Detecting Urban Emissions Changes and Events with a Near-Real-Time-Capable Inversion System

John Ware[1,2], Eric A. Kort[2], Riley Duren[3], Kimberly L Mueller[2,4], Kristal Verhulst[3], Vineet Yadav[3]

[1]Department of Physics, University of Michigan, Ann Arbor, Michigan, USA
[2]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, Michigan, USA
[3]NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
[4]National Institute of Standards and Technology (NIST); Gaithersburg, MD, United States of America

Key Points:

- LA CH4 flux estimates differ by driving meteorology but agree when calibrated for model sensitivity
- Aliso Canyon leak can be detected by inversions using operational meteorology
- Operational meteorology driven inversions significantly detect seasonal emission changes even with only one site

Corresponding author: John Ware, johnware@umich.edu

Abstract

In situ observing networks are increasingly being used to study greenhouse gas emissions in urban environments. While the need for sufficiently dense observations has often been discussed, density requirements depend on the question posed and interact with other choices made in the analysis. Focusing on the interaction of network density with varied meteorological information used to drive atmospheric transport, we perform geostatistical inversions of methane flux in the South Coast Air Basin, California in 2015-2016 using transport driven by a locally tuned Weather Research and Forecasting (WRF) configuration as well as by operationally-available meteorological products. We find total-basin flux estimates vary by as much as a factor of two between inversions, but the spread can be greatly reduced by calibrating the estimates to account for modeled sensitivity. Using observations from the full Los Angeles Megacities Carbon Project observing network, inversions driven by low-resolution generic wind fields are robustly sensitive ($p<0.05$) to seasonal differences in methane flux and to the increase in emissions caused by the 2015 Aliso Canyon natural gas leak. When the number of observing sites is reduced, the basinwide sensitivity degrades, but flux events can be detected by testing for changes in flux variance, and even a single site can robustly detect basin-wide seasonal flux variations. Overall, an urban monitoring system using an operational methane observing network and off-the-shelf meteorology could detect many seasonal or event-driven changes in near real time – and, if calibrated to a model chosen as a transfer standard, could also quantify absolute emissions.

1 Introduction

Recent years have seen increased efforts to quantify greenhouse gas emissions at or below the scale of individual cities. In complement to process-based inventories [*Gurney et al.*, 2012], aircraft campaigns [*Mays et al.*, 2009; *Wecht et al.*, 2014], and analysis of satellite data [*Kort et al.*, 2012; *Ye et al.*, 2017] among other methods, a common approach has been to deploy a network of sensors within and around a city [*McKain et al.*, 2012; *Breon et al.*, 2014; *McKain et al.*, 2015; *Richardson et al.*, 2016; *Shusterman et al.*, 2016; *Pugliese*, 2017; *Verhulst et al.*, 2017]. The density and placement of sensors within a network, together with the local meteorology and the spatiotemporal pattern of emissions, determines the extent to which the network is reliably sensitive to emissions over the whole region of interest and within the relevant time scale. Prospective network design studies [e.g., *Kort et al.*, 2013; *Turner et al.*, 2016; *Lopez-Coto et al.*, 2017] have attempted to ensure adequate sensitivity, but the standard of adequacy is necessarily relative to some particular purpose or question.

Much urban monitoring work focuses on improving the precision of absolute flux estimates, setting goals such as "to quantify CO_2 and CH_4 emission rates at 1 km^2 resolution with a 10% or better accuracy and precision" [*Davis et al.*, 2017]. Such precision may be a long way off or may not be achievable in every setting; however, a variety of other questions of interest can be answered without precisely constraining the absolute fluxes. For example: what seasonal variations and/or year-over-year trends exist in emissions rates, and what fraction of emissions can be attributed to the urban biosphere or to specific anthropogenic source sectors? An operational monitoring system might be able to detect an unusual excursion in the urban flux, and even to suggest a source location, even if the baseline flux is not known accurately.

In addition, network density interacts with a host of other factors that also impact the precision and confidence with which the above questions can be answered, including: representation of background concentrations and of the biosphere flux contribution, the statistical method to be used and the choices made in implementing that method (such as the specification of covariance parameters and the choice of a prior), and modeling of meteorology and of transport processes. This complex web of factors, and their interactions and contributions to the overall uncertainty in modeled posterior fluxes, are only beginning to be understood, especially in the urban setting. In this study, we focus on the meteorological driver of transport and how it impacts the inverse results. Future work should consider other factors, including the interaction of data density and driving meteorology with the choice of inversion methodology.

Representation of atmospheric transport is believed to be an important source of error in estimating GHG fluxes using atmospheric (in situ or column) observations [*McKain et al.*, 2012; *Feng et al.*, 2016]. However, there is no generally-adopted scheme for quantifying the effects of transport error. In inversions, some authors simply increase the model-data mismatch covariance across the board to account for transport error [e.g., *Breon et al.*, 2014]. *Lin and Gerbig* [2005] proposed using the increase in the variance of modeled concentrations when the observed error statistics of the wind components are incorporated as additional stochastic variability in the transport model. Recently, *Gourdji et al.* [2018]

showed that some of the effects of wind speed error could be mitigated by specifying an additional covariance proportional to the discrepancy in wind speed between model and observations.

Along with quantifying transport error, it is difficult to validate transport models or meteorological models in their role as drivers of transport in estimating fluxes for a particular question. On their own, meteorological models can be validated against point observations, most commonly of wind speed and direction and/or mixing depth. Validation of this kind is often used to tune model parameters or to choose a boundary-layer physics scheme or other model configuration [e.g., *Nehrkorn et al.*, 2013; *Feng et al.*, 2016], but does not directly address the fidelity of the transport or the impact on flux estimation. *Deng et al.* [2017] performed a semi-direct evaluation of coupled weather-transport models by comparing the marginal posterior likelihoods of the resulting CO_2 flux estimates. Direct validation of transport using controlled release of an inert tracer is also possible [e.g., *Harrison et al.*, 2012] but rarely included in urban studies.

In this study, rather than focus on the optimization of meteorological representation to achieve highest accuracy, highest resolution inversion results, we instead assess whether non-optimized, rapidly available meteorological products can successfully underpin an atmospheric inversion system. We focus on questions of whether such a system can detect anomalous high emissions events, and whether seasonal flux behaviors can be robustly inferred. If a rapidly available meteorological product can successfully underpin such a system, this indicates near-real time inversions driven by such a product could be conducted and expected to produce statistically useful results in near-real time.

To pursue such an approach, we consider Los Angeles as an ideal test case. California has had extensive study and validation of transport models [*Angevine et al.*, 2012, 2013; *Zhao et al.*, 2009; *Bagley et al.*, 2017]. A statewide assessment of transport is summarized in *Bagley et al.* [2017], and a regional assessment in the greater Los Angeles area in this study indicated little seasonally dependent bias. For Los Angeles specifically, previous work has assessed meteorological representation, determining what could be considered an optimal approach to high-resolution simulations and performing substantive validation [*Feng et al.*, 2016; *Angevine et al.*, 2013].

With this meteorological underpinning, *Yadav et al.* [2018] performs inversions in Los Angeles evaluating what can be learned with such an optimized system. In this study, rather than focusing on developing and validating an optimal transport representation, we use the *Yadav et al.* [2018] results as a 'base' case. We compare estimated fluxes from geostatistical inversions driven by this optimized base system with fluxes estimated from geostatistical inversions driven by three broadly-available models or reanalysis products: High-Resolution Rapid Refresh (HRRR), North American Regional Reanalysis (NARR), and the Global Data Assimilation System (GDAS). We evaluate how these different inversions perform at determining the absolute flux, detecting both anomalous high emissions events and seasonal flux variance across the basin, and evaluate the role of observation site density is achieving these objectives. Los Angeles provides an opportunistic location for these tests as the large leak from the Aliso Canyon storage facility, which released an estimated 97,100 Mg over four months beginning in October 2015 [*Conley et al.*, 2016], provides what could be considered a tracer release experiment for our purposes. Additionally, seasonal variation in methane emissions has been previously observed and reported [*Yadav et al.*, 2018], and also provide a challenge test case for our non-optimized meteorological drivers.

2 Approach

We perform geostatistical inversions of methane flux between July 1, 2015 and December 31, 2016, using transport driven by each of four meteorological models or reanalysis products: WRF, HRRR, NARR, and GDAS. Each product is used to drive the Lagrangian transport model STILT [*Lin et al.*, 2003; *Nehrkorn et al.*, 2010] in order to estimate the sensitivity of in situ CH_4 mole fraction measurements to emissions fluxes. We estimate fluxes using a geostatistical inversion system based on that developed by *Yadav et al.* [2018], with a spatial resolution of 0.03 degrees within the SoCAB and at a temporal resolution of four days. The study domain along the coast of Southern California, along with the locations of the observing sites and the Aliso Canyon gas storage facility, is shown in Figure 1.

One of the four meteorological drivers we consider, the Weather Research and Forecasting model (WRF) as configured by *Feng et al.* [2016], has been extensively validated by those authors against observations of wind speed and direction and of PBL height in the Los Angeles area, as well as by comparing forward-modeled CO_2 emissions from the detailed Hestia inventory to in situ and flask mole fraction observations. That validation provided the basis for the WRF runs used in *Yadav et al.* [2018], which are the same ones we use here. The inner WRF domain, which includes the region considered here, has a spatial resolution of 1.3 km and a time step of less than one minute. More details of the WRF setup are given in Supplementary Table S1.

To verify that this WRF configuration makes a reasonable base case for a locally-tuned driver of transport, we supplement the existing validation by *Feng et al.* [2016] by directly testing observable meteorological variables in the WRF configuration against those measured at 42 surface observation sites. Agreement is generally good. Across four-day periods between January 2015 and March 2016 (overlapping but not identically with our inversion timeframe), 10 m wind speed bias errors are below 0.5 m/s in 87% of cases, with RMS errors generally in the 1.5 to 2.0 m/s range. Bias errors in 2 m temperature are below 1 K in 92% of cases with RMS errors generally around 1.5 to 2.0 m/s. Despite strong seasonal variation in meteorology in Southern California, we find no discernible seasonality in RMS or bias errors of temperature or wind speed; see Supplemental Figure S1. While future improvements of transport representation are always possible, the combination of past validation and the meteorological comparison presented here establish that it is reasonable for us to treat the WRF system as a representative base case for a locally-tuned driver of transport.

In contrast, the NOAA High Resolution Rapid Refresh model (HRRR) [*Benjamin et al.*, 2016] has a resolution of 3 km over the continental United States and uses a WRF physics model assimilating radar data every 15min, but is not optimized for the local environment. HRRR output is available as of mid-2015, albeit with some gaps, most notably in August 2016 when the model was upgraded to Version 2. In addition, some STILT runs driven by HRRR fail before the full prescribed simulation period is complete. We exclude from the HRRR inversions any observations for which the necessary HRRR fields are not available or for which the HRRR-STILT sensitivity calculations cover 12 hours or less due to gaps in STILT-HRRR. The latter condition excludes 4.2% of observations, spanning every month of the study period but especially concentrated (6.9%) in November 2015 through March 2016. Although the increased failure rate coincides with the Aliso Canyon gas leak, we judge that it remains low enough to permit evaluation of the HRRR-STILT inversion.

The North American Regional Reanalysis (NARR) [*Mesinger et al.*, 2006] and the Global Data Assimilation System (GDAS) are much coarser, with resolutions of 32 km and 0.5 degrees respectively and timesteps of 3 hrs, but cover larger areas (North America and the whole globe). An advantage to using HRRR, NARR, and GDAS is that all are run in a routine operational mode; output can be downloaded from the NOAA READY archive in a format immediately suitable for transport modeling. For low cost, low latency flux estimation in any urban environment, these products are available off-the-shelf.

We would not expect coarse products like NARR and GDAS to accurately represent conditions on fine spatial scales within our estimation domain, which spans only about 200 km from east to west. The complex topography and sea breeze circulation pattern of the LA basin [*Lu and Turco*, 1994, 1995] further complicate the environment for transport modeling. *Lin et al.* [2017] emphasize the failure of transport driven by coarse meteorology to reproduce the diurnal cycle of CO_2 mole fraction in mountainous terrain. However, several factors may mitigate the effect of poorly resolved topography: while the SOCAB domain includes significant elevation changes, most of the observing sites are located in the valley; CH_4 flux generally has a less pronounced diurnal cycle than does CO_2 flux; and, as recommended by *Lin et al.* [2017] for coarse meteorology, we use only observations taken between 12:00 and 16:00 local time, when the terrain effects are minimized and the representation of vertical mixing is believed to be most reliable.

