript prepared
with version 4.2 of the L^AT_EX class copernicus.cls.
Date: 7 October 2016

Network design for quantifying urban CO₂ emissions: Assessing trade-offs between precision and network density

Alexander J. Turner^{1,2}, Alexis A. Shusterman³, Brian C. McDonald^{4,*},
Virginia Teige³, Robert A. Harley⁴, and Ronald C. Cohen^{3,5}

¹School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA.

²Environmental Energy and Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

³Department of Chemistry, University of California at Berkeley, Berkeley, CA, USA.

⁴Department of Civil and Engineering, University of California at Berkeley, Berkeley, CA, USA.

⁵Department 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)

1 **Abstract.** The majority of anthropogenic CO₂ emissions are attributable to urban areas. While
2 the emissions from urban electricity generation often occur in locations remote from consumption,
3 many of the other emissions occur within the city limits. Evaluating the effectiveness of strate-
4 gies for controlling these emissions depends on our ability to observe urban CO₂ emissions and
5 attribute them to specific activities. Cost effective strategies for doing so have yet to be described.
6 Here we characterize the ability of a prototype measurement network, modeled after the BEACO₂N
7 network in California's Bay Area, in combination with an inverse model based on WRF-STILT to
8 improve our understanding of urban emissions. The pseudo-measurement network includes 34 sites
9 at roughly 2 km spacing covering an area of roughly 400 km². The model uses an hourly 1 × 1 km²
10 emission inventory and 1 × 1 km² meteorological calculations. We perform an ensemble of Bayesian
11 atmospheric inversions to sample the combined effects of uncertainties of the pseudo-measurements
12 and the model. We vary the estimates of the combined uncertainty of the pseudo-observations and
13 model over a range of 20 ppm to 0.005 ppm and vary the number of sites from 1 to 34. We use these
14 inversions to develop statistical models that estimate the efficacy of the combined model-observing
15 system at reducing uncertainty in CO₂ emissions. We examine uncertainty in estimated CO₂ fluxes
16 at the urban scale, as well as for sources embedded within the city such as a line source (e.g., a
17 highway) or a point source (e.g., emissions from the stacks of small industrial facilities). Using our
18 inversion framework, we find that a dense network with moderate precision is the preferred setup
19 for estimating area, line, and point sources from a combined uncertainty and cost perspective. The

20 dense network considered here (modeled after the BEACO₂N network [with an assumed mismatch](#)
21 [error of 1 ppm at hourly temporal resolution](#)) could estimate weekly CO₂ emissions from an urban
22 region with less than 5% error, given our characterization of the combined observation and model
23 uncertainty.

24 **1 Introduction**

25 Carbon dioxide (CO₂) is an atmospheric trace gas and the single largest anthropogenic radiative
26 forcer, with a radiative forcing of 1.82 W m⁻² in 2011 relative to preindustrial times (IPCC, 2013).
27 CO₂ has increased from 280 ppm in preindustrial times to greater than 400 ppm in the present,
28 largely due to changes in fossil fuel emissions. Over 70% of these fossil fuel CO₂ emissions in
29 the United States (US) are attributable to urban areas (EIA, 2015; Hutyra et al., 2014), yet cur-
30 rent bottom-up inventories still have large uncertainties. As such, quantifying and monitoring the
31 emissions from urban areas is crucial to strategies for reducing future increases in CO₂.

32 Numerous studies have performed top-down estimations of CO₂ emissions using observations
33 from urban surface monitoring networks of various sizes (e.g., Gratani and Varone, 2005; McKain
34 et al., 2012; Newman et al., 2013; Lauvaux et al., 2013; Breon et al., 2015; Turnbull et al., 2015).
35 However, it's not immediately clear how many sites are necessary to monitor the emissions from an
36 urban area. Kort et al. (2013) found that a surface monitoring network would need at least 8 sites
37 operating for 8 weeks to accurately estimate CO₂ emissions in Los Angeles. Yet most current urban
38 monitoring networks have fewer than 8 sites but operate for much longer than 8 weeks. For example,
39 Gratani and Varone (2005) used a single site in Rome, Newman et al. (2013) used a single site in
40 Los Angeles, Lauvaux et al. (2013) used two sites in Davos, Switzerland, McKain et al. (2012) used
41 a network of 5 sites in Salt Lake City, and Breon et al. (2015) used 5 sites in Paris. Recent work
42 from Turnbull et al. (2015) employed a denser network of 12 sites in Indianapolis.

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

50 The tradeoff between measurement network density and mismatch error has yet to be charac-
51 terized. Understanding these tradeoffs is crucial to reducing the uncertainty in emissions from ur-
52 ban regions and to developing cost-effective urban monitoring networks. Here we present a high-
53 resolution inventory of CO₂ fluxes and a numerical model that relates atmospheric observations to
54 high resolution surface fluxes. We then use this inventory and model in a series of observing system

55 simulation experiments (OSSEs) to investigate the tradeoff between reductions in the mismatch error
56 and increases in the measurement network density. We develop statistical models to characterize this
57 relationship for different types of sources in the San Francisco Bay Area, identify limiting regimes,
58 and recommend future observing strategies.

