# Factors Impacting Observation-Based Estimates of Urban Greenhouse Gas Emissions

by

John Frederick Ware

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2018

Doctoral Committee:

       Professor Charles Doering, Co-Chair
       Assistant Professor Eric A. Kort, Co-Chair
       Professor August Evrard
       Associate Professor Mark Flanner
       Professor David Gerdes

TEACHERS STUDENTS RESEARCHERS WORKERS

JOIN GEO

John Frederick Ware

johnware@umich.edu

ORCID iD: 0000-0003-2789-8351

drops of water turn a mill

singly none

# ACKNOWLEDGEMENTS

I could not possibly have done the work represented here without a supportive and loving community. To adequately express the indispensability of the support I received, and the depth of my gratitude, feels like an impossible task. I only hope that the people who've helped along the way will be patient with me one more time as I try my best to thank them.

Hannah Maier has heard, held, fed, accompanied, encouraged, forgiven, accepted, and loved me. The past few months especially, she carried every bit of weight I knew how to hand her. Hannah makes me happy and, most importantly, she makes me better. My mom wrote her own dissertation the year I finished college. She having been through it with the personality she and I share, and my dad having cared for her as she did, my parents understood my experience in a way not many people could – so when they believed I could do it, I believed too. Meryl Spencer promised she would see me through and kept her promise, foster-parenting my fears when I couldn't take care of them on my own. Everyone who worked alongside me helped make this pursuit feel a little bit less solitary; Jon Curtiss and Bryan Ramson were particularly consistent and compassionate companions. Among other lessons, my therapist Eileen helped me learn to accept the incredible generosity of the people in my life, to believe that they could help and to let them.

Eric Kort was the most thoughtful mentor I could have asked for, clearly putting great care into tailoring his advising to my needs. He was persistent but understanding and a constant source of excitement for our work. Eric accepted and valued me as

a whole person at the same time that he took seriously his charge to train me as an independent scientist. I'm grateful to the physics department, especially to my first advisor, Ratindranath Akhoury, for supporting my decision to go outside the department to work with Eric. The members of Eric's group and of the Megacities Carbon Project provided data, code, advice, ideas, and instruction. Kim Mueller taught me much of what I know about greenhouse gas inversion and kept me honest as I put it into practice, and she and Vineet Yadav shared their work for me to build upon. In addition, I relied on the labor of many others at the university, including IT, facilities, grounds, and custodial staff, frontline administrators, and medical and wellness personnel.

Finally: the officers, stewards, activists, members, and staff of my union, the Graduate Employees' Organization, and our allies in the community and in the labor movement, gave me a purpose that has changed the direction of my life. They, and our work together, kept me going when research left me anxious or discouraged. At the union office, I could always find passionate, thoughtful, committed people and the opportunity to help build something worthwhile. Science and movement-building share the sense of making small attempts every day to contribute to a greater project. I hope that I have been a good steward of the projects I have taken up, and that those who carry them on after me may do so with joy and perseverance.

Thank you, all.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ABSTRACT

Urban areas are responsible for a large and increasing fraction of anthropogenic greenhouse gas emissions. Accurate methods for quantifying and monitoring those emissions are needed to suggest and evaluate mitigation policies, as well as for fundamental carbon cycle science as anthropogenic carbon dioxide emissions become a dominant source of uncertainty in closing the global carbon budget. I present investigations into several factors that can impact our ability to characterize urban greenhouse gas emissions using observations in the atmosphere. An automated method is developed for estimating the mixing depth, a key meteorological variable affecting the sensitivity of mole fraction observations to emissions fluxes, using optical remote sensing instruments. In a long time series of mixing depth estimates in Pasadena, California, day-to-day variability is shown to be large in comparison to seasonal trends. Significant mixing depth biases are demonstrated in meteorological models, and the likely impacts on emissions estimation are discussed.

Optimized estimates of methane emissions in the South Coast Air Basin, California, are made using several flux inversion or regularization methods, with four sources of meteorological information, and with all or some of the mole fraction observations taken at nine within-basin observing sites associated with the LA Megacities Carbon Project. Using the full observational dataset in a geostatistical inversion, the capability to detect seasonal and event-driven emissions changes is demonstrated with generic meteorology, opening the door to near-real-time monitoring. Differences in absolute methane emissions flux magnitude according to the source of driving mete-

orological information are shown to be largely removable by calibration to a trusted model. The choice of inversion or regularization method is shown to have substantial impacts both on the estimated emissions time series and on the capacity to detect emissions changes, especially when the observational constraint is reduced.