Driven by each meteorological product, STILT simulates the transport of 800 particles 60 hours back in time from each observation. The 60-hour simulation time was chosen conservatively to ensure that all recent within-domain influences on the particles are captured. In addition to advection, STILT includes a stochastic component that can simulate particle motion on spatial and temporal scales shorter than that of the driving meteorology, which may help mitigate the effect of using temporally coarse products like NARR and GDAS.

Our inversions process data from the surface monitoring network maintained by the LA Megacities Carbon Project, which measures CH_4 mole fractions at nine locations within our domain: Granada Hills (GRA), Mount Wilson Observatory (MWO), Pasadena/Caltech (CIT), downtown LA at the University of Southern California (USC), Compton (COM), CSU Fullerton (FUL), UC Irvine (IRV), Ontario (ONT), and San Bernardino (SBC). Detailed information about each site is given in [*Verhulst et al.*, 2017]. Data availability for each site during the study period is shown in Supplementary Figure S2; an additional site at Canoga Park (CNP) was not used here because it came online only in Oc-

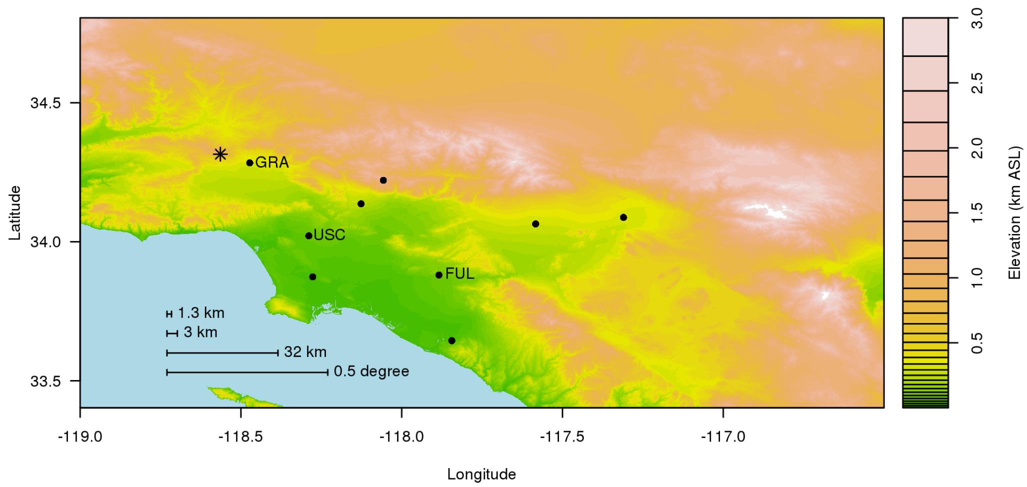


Figure 1. Colors: Elevation map of the study domain. Circles: locations of observing sites. The three sites included in the reduced network are indicated by their three-letter codes. The star in the western part of the domain indicates the location of the Aliso Canyon facility. Scale bars indicate the grid sizes for the WRF (1.3 km), HRRR (3 km), NARR (32 km), and GDAS (0.5°) meteorological fields, showing the coarse resolution of the latter fields relative to the domain.

tober 2016, at the end of our study period. Background concentrations are estimated as in *Verhulst et al.* [2017].

In order to test the impact of network density, we also perform inversions using a reduced network and using a single observing site (in addition to the background site). The single-site inversions use the network's most centrally-located site, at the University of Southern California in downtown Los Angeles (USC). The USC site was chosen to reflect a plausible design for a network consisting of only a single site, which would likely be designed to be sensitive to as much of the domain as possible at least part of the time. The reduced-network inversions use the sites at Fullerton (FUL), in the eastern part of the domain, and at Granada Hills (GRA), in the northwest near the Aliso Canyon facility, in addition to the USC site. These sites are selected to cover a broad domain in the basin and because observations are available for these three sites for the vast majority of the study period. In both the single-site and reduced-network cases, we would expect inversion performance to suffer if sites covering less of the domain were chosen. A complete description of the observing network is available in *Verhulst et al.* [2017].

In all inversions, we employ the geostatistical inversion methodology developed by *Yadav et al.* [2018]. In addition to a model linearly proportional to the distribution of emissions in the CALGEM inventory [*Zhao et al.*, 2009; *Jeong et al.*, 2012], we include a spatially constant model component, since we expect that the inversions using coarse meteorology may be unable to resolve the location of detected fluxes. Note that no input singles out either the location or the time period of the Aliso Canyon natural gas leak. In other words, this inversion makes use of no prior knowledge of the leak. We constrain the methane fluxes to nonnegative values using a bounded version of limited memory BFGS optimization [*Byrd et al.*, 1995], which is well suited to rapidly minimizing functions of many variables and thus facilitates rapid, near-real time calculations. This is different from the Lagrange multiplier approach used in *Yadav et al.* [2018]. Additional subtle differences between the WRF inversion case here and that of *Yadav et al.* [2018] are that we exclude periods in which STILT transport fails using any of our meteorological products (as described above), our focused time series is slightly different, and we do not include the Canoga Park site when it comes online late in the time period. These differences are driven by either the motivation to construct a fast, operational system or to ensure we make fair 1:1 comparisons across meteorological products.

The nonnegativity constraint on fluxes makes the posterior emissions probability non-Gaussian, which prevents us from calculating posterior uncertainties analytically. Uncertainties can be computed as in *Yadav et al.* [2018] by generating realizations from the posterior covariance distribution. However, each inversion covers only two consecutive four-day periods, the first of which is discarded as a spin-up window. As a result, the posterior uncertainty may not fully account for variation due to changes in the (actual or modeled) sensitivity of the observations to localized surface fluxes. That variation is especially important for our purposes, since we test the detectability of localized flux events and since we use coarse meteorological products in which the footprint of sensitivity may be misplaced even when its magnitude is correct. We therefore rely on the spread of flux estimates across a number of consecutive four-day periods, rather than a calculated uncertainty for any given period, as an estimate of variance when testing for flux changes (see section 3.2). For future near-real-time applications, this method has the additional advantage of saving the computing time needed to generate the realizations.

3 Analysis

3.1 Basin Total Flux

Estimated whole-basin methane fluxes from each of the four inversions are shown in Figure 2. The Aliso Canyon event and seasonal cycle, known features we are using to test operational meteorologies, appear evident in all inversions and we assess this statistically in section 3.2. All inversions show emissions up-ticks prior to the start of the Aliso Canyon event, which could be indicative of the leak beginning before the noted start date, or part of the seasonal increase in emissions. While this study does not attribute this feature, note that it is not explained by the timing of the four-day periods used in the inversion, since the increase begins in periods which do not overlap the reported leak. Considering emissions magnitudes, when the full observing network is included, estimates using transport driven by WRF and NARR average 53 and 47 Mg/hr outside the Aliso Canyon leak period, respectively, in broad agreement with the 35 to 50 Mg/hr range of baseline emissions estimates in other studies [e.g. *Wennberg et al.*, 2012; *Peischl et al.*, 2013; *Wecht et al.*, 2014; *Wong et al.*, 2015]. That our estimates fall at the upper end of that range is not surprising given that much of the previous work relied on observations taken in May-June 2010, not during the peak of the seasonal emissions cycle (see section 3.2). Estimates using HRRR are considerably higher than those using WRF, by about 96% on average over the 18-month study period, and estimates using GDAS are somewhat lower, by about 16% on average.

Much of the difference in estimated flux is explained by the difference in overall mean total sensitivity assigned by each model to the measurement network. We compute the mean total sensitivity $\mathrm{H}_{\mathrm{mean}}$ for each model over the 18-month period of the study by summing the sensitivity of the nine measurement sites, then taking the mean over spatial flux grid cells and over observation times. In order to make a direct comparison, we exclude (for all models) observations for which HRRR fields are missing or for which HRRR-STILT runs failed; see Section 2. Treating WRF as a transfer standard, we perform an empirical calibration, scaling the posterior fluxes $\mathbf{s}_j$ from the NARR, HRRR, and GDAS-driven inversions (j) by the ratios of the sensitivities computed using those models relative to those using WRF:

$$\mathbf{s}_{\mathrm{cal},j} = \frac{\mathrm{H}_{\mathrm{mean},j}}{\mathrm{H}_{\mathrm{mean,WRF}}} \times \mathbf{s} \tag{1}$$

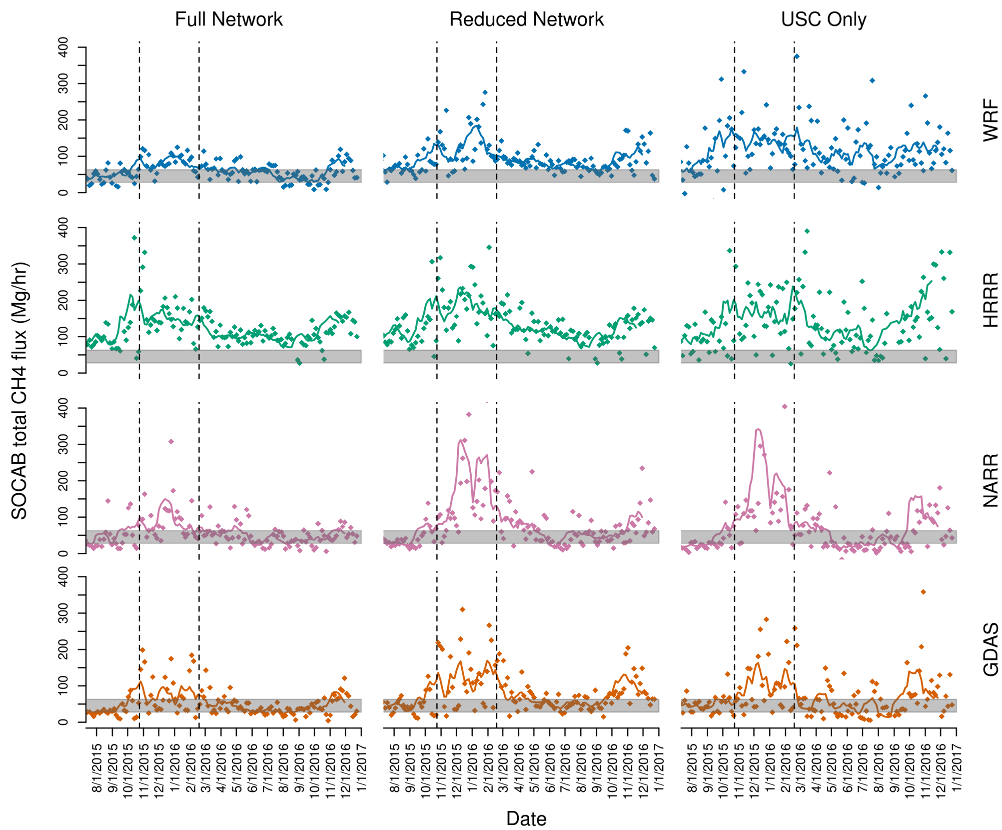


Figure 2. Points: estimated total CH_4 flux time series for the South Coast Air Basin (SoCAB), at four-day time intervals, according to inversions using transport driven by each of four meteorological models and using the full observing network (9 sites), a reduced network (3 sites), or a single observing site. Curves: 28-day running means of each time series for visual reference (not used in the analysis). The shaded band indicates the typical range of estimates in past studies. The dashed vertical lines indicate the start and end dates of the Aliso Canyon natural gas leak.

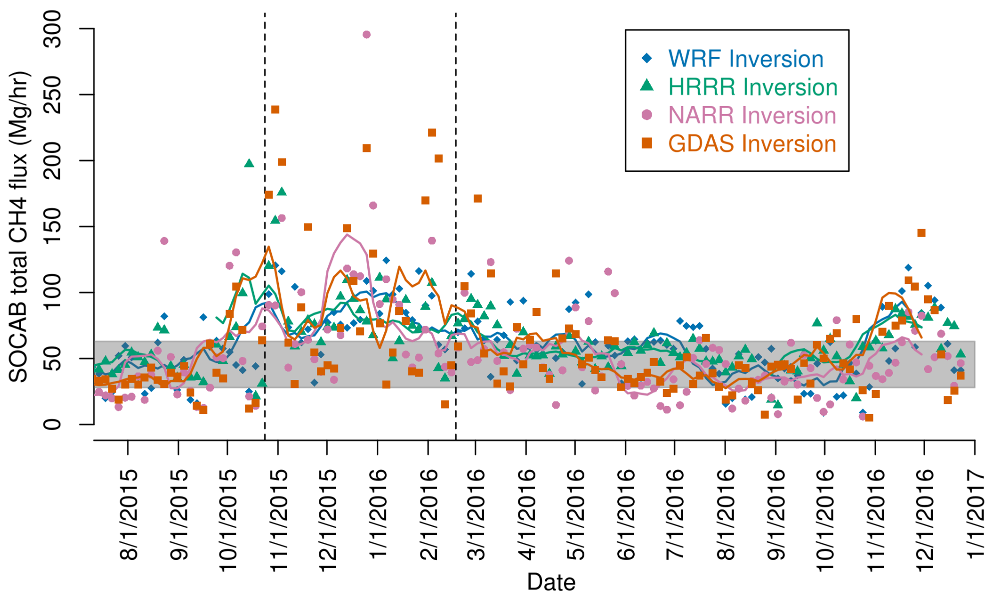


Figure 3. Estimated SoCAB total CH_4 flux time series in inversions using the full observing network after calibration by scaling the fluxes by the relative total sensitivity assigned to the observing network by each driver of the transport model. The calibration brings the estimates into close agreement overall. Curves: 28-day running means of each time series for visual reference (not used in the analysis). The shaded band indicates the typical range of estimates in past studies. The dashed vertical lines indicate the start and end dates of the Aliso Canyon natural gas leak.