59 **2 Constructing a high resolution regional CO₂ inventory**

60 McDonald et al. (2014) demonstrated that 1×1 km² spatial resolution is necessary to resolve the
61 gradients in urban CO₂ fluxes from highways. However, most of the existing CO₂ anthropogenic
62 inventories are not available at this resolution. For example, EDGAR (European Commission, 2011)
63 and VULCAN (Gurney et al., 2009) are only available at $0.1^\circ \times 0.1^\circ$ and 10×10 km², respectively.
64 A notable exception is the Odiac fossil fuel CO₂ inventory (Oda and Maksyutov, 2011) which is
65 based on satellite-observed nightlight data and available globally at 1×1 km² resolution. High reso-
66 lution fossil fuel CO₂ emissions are available for select cities and sectors such as Paris through the
67 AirParif inventory (Breon et al., 2015, <http://www.airparif.asso.fr/en/index/index>) and Indianapo-
68 lis, Los Angeles, Salt Lake City, and Phoenix through the HESTIA project (Gurney et al., 2012,
69 <http://hestia.project.asu.edu/>); three recent studies (Gately et al., 2013; McDonald et al., 2014; Gately
70 et al., 2015) developed high resolution CO₂ emissions from vehicular traffic.

71 The Bay Area Air Quality Management District (BAAQMD) provides detailed annual county-
72 level CO₂ emissions information for San Francisco and California’s Bay Area (Mangat et al., 2010).
73 The BAAQMD found that the transportation sector accounted for 36% of the Bay Area anthro-
74 pogenic emissions, industrial and commercial for 36%, electricity for 16%, residential fuel usage
75 for 7%, off-road equipment for 3.0%, and agriculture for 1%. The BAAQMD also reports CO₂
76 emissions for 4,375 point sources in the Bay Area. We geocode these point sources based on the
77 addresses provided by the BAAQMD. These point sources capture the emissions from the indus-
78 trial, commercial, and electricity sectors. We map residential fuel usage to population using block
79 level population data from the 2010 US Census and apply a temporal temperature scaling based on
80 Deschênes and Greenstone (2011); the resulting temporal scaling effect is small due to the temperate
81 climate in the East Bay region of the SF Bay Area.

82 Here we use the traffic CO₂ emissions from the fuel-based inventory for vehicle emissions (FIVE)
83 developed by McDonald et al. (2014). The FIVE traffic CO₂ inventory provides a representative
84 week of hourly CO₂ emissions for San Francisco and other nearby Bay Area cities at 10 km, 4
85 km, 1 km, and 500 m resolution. This representative week can be scaled to different years based
86 on the state fuel sales (see McDonald et al. (2014) for additional details). The FIVE inventory is
87 constructed by partitioning CO₂ emissions using state-level fuel data to individual roads with road-
88 specific traffic count data and temporal patterns from weigh-in-motion data. In this manner, CO₂
89 emissions from the FIVE inventory will be consistent with state and national CO₂ budgets and can

90 easily be scaled to different years.

91 Combining the industrial, commercial, electricity, residential, and traffic emissions account for
92 95.8% of the anthropogenic CO₂ emissions in the Bay Area. We do not have high resolution proxy
93 data for the off-road equipment or agriculture sectors in the Bay Area and have chosen to assume
94 their contributions are smaller than the uncertainty in the total budget; therefore we neglect these
95 sectors in the construction of our inventory.

96 CarbonTracker CT2013B (<http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/>; Peters et al., 2007)
97 provides 3 hourly fossil fuel, ocean, biogenic, and fire CO₂ fluxes at 1° × 1° resolution. These fluxes
98 are optimized to agree with atmospheric CO₂ observations. We regrid these fluxes to 1 × 1 km²
99 spatial resolution (see Supplemental Section S3) and use the fire, ocean, and biogenic sectors to
100 account for our natural fluxes.

101 Fig. 1 shows snapshots of the CO₂ fluxes from our inventory at 4 different times of day and
102 the a-temporal fluxes from EDGAR v4.2 FT2010 (European Commission, 2011). From Fig. 1 we
103 can see the inventory clearly resolves the large CO₂ gradients from highways, confirming that 1 × 1
104 km² spatial resolution is sufficient to resolve urban CO₂ fluxes from highways. The bottom panel
105 of Fig. 1 shows a time series of Bay Area CO₂ fluxes broken down by source. The diurnal cycle
106 in our inventory is largely driven by the traffic emissions with modest uptake from the biosphere
107 during the middle of the day. Other anthropogenic sources were assumed to have a negligible diurnal
108 cycle (Nassar et al., 2013). In what follows, we use EDGAR as the prior and the high spatio-temporal
109 resolution inventory as the “truth”.

110 [Fig. 1 about here.]

111 3 The Berkeley Atmospheric CO₂ Observation Network (BEACO₂N)

112 The Berkeley Atmospheric CO₂ Observation Network (“BEACO₂N”, see <http://beacon.berkeley.edu>
113 and Shusterman et al., 2016) was founded in 2012 as a web of approximately 25 carbon dioxide
114 sensing “nodes” stationed atop schools and museums in the Oakland, CA metropolitan area (see
115 Table 1). With sensors installed on an approximately 2 km square grid, BEACO₂N is the only
116 surface-level (3 to 130 m a.g.l.) greenhouse gas monitoring system with roughly the same spatial
117 resolution as the emissions inventories described above. Each node requires only a standard, 120V
118 power source and is sited on pre-existing structures based on voluntary, no-cost partnerships. The
119 BEACO₂N configuration therefore represents a reasonable expectation and is one model for future
120 monitoring networks aimed at constraining CO₂ fluxes at neighborhood scales within an urban dome.