# CHAPTER I

# Introduction

The work presented here focuses on *factors impacting observation-based estimates of urban greenhouse gas emissions*. In this introductory chapter, I provide the necessary background to the more detailed discussions in chapters II, III, and IV. Section 1.1 lays out the motivation for studying urban greenhouse gas emissions. Section 1.2 explains the distinction between atmospheric observation-based estimates and those relying on source inventories, including some of the advantages and disadvantages of each approach. The next two sections discuss two of the difficulties affecting observation-based estimates, and the tools used to face those difficulties, in more detail: Section 1.4 covers underdetermination and statistical inverse methods, and Section 1.3 covers atmospheric transport and transport modeling. In Section 1.5, I discuss the context of my work within the Los Angeles Megacities Carbon Project and provide background on the 2015-2016 Aliso Canyon natural gas leak, which is used in Chapters III and IV as a natural experiment. Finally, Section 1.6 provides an overview of the remainder of the dissertation.

## 1.1    Motivation

Urban areas are responsible for a large and increasing share of the world's greenhouse gas emissions: more than 70% of energy-related emissions as of 2010 (*Rosen-*

*zweig et al.*, 2010). Cities are also centers of effort to mitigate climate change, with many acting more aggressively than nation-states or the international community (*Rosenzweig et al.*, 2010). In the US, a number of cities have developed Climate Action Plans, setting goals and specifying actions to be taken to reduce their impacts on the climate (*Aznar et al.*, 2015).

These mitigation efforts will be more successful if they are informed by a clear understanding of urban emissions. We need to know what source of greenhouse gases within cities are most important, which are growing and how they are changing, and what policy interventions are likely to be effective – and we need to be able to track interventions as they are implemented, to measure their effectiveness in practice. At a minimum, jurisdictions that set emissions targets need to know whether they are on pace to meet those targets. As I discuss in chapter III, we may be able to track relative changes over time with better precision than we can quantify absolute emissions fluxes. Verification may therefore be especially feasible for the many mitigation goals which call for a certain percentage decrease in emissions relative to a baseline year.

Continuous monitoring of urban emissions can also contribute directly to emissions mitigation. *Duren and Miller* (2012) have proposed a 'tiered' observing system in which a broad network of sensors across an urban area is used to establish a baseline and to detect emissions changes over a wide area. When anomalous emissions are detected, aircraft or vehicle-based sensors can be deployed to locate and characterize the source and inform the relevant stakeholders. Identification of leaks in the natural gas distribution system is one especially promising possibility.

A better understanding of urban greenhouse gas emissions would also have basic scientific value. For example, the net exchange $N_L$ of $CO_2$ between the atmosphere and the terrestrial biosphere is typically calculated as the residual required to close

2

the total atmospheric carbon budget:

$$N_L = \frac{\mathrm{d}C}{\mathrm{d}t} - N_O - E_F - E_L \tag{1.1}$$

where $C$ is the carbon content of the atmosphere, $N_O$ is the net exchange with the oceans, and $E_F$ and $E_L$ are the rates of anthropogenic emissions due, respectively, to fossil fuel combustion and land use change. As *Ballantyne et al.* (2015) have shown, while the proliferation of global monitoring sites has greatly reduced the uncertainty in $\mathrm{d}C/\mathrm{d}t$ over the past several decades, that improvement has been largely offset by increased uncertainty in fossil fuel emissions. As a result, $N_L$ is barely better constrained today than it was fifty years ago.

As the predominant sources of fossil fuel emissions shift to developing countries (and become increasingly concentrated in the world's most rapidly growing megacities), total emissions uncertainty is likely to continue to increase. We can therefore expect poorly characterized anthropogenic $CO_2$ emissions to remain an important factor inhibiting study of the biogeochemical carbon cycle. The same emissions driving global climate change are also obscuring the ecosystem response to that change.

## 1.2   Methods for Urban Greenhouse Gas Quantification

Broadly speaking, greenhouse gas emissions can be quantified in two ways: with "bottom-up" source inventories or with "top-down" atmospheric observations. Inventories make use of information about the sources of emissions, including reported emissions from industrial facilities, traffic data, population density, electricity and heating fuel use, etc. Recently, some inventories have moved from global (e.g., *Oda and Maksyutov*, 2011; *Crippa et al.*, 2016) or regional scales (e.g., *Gurney et al.*, 2009; *Maasakkers et al.*, 2016) to very detailed (as fine as building-scale) representations of urban areas (*Gurney et al.*, 2012).

The inventory approach can be direct, but it relies on estimates of emissions from the typical source (vehicle, fixture, facility, etc.) of a particular kind. Such estimates may be unrepresentative if the distribution of emissions is fat-tailed (emissions are dominated by a small subset of sources). For example, inventories systematically underestimate methane emissions, seemingly because a substantial fraction of emissions come from a small number of high-emitting fixtures, wells, facilities, etc. (*Brandt et al.*, 2014). Inventories can also fail if local sources in a particular area differ systematically from what is typical, or if source characteristics change over time in a way that is not accounted for.