After calibration, the mean posterior emissions $\mathbf{s}_{\mathrm{cal},j}$ come into much closer alignment overall. The difference in mean flux over the full 18-month study period relative to the WRF inversion is reduced to 17% with HRRR and 1% with GDAS and increases modestly to 3% with NARR. The scaled time series are shown in Figure 3. As we look at increasingly shorter time scales, more scatter remains between the calibrated flux estimates. The mean residual difference between monthly mean fluxes from the WRF inversion and calibrated estimates over the same periods from the other inversions is about 20% with HRRR and NARR and about 25% with GDAS. Individual four-day flux estimates after calibration are moderately well correlated overall, r = 0.47 to 0.50, but often diverge (see Supplementary Figure S3).

If the sensitivity bias could be corrected using direct observations, our results suggest that accurate flux estimates might be possible, at least one monthly and longer time scales, using more widely available models than is generally assumed. However, several of the meteorological factors most clearly linked to the sensitivity fail to explain the difference. STILT computes sensitivity to surface fluxes by tracking the amount of time simulated air parcels spend in contact with the surface. The sensitivity H_{ij} of the ith observation to the jth flux region is given by [*Lin et al.*, 2003]

$$H_{ij} = \frac{m_{air}}{\rho_j}\frac{\tau}{z_j}; \qquad \tau = \frac{1}{N_i}\sum_{p_i=1}^{N_i} \Delta t_{p_i,j} \tag{2}$$

where z_j is the mixing depth, accounting for the effect of dilution, and τ is the average time spent by the parcels within the bottom one-half of the mixing layer above the flux region. The average is taken over N_i simular parcels released backwards from the ith observation and indexed by p_i. On the basis of these relations, we would expect the intermodel differences in sensitivity to be explained by systematic differences either in the mixing height or in the residence time, i.e., the time for air to travel from the edge of the study domain to the observing site, as driven by the wind speed.

In the STILT runs driven by each model or reanalysis product, we computed the mean time spent in the domain by measured air parcels before encountering an observation site (residence time) as well as the time-averaged mixing depth along the parcel's path. The same filtering was applied as in computing the mean sensitivities. As shown in Table 1, the results do not explain the differences in sensitivity. On average, mixing depths in HRRR are almost the same as those in WRF, and residence times are only modestly shorter – yet the sensitivity is much less. On the contrary, mixing depths in NARR are 80% higher on average than those in WRF, yet the sensitivity is very similar.

Since parcels may be within the horizontal extent of the domain but above the bottom half of the mixing layer (and therefore considered by STILT to be insensitive to surface fluxes), we also computed the fraction of their residence time that measured parcels spent near the surface. As shown in Table 1, this 'near-surface fraction' differs from WRF by no more than 13% in any of the other models. The expected combined effect of the mixing depth, residence time, and near-surface fraction is summarized on the fourth line of Table 1, in which we compute the relative sensitivity predicted by those mean variables according to

$$\frac{\mathrm{H_{mean}}}{\mathrm{H_{mean,WRF}}} = \frac{z_{WRF}}{z} \times \frac{\tau}{\tau_{\mathrm{WRF}}} \times \frac{f}{f_{\mathrm{WRF}}} \qquad \text{(predicted)} \tag{3}$$

where f is the near-surface fraction. The resulting prediction fails to capture the actual differences in total mean sensitivity, which are given on the last line of Table 1.

	WRF	HRRR	NARR	GDAS
Mixing Depth (m)	615	612 / 99%	1109 / 180%	573 / 93%
Residence Time (min)	315	278 / 88%	250 / 79%	308 / 98%
Near-Surface Fraction	0.57	0.49 / 87%	0.65 / 115%	0.45 / 80%
Predicted Relative Sensitivity	-	/ 77%	/ 51%	/ 84%
Actual Relative Sensitivity	-	/ 53%	/ 96%	/ 120%

Table 1. First three rows: mean values of meteorological variables expected to contribute to sensitivity, for STILT driven by each of four models or reanalysis products. These variables are described in section 3.1, and percentages are relative to the same variables in WRF. Fourth row: expected ratios of the sensitivity in HRRR, NARR, and GDAS, relative to that in WRF, given the above variables. Fifth row: actual ratios of the sentivity in HRRR, NARR, and GDAS to that in WRF. The actual relative sensitivities are not accurately predicted on the basis of the mean meteorological variables.

Therefore, although basin-wide, 18-month-average sensitivity explains the gross differences in estimated flux between the inversions, the basin-wide, 18-month-average differences in the relevant underlying meteorological variables do not control the sensitivity in the same way. In the transport model, the whole basin is not treated as a single region; rather, Equation 2 applies separately in each 0.03-degree grid cell and for each four-day period, and the fine-scale interactions between the variables have a substantial effect.

An important implication is that our modeled average sensitivities could not be calibrated to ground truth by debiasing the underlying meteorological variables in a basin-averaged manner. For example, using lidar observations in Pasadena, California (colocated with one of the LA Megacities observing sites), *Ware et al.* [2016] showed that NARR persistently overestimates the mixing depth at that location, by more than a factor of two on average, and that any local mixing depth bias in WRF was likely much smaller. Indeed, we can see in Table 1 that mixing depths in NARR are very high on average over the whole domain. However, if the estimated fluxes in the NARR inversion were scaled to correct for this bias as suggested by *Ware et al.* [2016], the result would be to introduce a large positive bias into the fluxes. Of course, wind speed and mixing depth observations can be used to evaluate and improve meteorological drivers of transport, as was done for the WRF configuration employed here by *Feng et al.* [2016] – but our results show that a mean calibration factor constructed from those observations could not be reliably correct.

We might expect that the mean meteorological variables would better predict the total sensitivity over shorter time periods, since correlations between the variables might be less important. However, we find that this is not the case on monthly timescales (see Supplementary Table S2), nor do calibration factors constructed from monthly average total sensitivities perform as well as the calibration factors calculated over the full 18-month study period. Calibration factors computed seasonally do somewhat better, but in most cases, seasonal mean fluxes come into closer agreement after applying the full 18-month calibration than after applying seasonal calibration. Overall, the calibration method seems to be most effective when applied over a year or more.

One alternative to computing calibration factors from meteorological observations could be to run a trusted custom model for a limited period, compute a calibration using the mean sensitivity for that period, then continue estimating fluxes using an operational product. Though the time period of our study is too limited for a conclusive demonstration, our experience suggests that this approach could be successful. We computed calibration factors for each of HRRR, NARR, and GDAS based on the first twelve months of the study period, then applied those factors to the flux estimates for the last six months, July-December 2016. That calibration reduced the difference in mean flux between HRRR- and WRF-driven inversions from 103% to 2% and between GDAS- and WRF-driven inversions from 15% to 6%, though it increased the difference between NARR- and WRF-driven inversions modestly, from 16% to 22%.

3.2 Anomaly and Trend Detection

We evaluate the ability of each inversion system to detect changes in the total basin flux, both seasonally and due to an unusual event or change. We test significance using Welch's unequal-variances t-test, which has similar power to a standard t-test and is appropriate whether or not the samples to be compared have the same variance. The significances (p-values) for all the tests described in this section are given in Table 2.

In all of the inversions using the full observing network, we observe a seasonal trend in CH_4 emissions. Emissions in November-December 2016 are estimated to be 38% (NARR inversion) to 83% (GDAS inversion) higher than those in July-August. These periods were selected so as not to overlap the timeframe of the Aliso Canyon leak, in order to isolate the 'normal' seasonal difference. The estimated difference is significant at the 95% level or better in all four inversions. The consistent detection and timing of the seasonal change, regardless of the meteorology used to drive transport, reinforce its status as a robust and substantial feature of Los Angeles methane emissions.

We also test the detectability of the increase in flux during the Aliso Canyon leak period. To remove the impact of the seasonal dependence, we compare the period October 24 through December 27, 2015 to the corresponding period in 2016 (in an operational setting, the comparison would generally be to previous years). The difference is significant at the 95% level in Welch's t-test in the HRRR, NARR, and GDAS inversions but much less significant (p=0.17) in the WRF inversion. Note that this test of event detectability is distinct from quantifying the rate of a known leak as in *Yadav et al.* [2018].

Our ability to observe the Aliso Canyon gas leak using the LA Megacities observing network is limited by its position near the edge of the inversion domain, such that its emissions are observable only intermittently. However, as is apparent in Figure 2, this intermittency can result in an increase in the variance of the retrieved fluxes, which may be significant even, or indeed especially, when the change in mean is not. In fact, in an F-test for difference of variance comparing October-December 2015 to 2016 as above, the increase in retrieved flux variance during the Aliso Canyon period is nearly as significant or more significant than the change in mean flux in the inversions driven by HRRR, NARR, and GDAS. The increase in variance is not significant (p=0.32) in the inversion driven by WRF, which shows the least variability relative to the estimated flux values. These results highlight the complimentary value of the two approaches, particularly for less-optimized meteorology.

That the inversion driven by WRF does not significantly detect the Aliso Canyon event using our tests may be surprising. One plausible explanation is that, during the leak period, the WRF inversion produces consistent but only moderately elevated flux estimates. This moderate increase is not sufficient to distinguish itself from the corresponding increase in late 2016. By contrast, the other inversions produce exceptionally high estimates for some four-day periods. Even though estimates for other periods are not elevated, the average increase is sufficient for detection.

a.) Seasonal Difference, Welch's t-test

	WRF	HRRR	NARR	GDAS
Full Network	0.047*	<0.001*	0.048*	0.012*
Reduced Network	0.024*	<0.001*	0.075	<0.001*
USC Site Only	0.53	0.0012*	0.025*	0.015*

b.) Aliso Canyon Period, Welch's t-test

	WRF	HRRR	NARR	GDAS
Full Network	0.17	0.025*	0.016*	0.039*
Reduced Network	0.63	0.004*	0.051	0.30
USC Site Only	0.15	0.39	0.24	0.89

c.) Aliso Canyon Period, F-test for Difference of Variance

	WRF	HRRR	NARR	GDAS
Full Network	0.32	<0.001*	<0.001*	0.044*
Reduced Network	0.60	0.056	0.016*	0.021*
USC Site Only	0.45	0.21	0.36	0.82

Table 2. Summary of p-values of two-sided tests for changes in mean emissions (a and b) or variance of emissions (c), comparing summer to winter of 2016 (a) or the first 64 days of the Aliso Canyon gas leak in 2015 to the equivalent period in 2016 (b and c). Tests significant at the 95% level are indicated with an asterisk. Seasonal flux differences are detected in most cases even with reduced observations; the Aliso Canyon leak is detected with the full network in the non-WRF inversions and with the reduced network in some cases using the test of difference of variance.

The difference in variability between the WRF inversion and the others may be due to the assignment of covariance parameters according to Restricted Maximum Likelihood (RML) analysis. Rather than assign prior uncertainties by expert judgment, RML finds the combination of covariances that make the actual observations most likely, given the sensitivity footprints computed by the transport [*Michalak et al.*, 2004]. The variances of the observations and the spatial pattern of prior covariance are therefore intermediate statistical quantities which are calculated during the course of the inversion. In our WRF-driven inversion, RML assigns most of the prior covariance to the spatially constant pattern. The result is that the cost of attributing an observed excess mole fraction to a flux is mostly insensitive to the spatial distribution of the observation's sensitivity footprint. In the other inversions, although the magnitude of prior covariance is similar *on average*, RML assigns more weight to the spatial pattern proportional to the CALGEM inventory, so the penalty for assigning an excess flux is more spatially variable. This would tend to make the inversion more sensitive to the modeled wind direction, which may not be accurate. If the footprint of a high observed mole fraction falls over a source known to CALGEM, the flux estimate can be increased a great deal at little cost; but if the footprint falls over an area without sources in CALGEM, increasing the flux estimate is costly.