121 [Table 1 about here.]

122 BEACO₂N’s unprecedented spatial density is achieved by exploiting lower cost instrumentation
123 than has traditionally been utilized for ambient CO₂ detection. The non-dispersive infrared (NDIR)

124 absorption sensor used in each BEACO₂N node (<http://www.vaisala.com/en/products/carbondioxide/Pages/GMP343.aspx>)
 125 has been seen to possess adequate sensitivity to resolve diurnal as well as seasonal phenomena rele-
 126 vant to urban environments (Rigby et al., 2008) and costs one to two orders of magnitude less than
 127 the commercial cavity ring-down instruments commonly used in other networks. However, the low-
 128 cost NDIR sensor is more susceptible to factors such as temporal drift and environmental instability
 129 that can negatively impact data quality. This trade-off between mismatch error and network density
 130 is explored below.

131 4 Observing system simulation experiments

132 CO₂ concentrations were simulated at 34 sites in the BEACO₂N network with the Stochastic Time-
 133 Inverted Lagrangian Transport (STILT) model (Lin et al., 2003), coupled to the Weather Research
 134 and Forecasting (WRF) meso-scale meteorological model run at 1 × 1 km² grid resolution (WRF-
 135 STILT; Nehrkorn et al., 2010). WRF-STILT computes footprints (Δ CO₂ per surface flux, or ppm
 136 per μmol·m⁻²·s⁻¹; See Supplemental Section S1 and Lin et al. (2003) for additional details) for
 137 each observation that relate the hourly 1 km² CO₂ fluxes (\mathbf{x} ; an $m \times 1$ vector) to the observations
 138 (\mathbf{y} ; an $n \times 1$ vector):

$$139 \mathbf{y} = \mathbf{H}\mathbf{x} \tag{1}$$

140 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
 141 location of the sites and the average network footprint for Sept 15 to 22. [The spatial extent of](#)
 142 [the footprints found here are similar to those found in Bastien et al. \(2015\), who performed an](#)
 143 [adjoint-based sensitivity analysis of urban air pollution in the San Francisco Bay area \(see their](#)
 144 [Fig. 2\).](#)

145 [Fig. 2 about here.]

146 Our aim is to estimate hourly CO₂ fluxes at 1 km² over a one week period. As such, the model
 147 domain is 88 km × 101 km and we solve for 240 hours of fluxes (1 week plus 3 additional days
 148 of back trajectories). The resulting state vector has 2,133,120 elements ($m = m_t \cdot m_x \cdot m_y$ with
 149 $m_t = 240$, $m_x = 88$, $m_y = 101$) and the posterior fluxes will have hourly temporal resolution and 1
 150 km² spatial resolution. The dimension of n will depend on the number of sites in the observational
 151 network.

152 Here we use our high resolution CO₂ inventory (\mathbf{x}^* ; an $m \times 1$ vector) to generate synthetic obser-
 153 vations (\mathbf{y}^* ; an $n \times 1$ vector):

$$154 \mathbf{y}^* = \mathbf{H}\mathbf{x}^* + \boldsymbol{\varepsilon} \tag{2}$$

155 where $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of normally distributed noise with mean $\boldsymbol{\epsilon}_b$ and diagonal covariance ma-
 156 trix \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

157 are uncorrelated. Our base case inversion assumes the mean bias is zero: $\epsilon_b = \mathbf{0}$. We evaluate the
 158 sensitivity to this assumption in Section 6 and Supplemental Section S6.2. These synthetic observa-
 159 tions can then be used in a Bayesian inference framework to estimate the optimal CO₂ fluxes (c.f.
 160 Rodgers, 2000). Assuming the prior and likelihood distributions are Gaussian gives us a closed-form
 161 solution for the posterior CO₂ fluxes:

$$162 \hat{\mathbf{x}} = \mathbf{x}_p + (\mathbf{HB})^T (\mathbf{HBH}^T + \mathbf{R})^{-1} (\mathbf{y}^* - \mathbf{Hx}_p) \quad (3)$$

163 where \mathbf{x}_p is an $m \times 1$ vector of prior CO₂ fluxes, comprised of a coarse (10×10 km²) a-temporal
 164 EDGAR v4.2 FT2010 anthropogenic CO₂ inventory and natural fluxes from CarbonTracker CT2013B,
 165 regridded to 1×1 km². \mathbf{B} is the $m \times m$ prior error covariance matrix. The prior error covariance
 166 matrix can be expressed as a Kronecker product (cf. Meirink et al., 2008; Singh et al., 2011; Yadav
 167 and Michalak, 2013) of temporal and spatial covariance matrices: $\mathbf{B} = \mathbf{D} \otimes \mathbf{E}$ where \mathbf{D} is the tem-
 168 poral covariance matrix and \mathbf{E} is the spatial covariance matrix. The \mathbf{B} matrix has an uncertainty of
 169 100% at the native resolution and the spatial and temporal covariance matrices are fully populated
 170 (see Supplemental Section S2 for more details).