In general, the threshold for a flux event to be detectable by a given observing and inversion system depends not only on the magnitude of the event but also on its duration and variance. It also depends on the event's timing, because the mean flux and variance during the reference period used for comparison will vary according to the seasonal cycle. By way of an example, for a hypothetical event persisting at least from September 4 to October 26, 2017 (and compared to the corresponding period in 2016), we compute the sensitivity according to the better of Welch's t-test and the F-test for difference of variance for a range of estimated flux increases and variances. The results are shown in Figure 4 for the inversions driven by each of the four meteorological products. In this example, a flux increase estimated at 30-40% above the baseline by an inversion using WRF or NARR would be detected as significant if the variance were approximately unchanged. The same is true for an increase estimated at 20-30% by the inversion using GDAS or estimated at about 20% by the inversion using HRRR. Note, however, that the same thresholds do not persist at other times and that the threshold for the actual flux increase due to an event may be higher if the event is not consistently upwind of the observing sites.

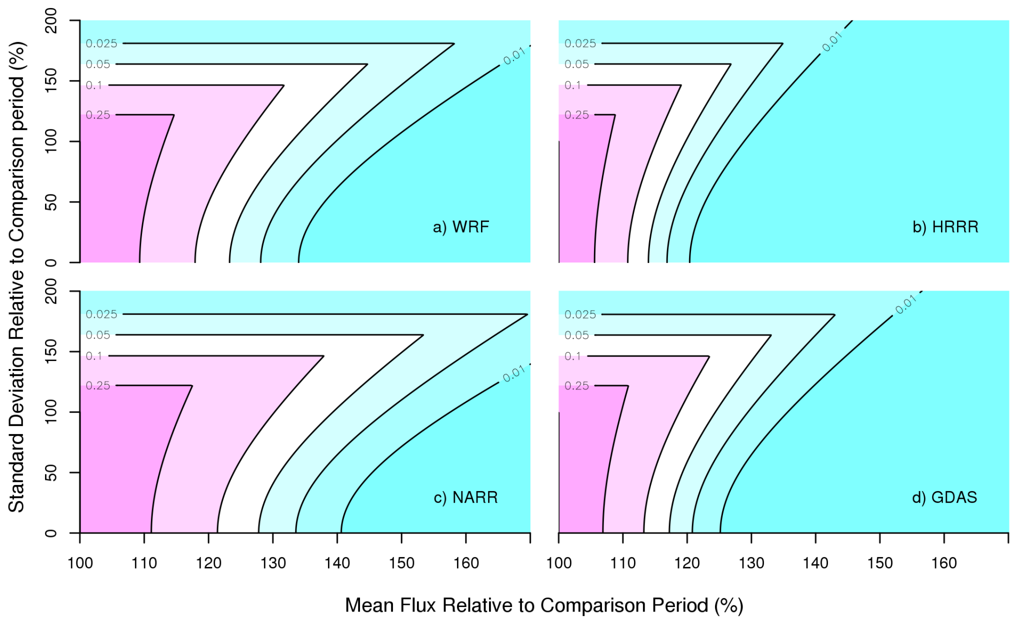


Figure 4. Sensitivity (p-values) of inversions using each meteorological driver to hypothetical flux events occurring between September 4 and October 26, 2017, as a function of the change in mean flux and variance relative to the same period in 2016. The inversions shown here use the full observing network (9 sites). Changes in mean flux are less significant when accompanied by high variance, but sufficiently large variance increases are themselves significant in an F-test.

3.3 Network Density

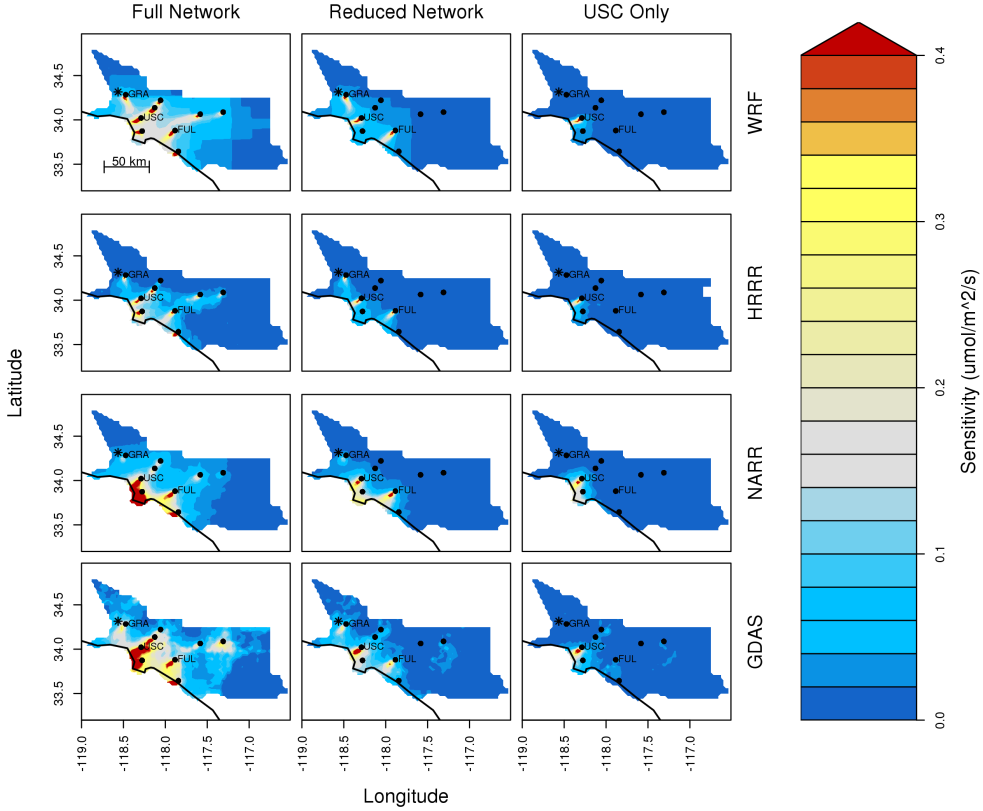
As the number of observing sites is reduced, the methane flux retrievals generally become noisier, exhibiting greater variance even in the absence of any known flux event. In almost all cases, robustly detecting the Aliso Canyon leak event is more difficult with only three observing sites than with the full network. However, the HRRR-driven inversion remains sensitive to the change in mean flux (p=0.004) and the NARR- and GDAS-driven inversions remain sensitive to the increase in variance (p=0.016 and p=0.021, respectively).

With only a single observing location, none of our inversions can detect a significant change either in the mean or in the variance of the fluxes during the Aliso Canyon leak. The USC site alone can constrain only a small part of the study domain, and even that part only inconsistently. Figure 5 illustrates the decrease in measurement constraint when the number of the number of observing sites is reduced.

By contrast, even a single measurement location is sufficient in most of our inversions (excepting that using WRF) to observe the seasonal cycle. Broad and consistent sensitivity may be less critical for this purpose than for detecting a point source event because the seasonal difference is likely to be widely distributed throughout the domain. Although our study period is too short to observe it, we might expect the same to apply to year-over-year secular changes.

4 Conclusions

Our results suggest that the ability of an in situ observing network to detect changes in emissions may be less sensitive to the choice of transport driver than are estimates of the absolute total flux. Much of the difference in absolute flux estimates between inversions driven by divergent meteorology seems to be attributable to biases in long-term sensitivity, which can be calibrated by comparison to a trusted model chosen as a transfer standard. Debiasing with weather observations (e.g. scaling results by observed bias in mixing depth) would not be successful as the sensitivity bias is not predicted by the mean values of the relevant underlying meteorological variables. However, an accurate total estimate is not a prerequisite for observing changes, including seasonally or in the case of leaks or other large anomalies. Although our study period is not long enough to directly observe, trends over the course of years could likely be characterized in the same way. We



483 **Figure 5.** Heat map: sensitivity of the full observing network (9 sites), a reduced network (3 sites), and the USC site alone to fluxes within the SoCAB during the first four days of the Aliso Canyon natural gas leak, October 24-27, 2015, as computed by STILT driven by each of four meteorological products. Circles: locations of observing sites. The three sites included in the reduced network are indicated by their three-letter codes. The star near the western edge of the domain indicates the location of the Aliso Canyon facility. The breadth and magnitude of sensitivity degrade as measuring locations are removed.

find that even with only a single observing site, seasonal flux changes emerge as robustly detectable with operational meteorology supporting an inversion, suggesting sparse urban networks can potentially provide valuable, rapid information.

The ability of a surface network to detect flux changes contributes to the functioning of a 'tiered' observing system [*Duren and Miller*, 2012] for megacities carbon emissions, which includes continuous monitoring at the urban scale, targeted deployments to characterize significant individual sources, and regional or boundary condition data from aircraft and satellites, as well as bottom-up inventories. A flux inversion system run operationally could provide the first notice of events worthy of more detailed investigation by other methods. The more quickly these events can be identified, the better opportunity we will have to quantify and characterize them as well as to inform stakeholders.

So far, the ability to usefully detect emissions events using urban concentration measurements has been limited by the long time delay, typically measured in years, between collecting initial data and producing a flux estimate. (An exception was the near-real-time monitoring performed by *Lauvaux et al.* [2013] in Davos, Switzerland in 2011-2012.) One major source of latency is the time, expense, and computational resources involved in meteorological modeling for transport. Others have begun demonstrating forward model simulations using operational meteorology *Pugliese* [2017]. We now have demonstrated that at least some operational monitoring goals utilizing atmospheric inversions can be met using a variety of meteorological products, including several that are made available on a routine basis and nearly in real time. Output from HRRR is posted on the NOAA READY archive each day, covering the previous day. Continuous archival of GDAS has recently been supplanted by Global Forecast System (GFS) short-term forecasts, which are initialized with GDAS but have twice the resolution both in space (0.25 degrees) and in time (3 hours). GFS zero-hour forecasts are finalized the same day, and since GFS covers the whole globe, they can be retrieved for the vicinity of any major city or other area of interest. Our work shows that the coarse spatial resolution of these products does not necessarily limit their utility in an urban setting.

Once the meteorological fields are ready, the remaining computational requirements can be modest. For this study, calculating influence footprints with STILT using HRRR meteorology took about fifteen minutes for each observation on a 2.2 GHz CPU with 128 GB of RAM. In total, running footprints for up to 16 observations in parallel, the footprints for a single inversion covering two consecutive four-day periods took about 5.5 hours to calculate. In an operational mode, each day's footprints could be run the next day, taking less than one hour. The geostatistical inversions themselves each took only about two minutes, although that time would be longer if we computed posterior covariances as in *Yadav et al.* [2018] or, especially, if we allowed off-diagonal terms in the prior covariances.

This suggests that the remaining obstacle for an operational near-real time inversion system lies not in latency of meteorological drivers, flux priors, or inversion calculation, but instead on the rapid collection of QA/QC'd network observations, and in cases where global models are used for background concentrations, the latency of those global model runs. Given that this work suggests fluxes can be estimated rapidly once concentration data is collected and quality-controlled, accelerating this step could see a near-real time system actually implemented.

Acknowledgments

The authors thank Thomas Nehrkorn for assistance with observational validation of WRF meteorology. Support was provided by the NIST Greenhouse Gas and Climate Science Measurements Program, including under grant 70NANB17H176, by the NOAA Atmospheric Chemistry, Carbon Cycle, and Climate Program, and by NASA under grant NNN12AA01C. Portions of this work were performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.

Mole fraction data used in this study are provided in a supplementary spreadsheet. Up to date data from LA basin including other time periods can be found at the Megacities Carbon Project web portal at https://megacities.jpl.nasa.gov/portal/. Meteorological fields from HRRR, NARR, and GDAS are available on the NOAA READY archive at https://www.ready.noaa.gov/archives.php.

References

Angevine, W. M., L. Eddington, K. Durkee, C. Fairall, L. Bianco, and J. Brioude (2012), Meteorological model evaluation for calnex 2010, *Monthly Weather Review*, *140*(12), 3885–3906.

Angevine, W. M., J. Brioude, S. McKeen, J. S. Holloway, B. M. Lerner, A. H. Goldstein, A. Guha, A. Andrews, J. B. Nowak, S. Evan, et al. (2013), Pollutant transport among california regions, *Journal of Geophysical Research: Atmospheres*, *118*(12), 6750–6763.

Bagley, J. E., S. Jeong, X. Cui, S. Newman, J. Zhang, C. Priest, M. Campos-Pineda, A. E. Andrews, L. Bianco, M. Lloyd, et al. (2017), Assessment of an atmospheric transport model for annual inverse estimates of california greenhouse gas emissions, *Journal of Geophysical Research: Atmospheres*, *122*(3), 1901–1918.

Benjamin, S. G., S. S. Weygandt, J. M. Brown, M. Hu, C. R. Alexander, T. G. Smirnova, J. B. Olson, E. P. James, D. C. Dowell, G. A. Grell, et al. (2016), A North American hourly assimilation and model forecast cycle: The Rapid Refresh, *Monthly Weather Review*, *144*(4), 1669–1694.

Breon, F. M., G. Broquet, Puygrenier, F. Chevallier, I. Xueref-Rémy, M. Ramonet, E. Dieudonne, M. Lopez, M. Schmidt, O. Perrussel, and P. Ciais (2014), An attempt at estimating Paris area CO_2 emissions from atmospheric concentration measurements, *Atmospheric Chemistry and Physics*, *14*, 9647–9703, doi:10.5194/acpd-14-9647-2014.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995), A limited memory algorithm for bound constrained optimization, *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208.

Conley, S., G. Franco, I. Faloona, D. R. Blake, J. Peischl, and T. Ryerson (2016), Methane emissions from the 2015 Aliso Canyon blowout in Los Angeles, CA, *Science*, p. aaf2348.

Davis, K. J., A. Deng, T. Lauvaux, N. L. Miles, S. J. Richardson, D. P. Sarmiento, K. R. Gurney, R. M. Hardesty, T. A. Bonin, W. A. Brewer, et al. (2017), The Indianapolis Flux Experiment (INFLUX): A test-bed for developing urban greenhouse gas emission measurements, *Elementa: Science of the Anthropocene*, *5*.

Deng, A., T. Lauvaux, K. J. Davis, B. J. Gaudet, N. Miles, S. J. Richardson, K. Wu, D. P. Sarmiento, R. M. Hardesty, T. A. Bonin, et al. (2017), Toward reduced transport errors in a high resolution urban CO_2 inversion system, *Elementa: Science of the Anthropocene*, *5*.

Duren, R. M., and C. E. Miller (2012), Measuring the carbon emissions of megacities, *Nature Climate Change*, *2*(8), 560.

Feng, S., T. Lauvaux, S. Newman, P. Rao, R. Ahmadov, A. Deng, L. I. Díaz-Isaac, R. M. Duren, M. L. Fischer, C. Gerbig, K. R. Gurney, J. Huang, S. Jeong, Z. Li, C. E. Miller, D. O'Keeffe, R. Patarasuk, S. P. Sander, Y. Song, K. W. Wong, and Y. L. Yung (2016), Los Angeles megacity: a high-resolution land–atmosphere modelling system for urban CO_2 emissions, *Atmospheric Chemistry and Physics*, *16*(14), 9019–9045, doi: 10.5194/acp-16-9019-2016.

Gourdji, S., V. Yadav, A. Karion, K. Mueller, S. Conley, T. Ryerson, T. Nehrkorn, and E. Kort (2018), The Aliso Canyon natural gas leak as a natural tracer experiment: Reducing errors in aircraft atmospheric inversion estimates of point-source emissions, *Environmental Research Letters*.

Gurney, K. R., I. Razlivanov, Y. Song, Y. Zhou, B. Benes, and M. Abdul-Massih (2012), Quantification of fossil fuel CO_2 emissions on the building/street scale for a large US city, *Environmental Science & Technology*, *46*(21), 12,194–12,202.

Harrison, R., M. Dall'Osto, D. Beddows, A. Thorpe, W. Bloss, J. Allan, H. Coe, J. Dorsey, M. Gallagher, C. Martin, et al. (2012), Atmospheric chemistry and physics in the atmosphere of a developed megacity (London): an Overview of the REPARTEE experiment and its conclusions, *Atmospheric Chemistry and Physics*, *12*(6), 3065–3114.

Jeong, S., C. Zhao, A. E. Andrews, L. Bianco, J. M. Wilczak, and M. L. Fischer (2012), Seasonal variation of CH_4 emissions from central California, *Journal of Geophysical Research: Atmospheres*, *117*(D11).

Kort, E. A., C. Frankenberg, C. E. Miller, and T. Oda (2012), Space-based observations of megacity carbon dioxide, *Geophysical Research Letters*, *39*(17).

Kort, E. A., W. M. Angevine, R. Duren, and C. E. Miller (2013), 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*(3), 1577–1584.

Lauvaux, T., N. L. Miles, S. J. Richardson, A. Deng, D. R. Stauffer, K. J. Davis, G. Jacobson, C. Rella, G. Calonder, and P. L. DeCola (2013), 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*(12), 2654–2668, doi:10.1175/jamc-d-13-038.1.

Lin, J., and C. Gerbig (2005), Accounting for the effect of transport errors on tracer inversions, *Geophysical Research Letters*, *32*(1).

Lin, J., C. Gerbig, S. Wofsy, A. Andrews, B. Daube, K. Davis, and C. Grainger (2003), 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*(D16).

Lin, J. C., D. V. Mallia, D. Wu, and B. B. Stephens (2017), How can mountaintop co 2 observations be used to constrain regional carbon fluxes?, *Atmospheric Chemistry and Physics*, *17*(9), 5561–5581.

Lopez-Coto, I., S. Ghosh, K. Prasad, and J. Whetstone (2017), Tower-based greenhouse gas measurement network design âĂŤ- The National Institute of Standards and Technology North East Corridor Testbed, *Advances in Atmospheric Sciences*, *34*(9), 1095–1105.

Lu, R., and R. P. Turco (1994), Air pollutant transport in a coastal environment. Part I: two-dimensional simulations of sea-breeze and mountain effects, *Journal of the Atmospheric Sciences*, *51*, 2285–2308, doi:10.1175/1520-0469(1994)051<2285:APTIAC>2.0.CO;2.

Lu, R., and R. P. Turco (1995), Air pollutant transport in a coastal environment—II. Three-dimensional simulations over Los Angeles basin, *Atmospheric Environment*, *29*, 1499–1518, doi:10.1016/1352-2310(95)00015-Q.

Mays, K. L., P. B. Shepson, B. H. Stirm, A. Karion, C. Sweeney, and K. R. Gurney (2009), Aircraft-based measurements of the carbon footprint of Indianapolis, *Environmental Science & Technology*, *43*(20), 7816–7823.

McKain, K., S. C. Wofsy, T. Nehrkorn, J. Eluszkiewicz, J. R. Ehleringer, and B. B. Stephens (2012), Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region, *Proceedings of the National Academy of Sciences*, *109*, 8423–8428, doi:10.1073/pnas.1116645109.

McKain, K., A. Down, S. M. Raciti, J. Budney, L. R. Hutyra, C. Floerchinger, S. C. Herndon, T. Nehrkorn, M. S. Zahniser, R. B. Jackson, et al. (2015), Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts, *Proceedings of the National Academy of Sciences*, *112*(7), 1941–1946.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. H. Berbery, et al. (2006), North American regional reanalysis, *Bulletin of the American Meteorological Society*, *87*(3), 343–360.

Michalak, A. M., L. Bruhwiler, and P. P. Tans (2004), A geostatistical approach to surface flux estimation of atmospheric trace gases, *Journal of Geophysical Research: Atmospheres*, *109*(D14).

Nehrkorn, T., J. Eluszkiewicz, S. C. Wofsy, J. C. Lin, C. Gerbig, M. Longo, and S. Freitas (2010), Coupled Weather Research and Forecasting–Stochastic Time-Inverted Lagrangian Transport (WRF–STILT) model, *Meteorology and Atmospheric Physics*, *107*(1-2), 51–64.

Nehrkorn, T., J. Henderson, M. Leidner, M. Mountain, J. Eluszkiewicz, K. McKain, and S. Wofsy (2013), WRF simulations of the urban circulation in the Salt Lake City area for CO_2 modeling, *Journal of Applied Meteorology and Climatology*, *52*(2), 323–340.

Peischl, J., T. Ryerson, J. Brioude, K. Aikin, A. Andrews, E. Atlas, D. Blake, B. Daube, J. Gouw, E. Dlugokencky, et al. (2013), Quantifying sources of methane using light alkanes in the Los Angeles basin, California, *Journal of Geophysical Research: Atmospheres*, *118*(10), 4974–4990.

Pugliese, S. C. (2017), Observational constraints on air quality and greenhouse gases in the greater Toronto area, Ph.D. thesis, University of Toronto (Canada).

Richardson, S., N. Miles, K. Davis, T. Lauvaux, D. Martins, et al. (2016), CO_2, CO, and CH_4 surface in situ measurement network in support of the Indianapolis FLUX (INFLUX) Experiment, *Elementa: Science of the Anthropocene*.

Shusterman, A. A., V. E. Teige, A. J. Turner, C. Newman, J. Kim, and R. C. Cohen (2016), The BErkeley Atmospheric CO_2 Observation Network: initial evaluation, *Atmospheric Chemistry and Physics*, *16*(21), 13,449–13,463.

Turner, A. J., A. A. Shusterman, B. C. McDonald, V. Teige, R. A. Harley, and R. C. Cohen (2016), Network design for quantifying urban CO_2 emissions: assessing trade-offs between precision and network density, *Atmospheric Chemistry and Physics*, *16*(21), 13,465–13,475.

Verhulst, K. R., A. Karion, J. Kim, P. K. Salameh, R. F. Keeling, S. Newman, J. Miller, C. Sloop, T. Pongetti, P. Rao, et al. (2017), Carbon dioxide and methane measurements from the Los Angeles Megacity Carbon Project–Part 1: Calibration, urban enhancements, and uncertainty estimates, *Atmospheric Chemistry and Physics*, *17*(13), 8313–8341.

Ware, J., E. A. Kort, P. DeCola, and R. Duren (2016), Aerosol lidar observations of atmospheric mixing in Los Angeles: Climatology and implications for greenhouse gas observations, *Journal of Geophysical Research: Atmospheres*, *121*(16), 9862–9878.

Wecht, K. J., D. J. Jacob, M. P. Sulprizio, G. Santoni, S. C. Wofsy, R. Parker, H. Bösch, and J. Worden (2014), Spatially resolving methane emissions in California: constraints from the CalNex aircraft campaign and from present (GOSAT, TES) and future (TROPOMI, geostationary) satellite observations, *Atmospheric Chemistry and Physics*, *14*(15), 8173–8184.

Wennberg, P. O., W. Mui, D. Wunch, E. A. Kort, D. R. Blake, E. L. Atlas, G. W. Santoni, S. C. Wofsy, G. S. Diskin, S. Jeong, et al. (2012), On the sources of methane to the Los Angeles atmosphere, *Environmental Science & Technology*, *46*(17), 9282–9289.

Wong, K., D. Fu, T. Pongetti, S. Newman, E. Kort, R. Duren, Y.-K. Hsu, C. Miller, Y. Yung, and S. Sander (2015), Mapping CH_4:CO_2 ratios in Los Angeles with CLARS-FTS from Mount Wilson, California, *Atmospheric Chemistry and Physics*, *15*(1), 241–252.

Yadav, V., K. Mueller, K. Verhulst, R. Duren, T. Nehrkorn, J. Kim, R. F. Weiss, R. Keeling, S. Sander, M. Fischer, S. Newman, M. Falk, T. Kuwayama, T. Rafiq, J. Whetstone, A. Karion, and C. Miller (2018), Spatio-temporally resolved methane fluxes from the Los Angeles Megacity.

Ye, X., T. Lauvaux, E. A. Kort, T. Oda, S. Feng, J. C. Lin, E. Yang, and D. Wu (2017), Constraining fossil fuel CO_2 emissions from urban area using OCO-2 observations of total column CO_2, *Atmospheric Chemistry and Physics Discussions*, *2017*, 1–30, doi: 10.5194/acp-2017-1022.

Zhao, C., A. E. Andrews, L. Bianco, J. Eluszkiewicz, A. Hirsch, C. MacDonald, T. Nehrkorn, and M. L. Fischer (2009), Atmospheric inverse estimates of methane emissions from Central California, *Journal of Geophysical Research: Atmospheres*, *114*(D16).

Figure 1.