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

175 Fig. 3 shows an example of the estimated CO₂ fluxes. We can see that the posterior fluxes cap-
 176 ture more of the spatial variability in the CO₂ fluxes than the prior fluxes in the region where the
 177 network is deployed. We find substantial improvements in the diurnal cycle (see panel d). Previ-
 178 ous work has used the posterior covariance matrix ($\mathbf{Q} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1})^{-1}$), averaging kernel
 179 matrix ($\mathbf{A} = \mathbf{I} - \mathbf{QB}^{-1}$), and the degrees of freedom for signal (DOFs = $\text{tr}(\mathbf{A})$) as metrics to eval-
 180 uate the information content of different observing systems (e.g., Kort et al., 2013; Wu et al., 2016).
 181 However, it is computationally infeasible to construct these $m \times m$ matrices for our application as
 182 $m > 10^6$ and storing them would require ~ 36 Tb of memory (assuming double precision, dense
 183 matrices).

184 [Fig. 3 about here.]

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

$$188 \eta = 1 - \frac{\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2}{\|\mathbf{x}_p - \mathbf{x}^*\|_2} \quad (4)$$

189 This error metric, η , was chosen as it has a similar form to the averaging kernel matrix but it also
 190 allows us to directly compare the posterior fluxes to the true fluxes. This relative error metric can
 191 be related to the flux error (see Supplemental Section S5). As such, we can use the error metric to

192 evaluate the ability of the observing system to resolve three types of emission sources: (1) area, (2)
 193 line, and (3) point sources, by examining a subset of grid cells in the domain (see Section S3 for
 194 more details). The area source (AS) examined here is the East Bay urban dome ($147 \pm 55 \text{ tC hr}^{-1}$;
 195 uncertainty is the $1\text{-}\sigma$ range of hourly fluxes from the high resolution inventory), the line source
 196 (LS) is Interstate 880 and the Bay Bridge ($45 \pm 20 \text{ tC hr}^{-1}$), and the point sources (PS) are 4 large
 197 CO_2 sources in the East Bay ($9 \pm 4 \text{ tC hr}^{-1}$). For comparison, Salt Lake City emits $\sim 300 \pm 50 \text{ tC}$
 198 hr^{-1} (McKain et al., 2012). The top panel of Fig. 2 shows these three source types.

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

207

[Fig. 4 about here.]

208 5 Simplified statistical models of error reduction

209 We develop statistical models to predict the error reduction and quantify the importance of the differ-
 210 ent factors governing the error reduction. We tested all combinations of models with the following
 211 7 parameters (127 possible combinations): $\sqrt{\sigma_m}$, $\sqrt{n_s}$, $\ln(\sigma_m)$, $\ln(n_s)$, σ_m , n_s , and a constant.
 212 These statistical models were evaluated using Akaike information criterion (AIC) and Bayesian in-
 213 formation criterion (BIC). The following statistical models were found to be best:

$$214 \hat{\eta}_{\text{AS}} = \beta_6 \sqrt{\sigma_m} + \beta_5 \sqrt{n_s} + \beta_4 \ln(\sigma_m) + \beta_3 \ln(n_s) + \beta_2 \sigma_m + \beta_0 \quad (5)$$

$$215 \hat{\eta}_{\text{LS}} = \beta_6 \sqrt{\sigma_m} + \beta_5 \sqrt{n_s} + \beta_4 \ln(\sigma_m) + \beta_3 \ln(n_s) + \beta_2 \sigma_m + \beta_1 n_s \quad (6)$$

$$216 \hat{\eta}_{\text{PS}} = \beta_6 \sqrt{\sigma_m} + \beta_5 \sqrt{n_s} + \beta_4 \ln(\sigma_m) + \beta_2 \sigma_m + \beta_0 \quad (7)$$

217 All the regression coefficients (β_i) in the statistical models yielded statistically significant ($p <$
 218 0.001) parameters based on F-tests (see the Supplemental Section S7 for the regression coefficients
 219 and model selection criterion).

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

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

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

239 6 Discussion

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

- 241 1. Achieving $\sigma_m = 1$ ppm adds value. There is relatively little additional benefit to reducing
242 mismatch error to 0.1 ppm, particularly for estimating line or point source emissions.
- 243 2. At $\sigma_m = 1$ ppm there is a benefit to increasing the number of sites, but this benefit increases
244 slower than $\sqrt{n_s}$.
- 245 3. At $\sigma_m = 5$ ppm there is little benefit from increasing the number of sites; reducing the noise
246 would add more value.

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

254 [Fig. 5 about here.]

255 6.1 Additional factors affecting observing system design

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

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

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

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

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

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

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

291 **6.2 Potential cost tradeoffs**

292 We consider two potential observing systems:

- 293 1. “Network A” ($n_s = 25$, $\sigma_m = 1$ ppm): A dense network with moderate-precision instruments.
294 This network is similar to the BEACO₂N network described in Section 3. We assume a cost
295 of \$5,000 per instrument giving a total cost of \$125,000. This network is shown as a purple
296 star in the left column of Fig. 4.
- 297 2. “Network B” ($n_s = 3$, $\sigma_m = 0.1$ ppm): A sparse network with of high-precision instruments.
298 This network uses cavity-ring down instruments. We assume a cost of \$50,000 per instrument
299 giving a total cost of \$150,000. This network is shown as a green star in the left column of
300 Fig. 4.