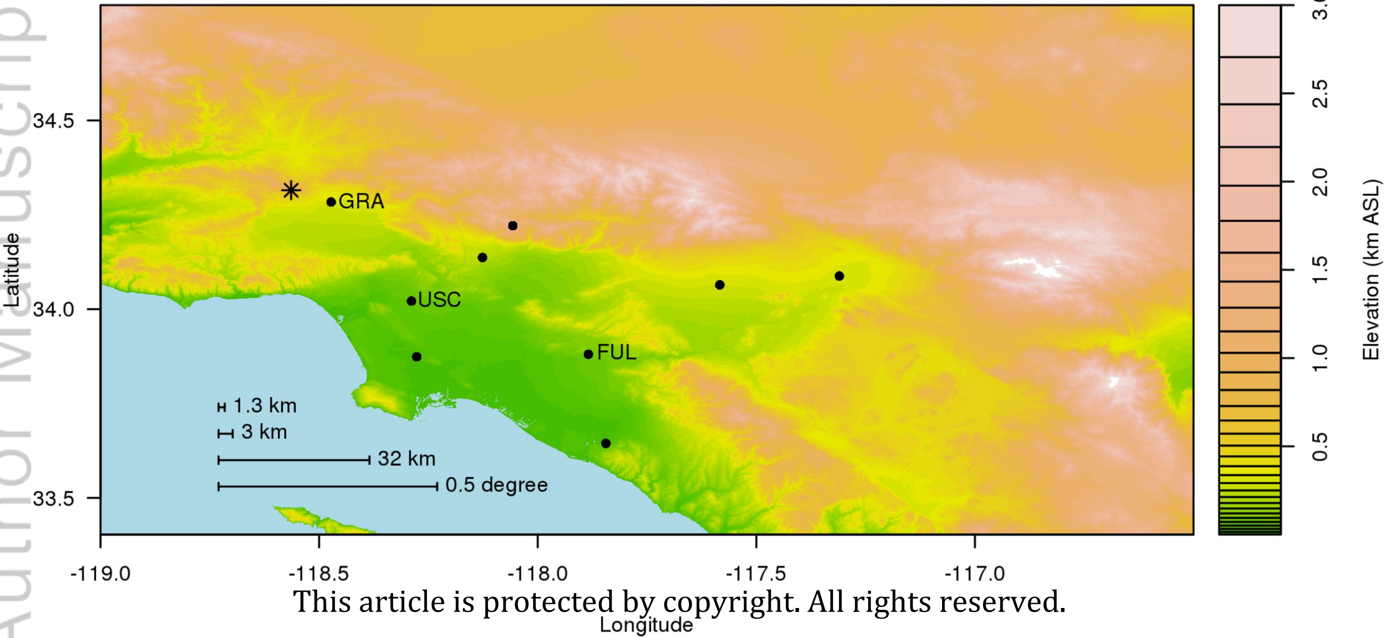

34.5
34.0
33.5
Latitude
GRA
USC
FUL
1.3 km
3 km
32 km
0.5 degree
-119.0
-118.5
-118.0
-117.5
-117.0
Longitude
3.0
2.5
2.0
1.5
1.0
0.5
Elevation (km ASL)

Figure 2.

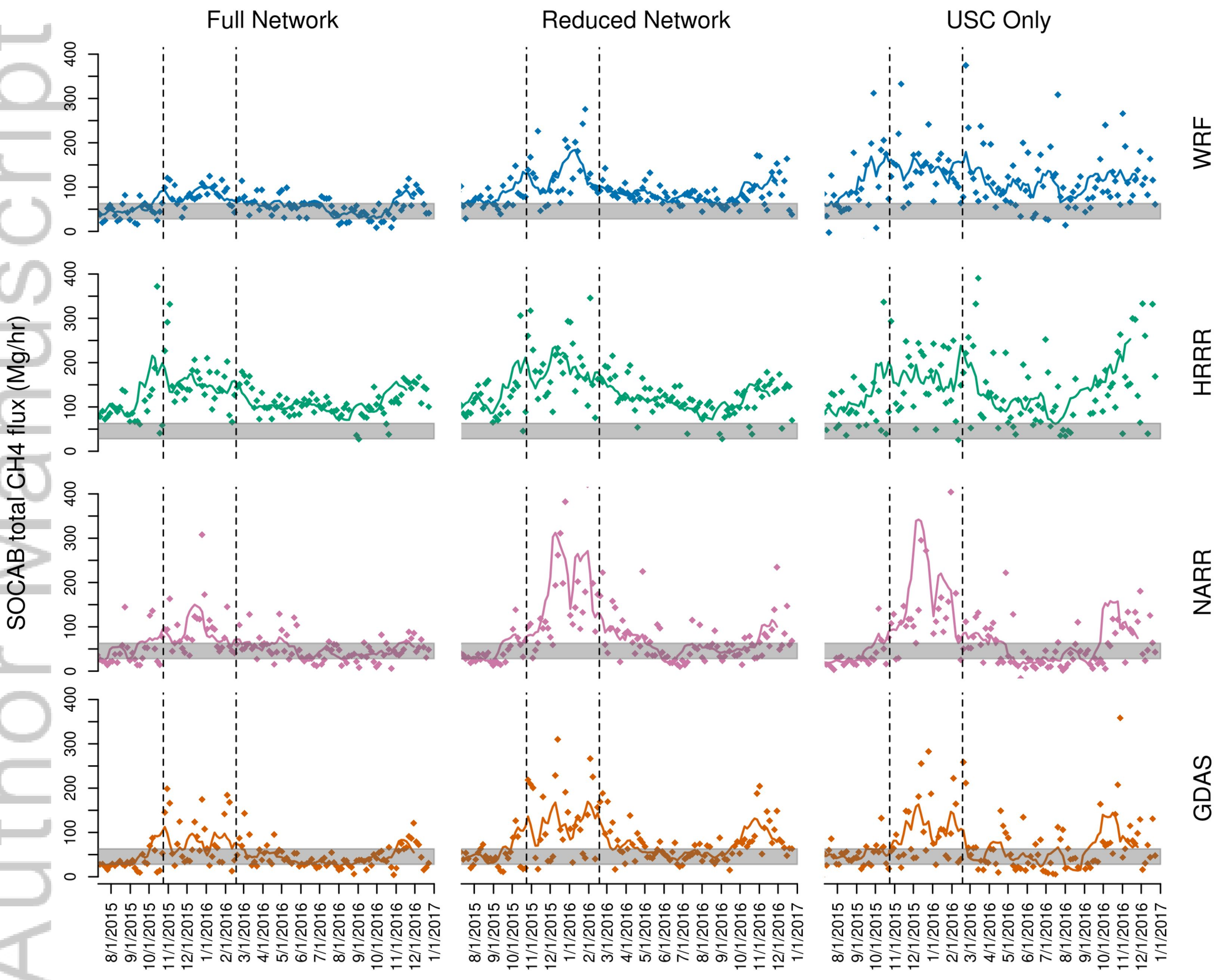
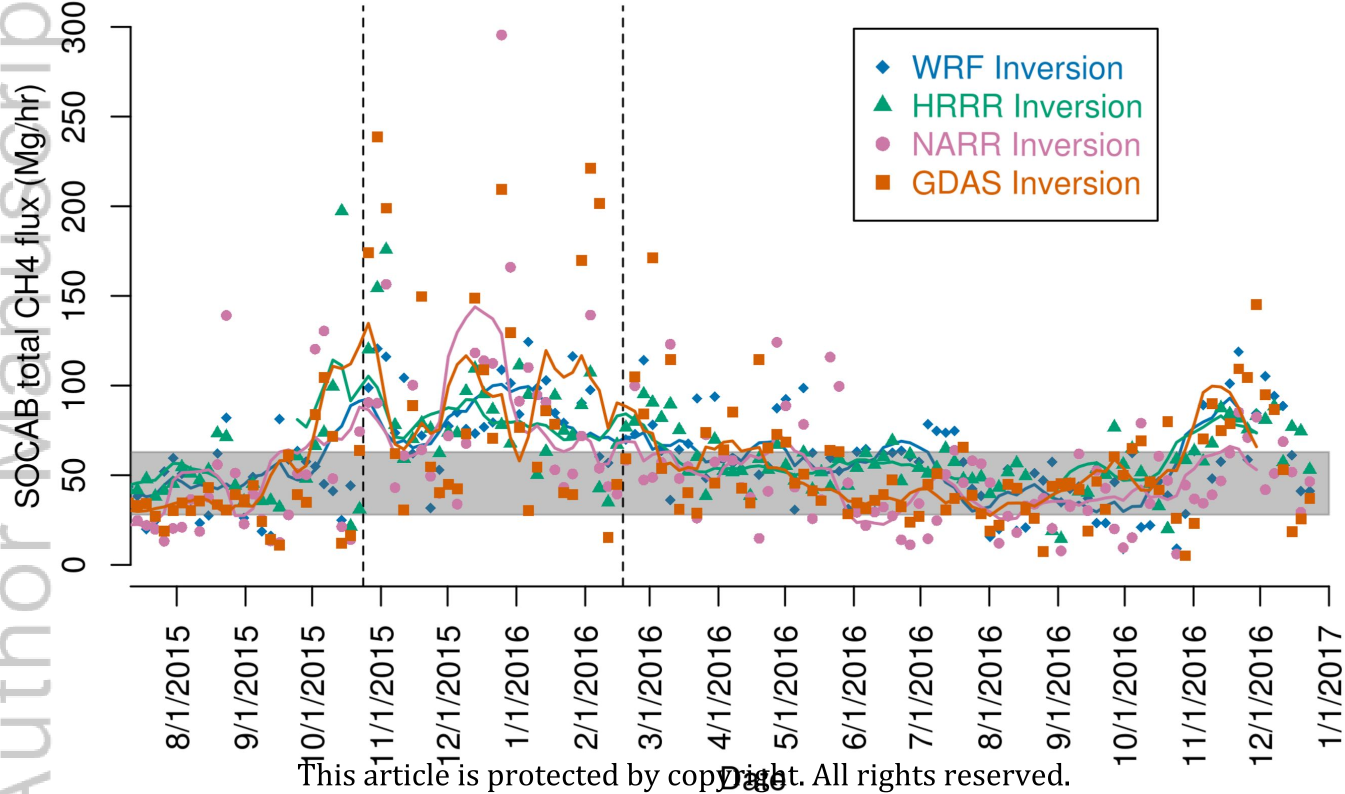

Full Network
Reduced Network
USC Only
WRF
HRRR
NARR
GDAS
SOCAB total CH4 flux (Mg/hr)
0
100
200
300
400
8/1/2015
9/1/2015
10/1/2015
11/1/2015
12/1/2015
1/1/2016
2/1/2016
3/1/2016
4/1/2016
5/1/2016
6/1/2016
7/1/2016
8/1/2016
9/1/2016
10/1/2016
11/1/2016
12/1/2016
1/1/2017
Date

Figure 3.



WRF Inversion
HRRR Inversion
NARR Inversion
GDAS Inversion
SOCAB total CH4 flux (Mg/hr)
0
50
100
150
200
250
300
8/1/2015
9/1/2015
10/1/2015
11/1/2015
12/1/2015
1/1/2016
2/1/2016
3/1/2016
4/1/2016
5/1/2016
6/1/2016
7/1/2016
8/1/2016
9/1/2016
10/1/2016
11/1/2016
12/1/2016
1/1/2017
Date

Figure 4.