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

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

308 6.3 The relationship between network density and transport error

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

321 7 Conclusions

322 Understanding the factors that govern our ability to estimate urban greenhouse gas emissions are
323 crucial to improving an observing system and reducing the uncertainty in emission estimates. Here
324 we have quantitatively mapped the errors in CO₂ emission estimates from different observing sys-
325 tems for three different types of sources in California’s Bay Area: area sources, line sources, and
326 point sources. Our results show that different observing systems may fall into noise or site-limited

327 regimes where reducing the uncertainty in the estimated emissions is governed by a single factor;
328 these regimes differ for the source types. Identifying the regime an observing system is in will help
329 inform future improvements to the observing system. A number of prior urban CO₂ experiments
330 have defined as a goal, the understanding of emissions to less than 10% (e.g., Kort et al., 2013; Wu
331 et al., 2016). We find that a BEACO₂N-like network could achieve this accuracy and precision with
332 1 week of observations, if the dominant source of error is instrument precision. This conclusion may
333 motivate a re-examining of the conventional instrument quality-oriented design of CO₂ observing
334 systems, according to the stated goal of a given network.

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

342 References

- 343 Bastien, L. A., McDonald, B. C., Brown, N. J., and Harley, R. A.: High-resolution mapping of sources con-
344 tributing to urban air pollution using adjoint sensitivity analysis: benzene and diesel black carbon, *Environ*
345 *Sci Technol*, 49, 7276–84, doi:10.1021/acs.est.5b00686, 2015.
- 346 Breon, F. M., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Remy, I., Ramonet, M., Dieudonn, E.,
347 Lopez, M., Schmidt, M., Perrussel, O., and Ciais, P.: An attempt at estimating Paris area CO₂ emissions
348 from atmospheric concentration measurements, *Atmospheric Chemistry and Physics*, 15, 1707–1724, doi:
349 10.5194/acp-15-1707-2015, 2015.
- 350 Deschnes, O. and Greenstone, M.: Climate Change, Mortality, and Adaptation: Evidence from Annual
351 Fluctuations in Weather in the US, *American Economic Journal: Applied Economics*, 3, 152–185, doi:
352 10.1257/app.3.4.152, 2011.
- 353 EIA, U.: Emissions of Greenhouse Gases in the U.S., Tech. rep., U.S. Energy Information Administration,
354 2015.
- 355 European Commission: Emission Database for Global Atmospheric Research (EDGAR), release version 4.2,
356 Tech. rep., Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL), 2011.
- 357 Gately, C. K., Hutyra, L. R., Wing, I. S., and Brondfield, M. N.: A bottom up approach to on-road CO₂
358 emissions estimates: improved spatial accuracy and applications for regional planning, *Environ Sci Technol*,
359 47, 2423–30, doi:10.1021/es304238v, 2013.
- 360 Gately, C. K., Hutyra, L. R., and Sue Wing, I.: Cities, traffic, and CO₂: A multidecadal assessment of
361 trends, drivers, and scaling relationships, *Proc Natl Acad Sci U S A*, 112, 4999–5004, doi:10.1073/pnas.
362 1421723112, 2015.
- 363 Gratani, L. and Varone, L.: Daily and seasonal variation of CO₂ in the city of Rome in relationship with the
364 traffic volume, *Atmospheric Environment*, 39, 2619–2624, doi:10.1016/j.atmosenv.2005.01.013, 2005.
- 365 Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., and de la Rue du Can,
366 S.: High resolution fossil fuel combustion CO₂ emission fluxes for the United States, *Environ Sci Technol*,
367 43, 5535–41, 2009.
- 368 Gurney, K. R., Razlivanov, I., Song, Y., Zhou, Y., Benes, B., and Abdul-Massih, M.: Quantification of fossil
369 fuel CO₂ emissions on the building/street scale for a large U.S. city, *Environ Sci Technol*, 46, 12 194–202,
370 doi:10.1021/es3011282, 2012.
- 371 Hutyra, L. R., Duren, R., Gurney, K. R., Grimm, N., Kort, E. A., Larson, E., and Shrestha, G.: Urbanization
372 and the carbon cycle: Current capabilities and research outlook from the natural sciences perspective, *Earth's*
373 *Future*, 2, 473–495, doi:10.1002/2014ef000255, 2014.
- 374 IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth As-
375 sessment Report of the Intergovernmental Panel on Climate Change, Tech. rep., 2013.
- 376 Kort, E. A., Angevine, W. M., Duren, R., and Miller, C. E.: Surface observations for monitoring urban fossil fuel
377 CO₂ emissions: Minimum site location requirements for the Los Angeles megacity, *Journal of Geophysical*
378 *Research: Atmospheres*, 118, 1577–1584, doi:10.1002/jgrd.50135, 2013.
- 379 Lauvaux, T., Miles, N. L., Richardson, S. J., Deng, A., Stauffer, D. R., Davis, K. J., Jacobson, G., Rella, C.,
380 Calonder, G.-P., and DeCola, P. L.: Urban Emissions of CO₂ from Davos, Switzerland: The First Real-
381 Time Monitoring System Using an Atmospheric Inversion Technique, *Journal of Applied Meteorology and*

382 *Climatology*, 52, 2654–2668, doi:10.1175/jamc-d-13-038.1, 2013.

383 Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-
384 field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted
385 Lagrangian Transport (STILT) model, *Journal of Geophysical Research-Atmospheres*, 108, ACH 2–1–ACH
386 2–17, doi:10.1029/2002jd003161, 2003.

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

389 McDonald, B. C., McBride, Z. C., Martin, E. W., and Harley, R. A.: High-resolution mapping of motor vehicle
390 carbon dioxide emissions, *Journal of Geophysical Research-Atmospheres*, 119, 5283–5298, doi:10.1002/
391 2013jd021219, 2014.

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

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

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

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

404 Newman, S., Jeong, S., Fischer, M. L., Xu, X., Haman, C. L., Lefer, B., Alvarez, S., Rappenglueck, B., Kort,
405 E. A., Andrews, A. E., Peischl, J., Gurney, K. R., Miller, C. E., and Yung, Y. L.: Diurnal tracking of anthro-
406 pogenic CO₂ emissions in the Los Angeles basin megacity during spring 2010, *Atmospheric Chemistry and
407 Physics*, 13, 4359–4372, doi:10.5194/acp-13-4359-2013, 2013.

408 Oda, T. and Maksyutov, S.: A very high-resolution (1 km×1 km) global fossil fuel CO₂ emission inventory
409 derived using a point source database and satellite observations of nighttime lights, *Atmospheric Chemistry
410 and Physics*, 11, 543–556, doi:10.5194/acp-11-543-2011, 2011.

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

415 Rigby, M., Toumi, R., Fisher, R., Lowry, D., and Nisbet, E. G.: First continuous measurements of CO₂ mixing
416 ratio in central London using a compact diffusion probe, *Atmospheric Environment*, 42, 8943–8953, doi:
417 10.1016/j.atmosenv.2008.06.040, 2008.

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

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

422 Singh, K., Jardak, M., Sandu, A., Bowman, K., Lee, M., and Jones, D.: Construction of non-diagonal back-
423 ground error covariance matrices for global chemical data assimilation, *Geoscientific Model Development*,
424 4, 299–316, doi:10.5194/gmd-4-299-2011, 2011.

425 Turnbull, J. C., Sweeney, C., Karion, A., Newberger, T., Lehman, S. J., Tans, P. P., Davis, K. J., Lauvaux, T.,
426 Miles, N. L., Richardson, S. J., Cambaliza, M. O., Shepson, P. B., Gurney, K., Patarasuk, R., and Razlivanov,
427 I.: Toward quantification and source sector identification of fossil fuel CO₂ emissions from an urban area:
428 Results from the INFLUX experiment, *Journal of Geophysical Research-Atmospheres*, 120, 292–312, doi:
429 10.1002/2014JD022555, 2015.

430 Wu, L., Broquet, G., Ciais, P., Bellassen, V., Vogel, F., Chevallier, F., Xueref-Remy, I., and Wang, Y.: What
431 would dense atmospheric observation networks bring to the quantification of city CO₂ emissions?, *Atmo-
432 spheric Chemistry and Physics*, 16, 7743–7771, doi:10.5194/acp-16-7743-2016, 2016.

433 Yadav, V. and Michalak, A. M.: Improving computational efficiency in large linear inverse problems:
434 an example from carbon dioxide flux estimation, *Geoscientific Model Development*, 6, 583–590, doi:
435 10.5194/gmd-6-583-2013, 2013.