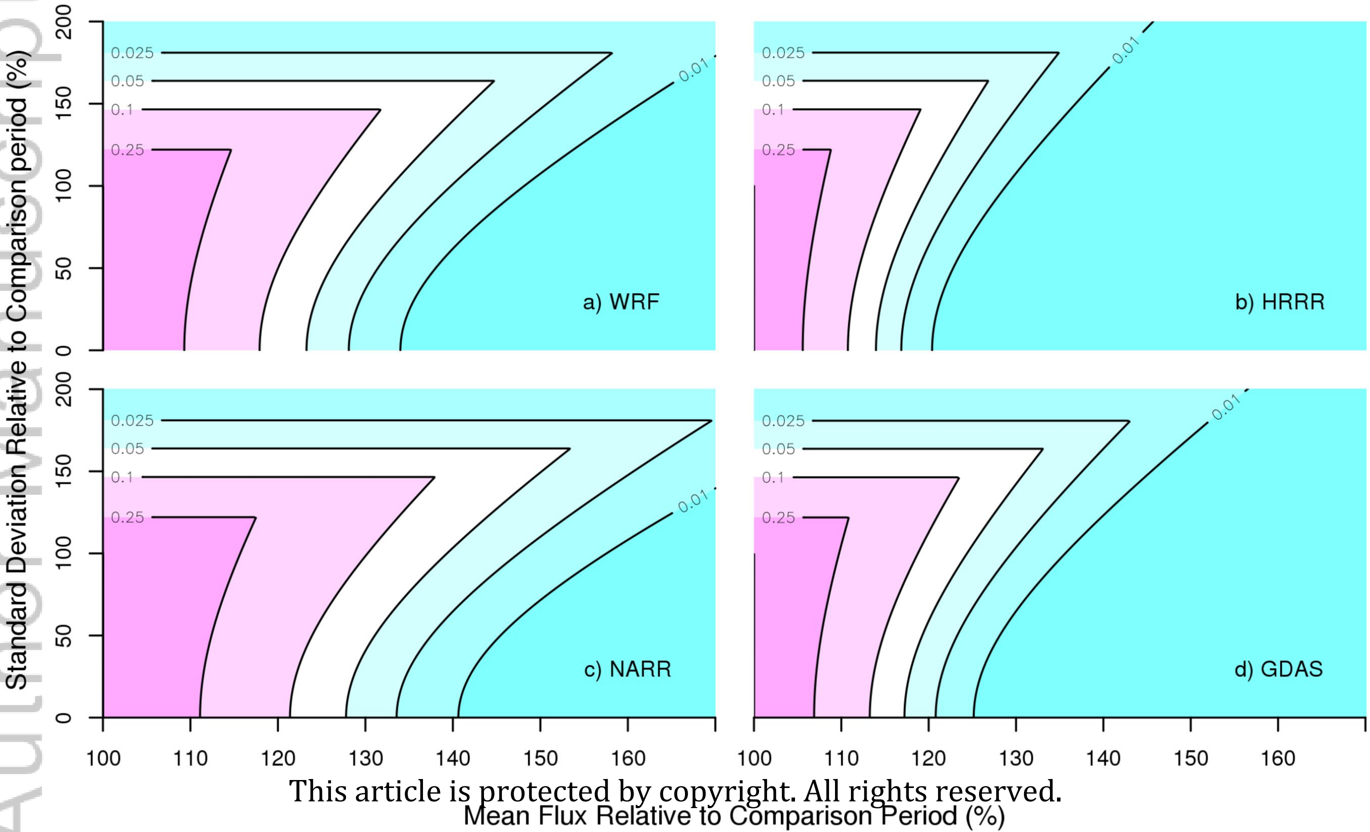
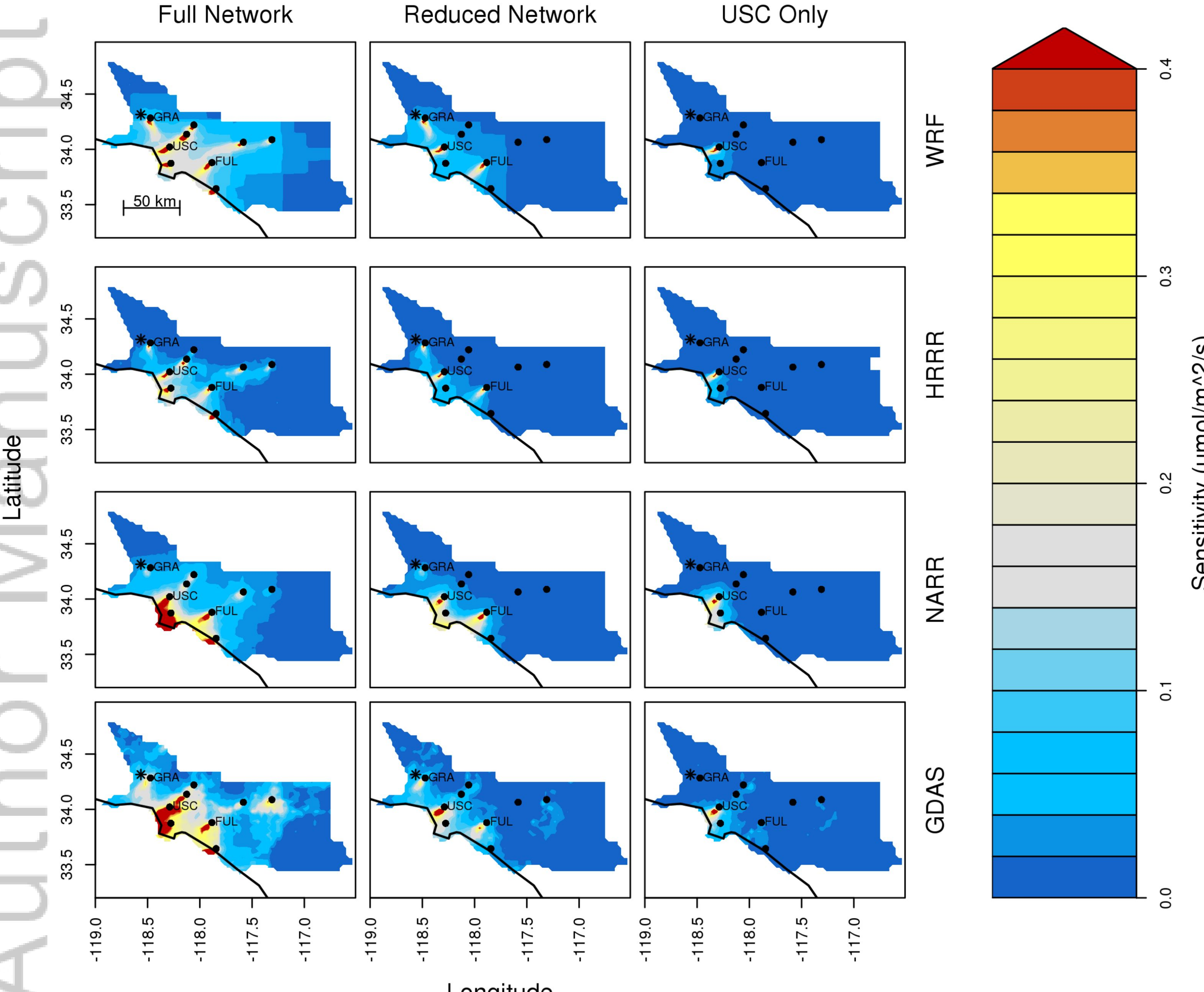
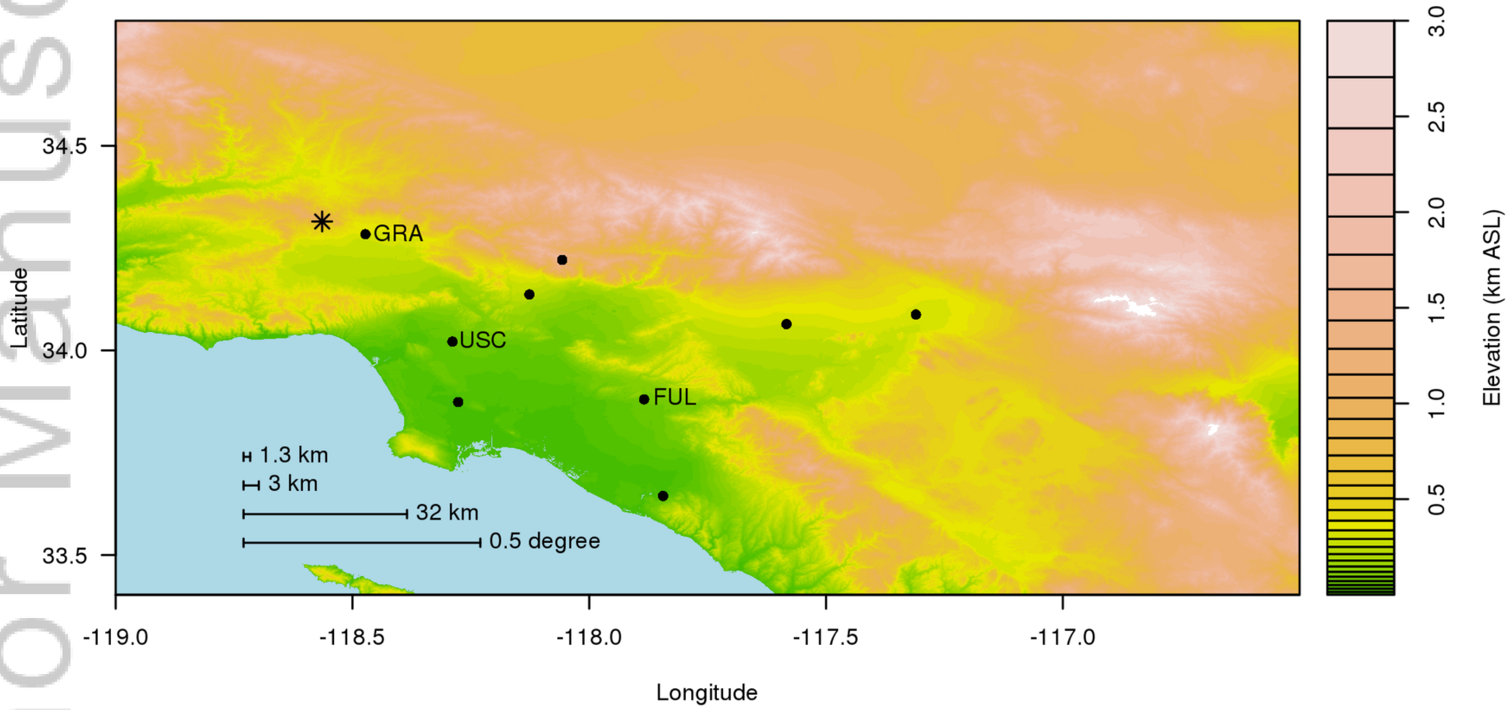

Standard Deviation Relative to Comparison period (%)
200
150
100
50
0
0.025
0.05
0.1
0.25
0.01
a) WRF
b) HRRR
c) NARR
d) GDAS
100
110
120
130
140
150
160
Mean Flux Relative to Comparison Period (%)

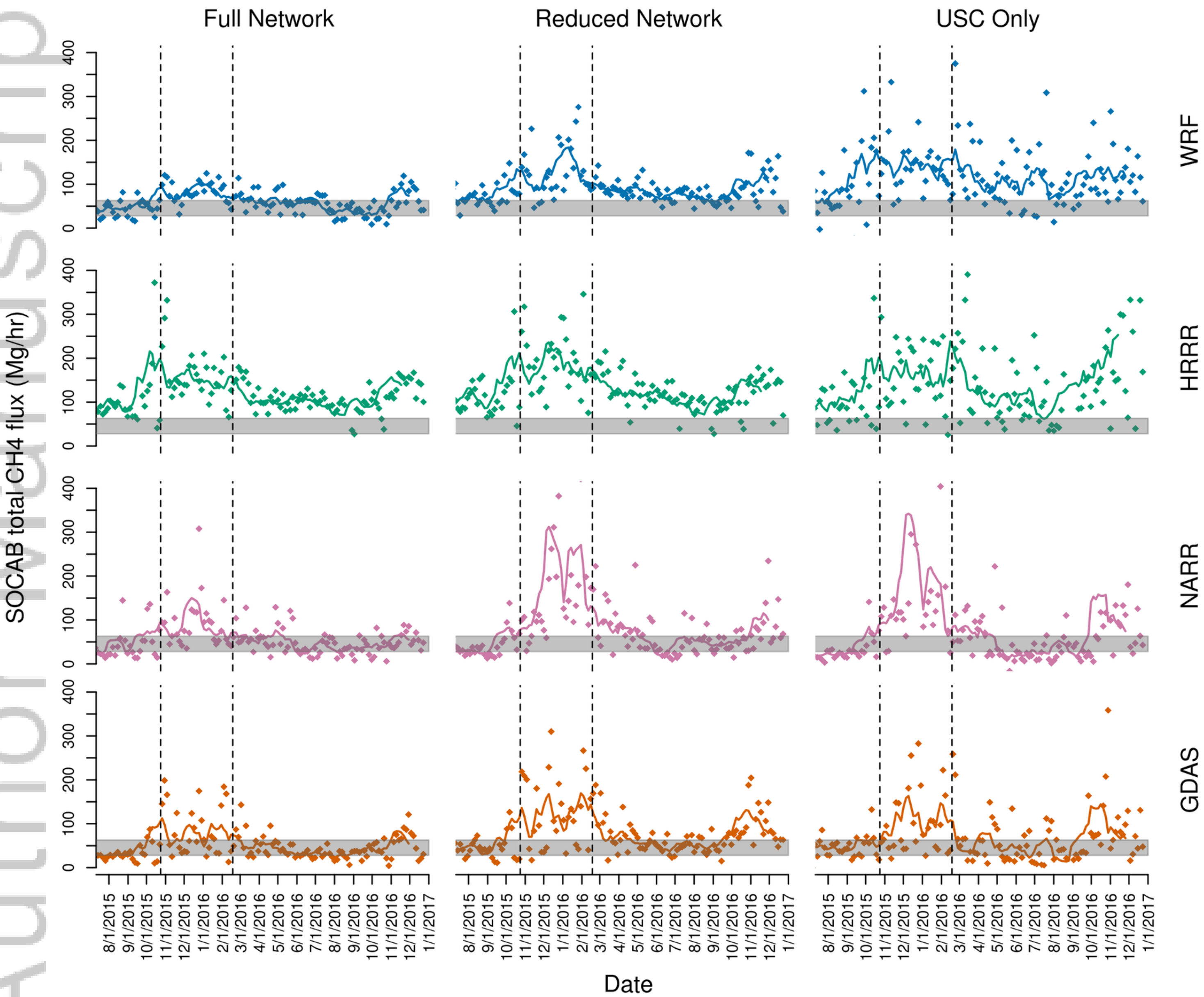
Figure 5.