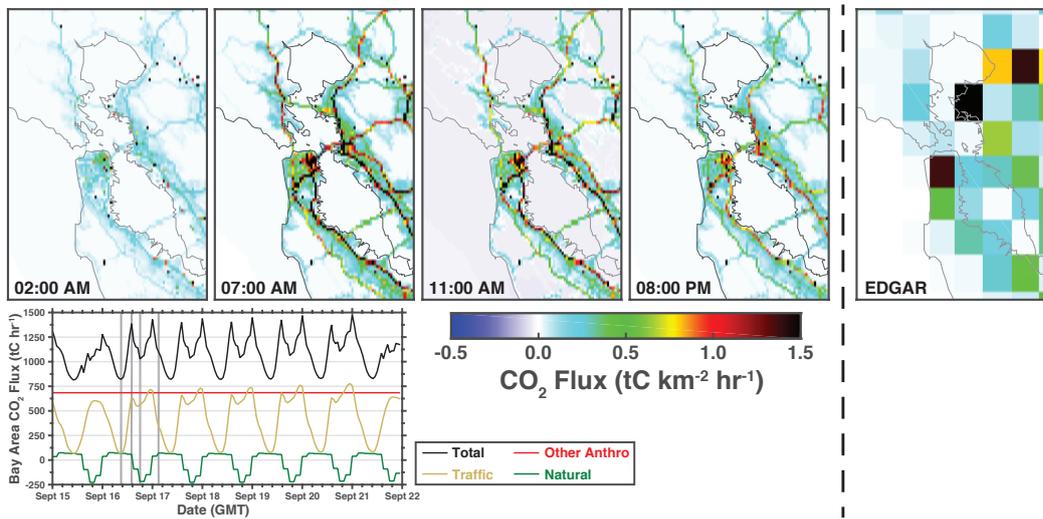


Fig. 1. September 2013 CO₂ fluxes from bottom-up inventories. Top row shows the fluxes in the Bay Area ($122.0357^{\circ} - 122.7683^{\circ}\text{W}$, $37.3771^{\circ} - 38.2218^{\circ}\text{N}$) at four representative hours (hour in local time). Right panel shows the a-temporal EDGAR v4.2 FT2010 CO₂ flux in the Bay Area. Bottom panel shows the total Bay Area CO₂ 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.

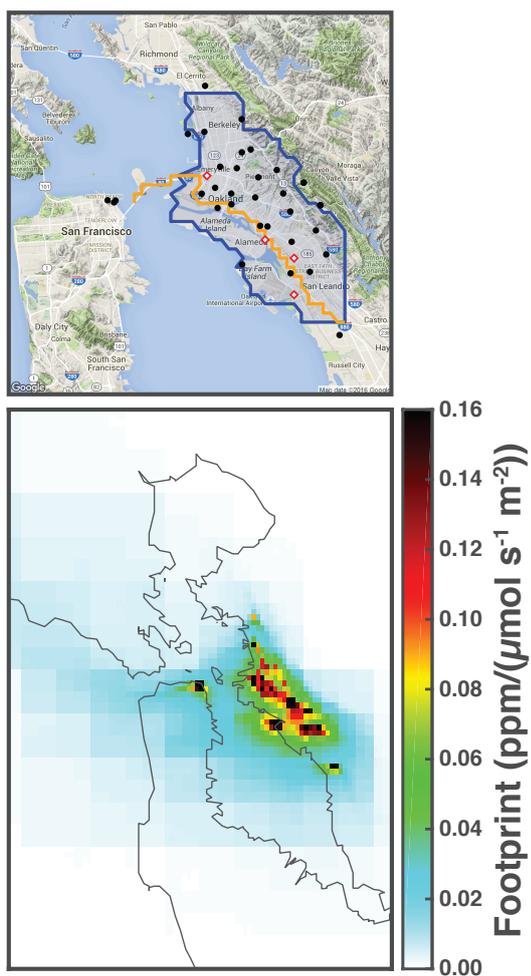


Fig. 2. Top panel shows the location of the sites (black circles), the area source (blue region), the line source (orange line), and point sources (red diamonds). Bottom panel shows the September 15 to 22 average footprint for the 34 sites in the network, see Table 1 for a list of the sites. The bottom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

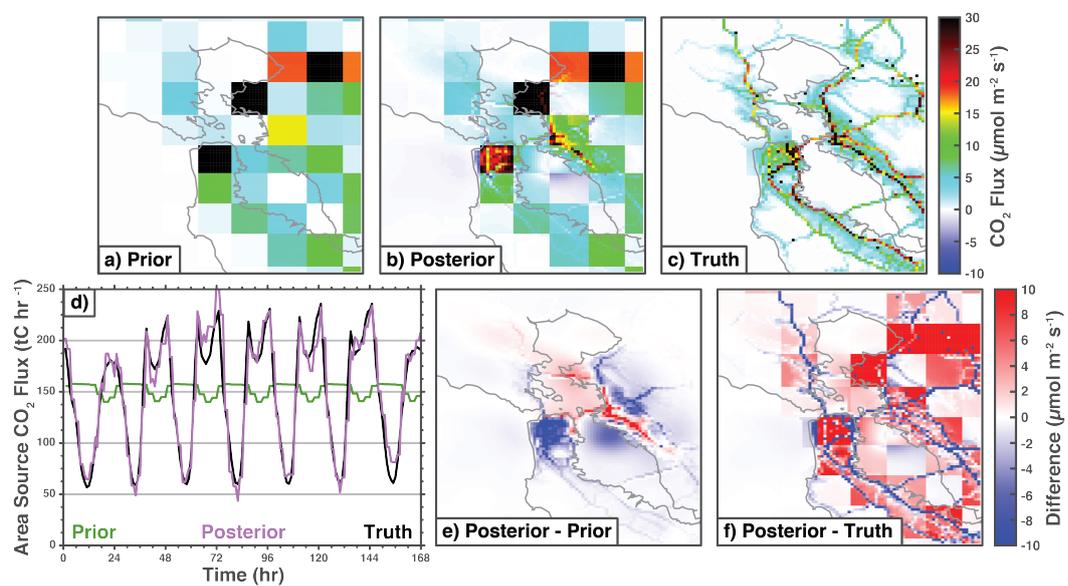


Fig. 3. Example of estimated CO₂ fluxes. Top row shows the average emissions from (a) the prior, (b) the posterior, and (c) the true emissions. Panel (d) shows a time series of the emissions from the area source with the prior (green), posterior (pink), and true emissions (black). Panel (e) shows the 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).