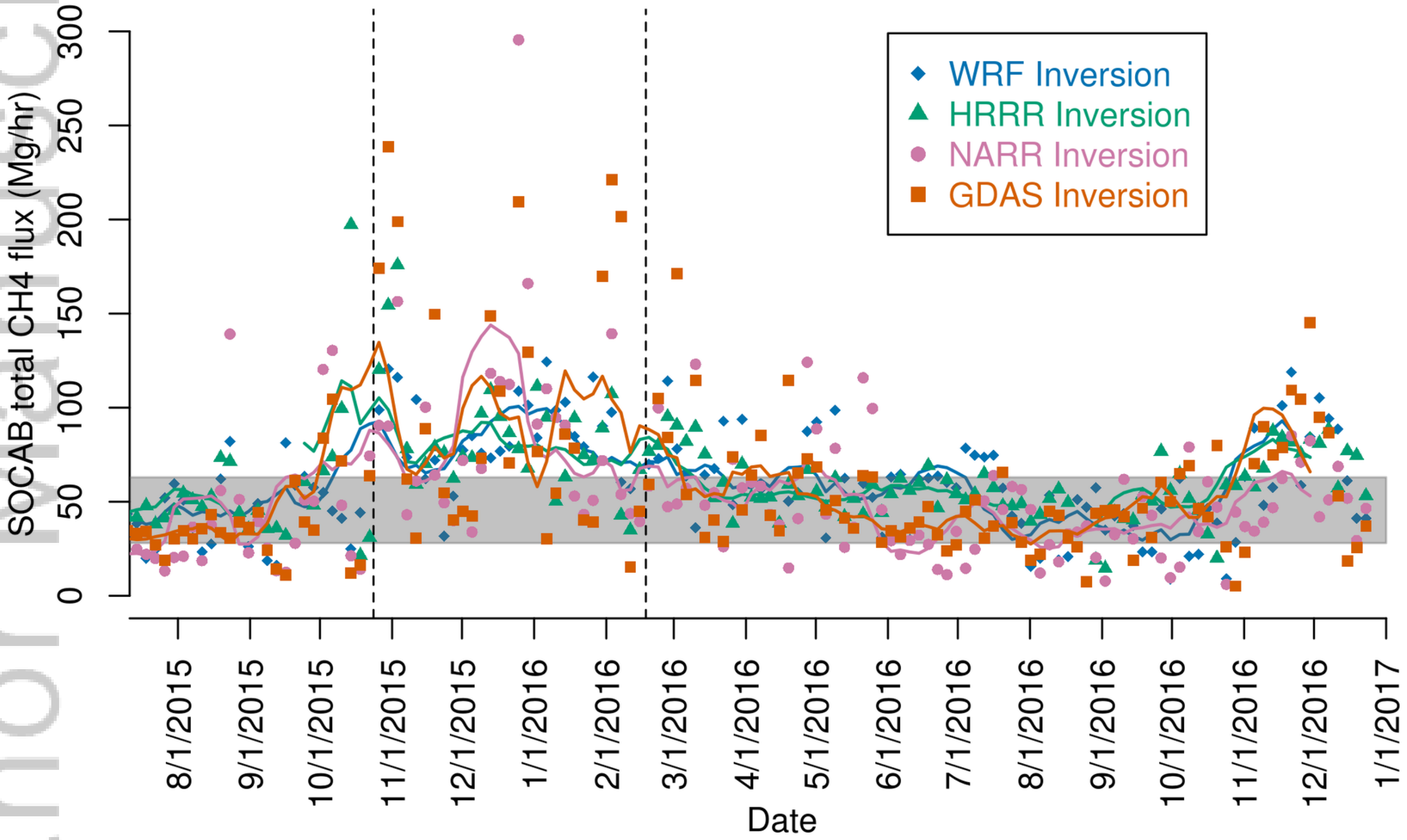
Full Network
Reduced Network
USC Only
WRF
HRRR
NARR
GDAS
GRA
USC
FUL
50 km
Latitude
Longitude
34.5
34.0
33.5
-119.0
-118.5
-118.0
-117.5
-117.0
Sensitivity (umol/m^2/s)
0.0
0.1
0.2
0.3
0.4



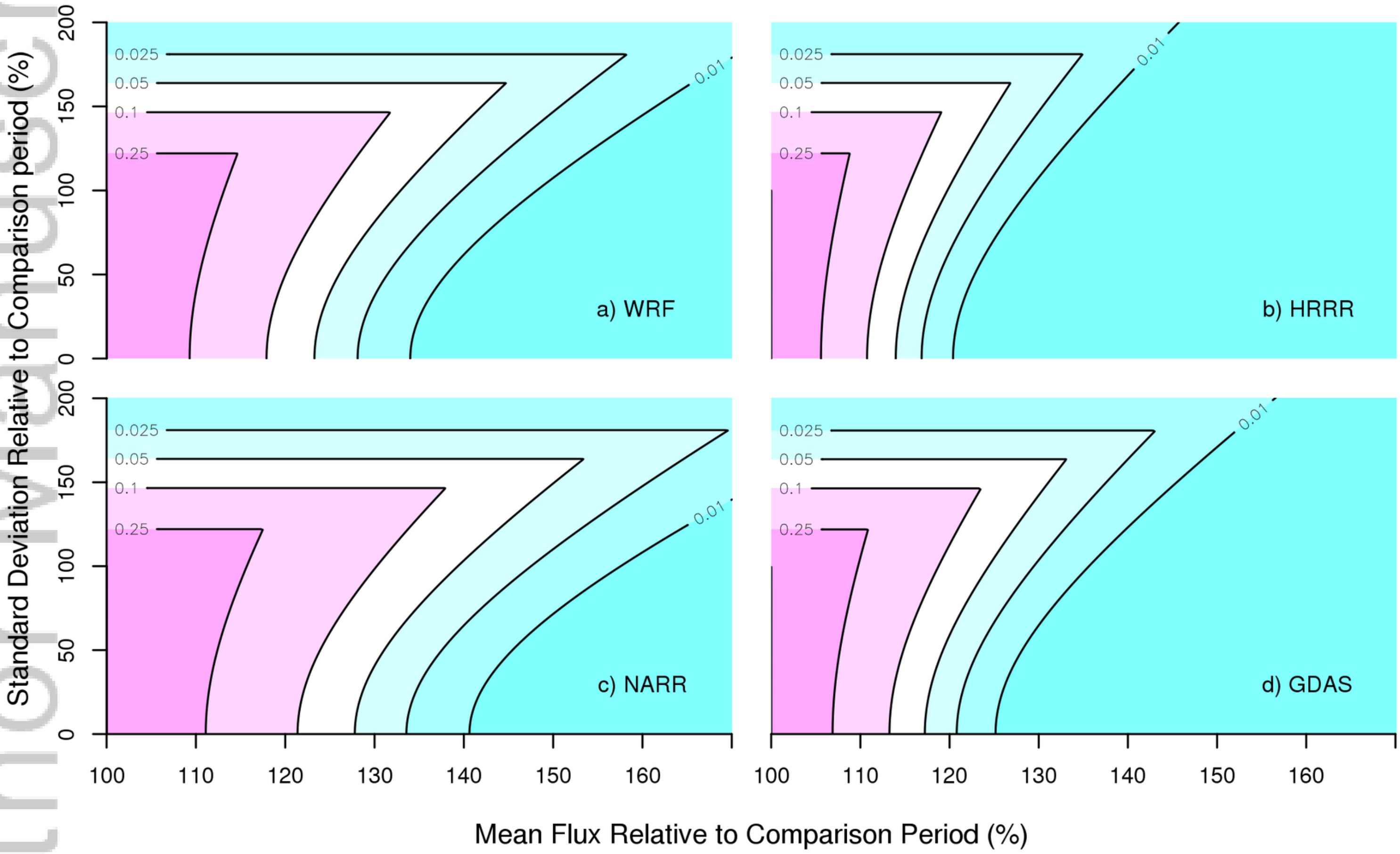
2018JD029224-f01-z-.png



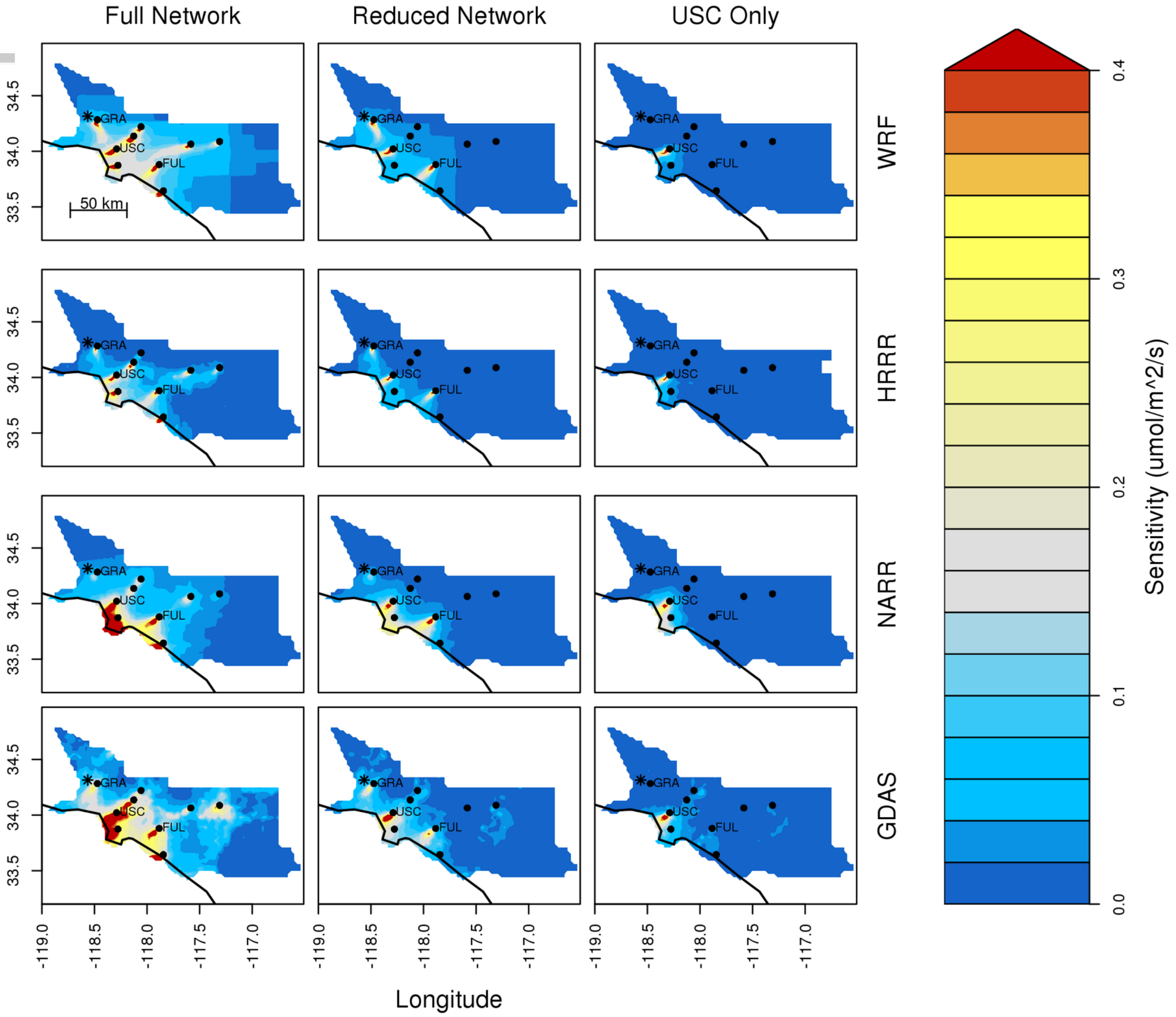
2018JD029224-f02-z-.png



2018JD029224-f03-z-.png



2018JD029224-f04-z-.png



2018JD029224-f05-z-.png