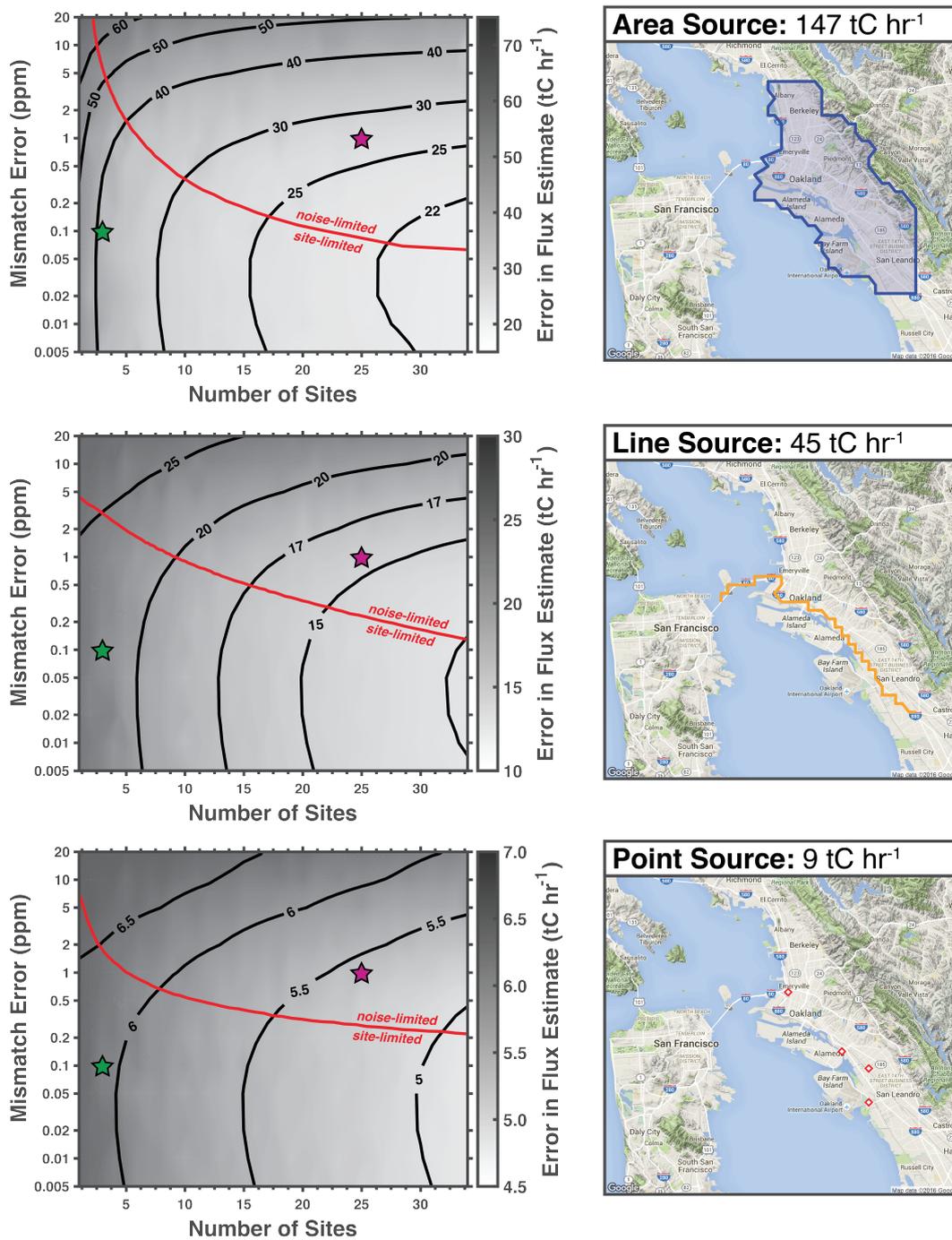


Fig. 4. Left column shows the error in the posterior CO₂ fluxes. Right column shows the fluxes being estimated. Top row is the area source, middle row is the line source, and bottom row is the point source. Inversions were performed using $n_s = [1, 2, \dots, 34]$ sites and $\sigma_m = [0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 10, 20]$ 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.

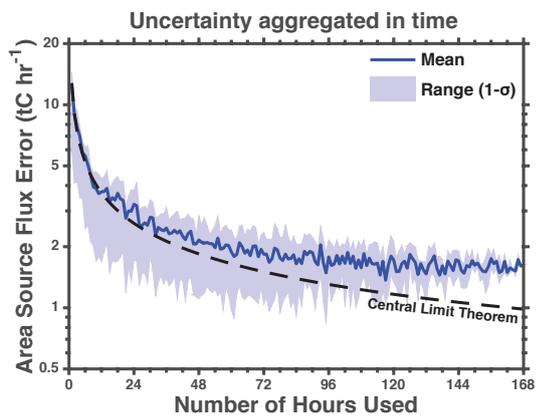


Fig. 5. Uncertainty aggregated in time for the best case inversion (see Fig. 3). The CO₂ flux estimate in this study has an hourly temporal resolution. The uncertainty in the emissions estimate declines as the estimate is averaged to longer temporal scales. Solid blue line is the mean uncertainty, shading 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.