The "top-down" observational approach to characterizing emissions is to measure the concentrations of those gases in the nearby atmosphere. A variety of measurement technologies can be employed. In urban environments these have included, to give a few examples: in situ spectroscopic instruments mounted on towers or rooftops (*McKain et al.*, 2012; *Breon et al.*, 2014; *McKain et al.*, 2015; *Richardson et al.*, 2016; *Pugliese*, 2017) or carried on research aircraft (*Mays et al.*, 2009; *Ryerson et al.*, 2013), low-cost chemical sensors (*Shusterman et al.*, 2016), flask capture of air samples for later isotopic analysis (*Clark-Thorne and Yapp*, 2003; *Pataki et al.*, 2003; *Newman et al.*, 2008) and remote sensing of atmospheric column abundance with terrestrial upward-pointing (*Wunch et al.*, 2009) or slant-path (*Wong et al.*, 2016) as well as satellite-borne (*Kort et al.*, 2012; *Worden et al.*, 2012) instruments.

Observational methods have the advantage that gases are measured even if their sources are unknown or unusual. However, the measured concentration is at a remove from the quantity of interest, emissions flux from the surface, which introduces two important difficulties. First, the relationship between flux and concentration is mediated by the wind-driven transport of the emitted species through the atmosphere. Atmospheric transport is a complex process involving motion at multiple scales, and models of transport are uncertain and difficult to validate. Second, even

if the flux-concentration relationship is assumed to be known, the available concentration observations typically do not suffice to uniquely determine the fluxes on the spatial and temporal scales of interest.

## 1.3   Atmospheric Transport

In some cases, fluxes from a region of interest can be inferred from concentration measurements using a mass balance approach. The simplest form of mass balance uses a one-box model: concentrations are measured upwind and downwind of an area of interest, or box, and the difference is attributed to sources within the box. There are two relevant meteorological parameters, which are assumed constant: the horizontal wind speed $v$ and the mixing depth $z_i$. The wind speed and the size of the box in the direction parallel to the wind together determine the ventilation time. The lower the wind speed, the more time an air mass takes to cross the box and so the greater the concentration increase for a given emissions flux.

The mixing depth is the vertical extent of the mixing layer, which is the part of the atmosphere in regular contact with the surface on a time scale of about an hour. Vertical eddies distribute atmospheric constituents, including gases and particles emitted from the surface, through the mixing layer. Mixing depths can vary from less than 100 m at night, when mixing is driven only by the interaction of horizontal wind with the rough surface, to as much as several kilometers when a dry surface is heated by solar radiation. Assuming that the ventilation time is long enough for the air to be well mixed, the mixing depth determines the height of the box in a box model. The taller the mixing layer, the more diffuse the emitted species.

In sum, the observed difference in concentration between the upwind and downwind faces of a homogeneous box of length $L$ in the direction parallel to the horizontal

wind, containing a surface emissions flux $s$, is

$$\Delta C = \frac{m_{air}}{z_i \rho} \frac{L}{v} s \tag{1.2}$$

where $m_{air}$ and $\rho$, the molar mass and density of air, are needed to convert the concentration to a unitless mixing ratio (e.g. parts per million) as is typically measured.

The mass balance approach is appropriate when the sources of interest are either isolated, such as a remote oil field, or approximately homogeneous, such as a wilderness. In addition, the local meteorology must be fairly simple: both the wind speed and direction and the mixing depth must be constant over the extent of the box. Some cities meet these criteria at least some of the time. For example, Indianapolis, Indiana is surrounded by cropland in all directions and is located in a flat area in which the wind is often from the west. Mass balance estimates of total $CO_2$ and $CH_4$ emissions from Indianapolis have been made using concentration observations from aircraft (*Mays et al.*, 2009; *Cambaliza et al.*, 2015).

Many urban environments, however, are not amenable to mass balance. Large point sources within cities, such as power plants and landfills, are typically surrounded by other significant but nonuniform emitters: other industrial facilities, road networks, residential and commercial buildings, natural gas distribution infrastructure, and so on. Even cities considered as a whole often have indistinct boundaries, transitioning continuously into other settled areas. The local meteorology may also be complex. For example, especially in cities surrounded by mountains (e.g. Salt Lake City, see *McKain et al.*, 2012), an 'urban dome' may form. Air can remain trapped for an extended period, accumulating emitted species, before being flushed out.

The work presented here focuses on the Los Angeles area, in which the proximity of both the mountains to the east and the Pacific coast to the west generally results in a sea breeze pattern of circulation (*Lu and Turco*, 1994, 1995). During the day,

the temperature contrast between the land and ocean surface drives onshore winds. Air flowing upslope may have enough energy to continue over the mountains onto the continent, or may reach a maximum altitude and then return back over the city as an elevated layer. Other patterns are also sometimes observed; the most important is the Santa Ana condition, in which hot, dry winds flow down into the city from the desert to the east.

### 1.3.1 Lagrangian Modeling

In order to infer emissions fluxes from concentrations in such a complex environment, physically realistic models of atmospheric transport are employed. In this work, we use the Lagrangian transport model STILT (Stochastic Time-Inverted Lagrangian Transport) (*Lin et al.*, 2003). In constrast to Eulerian models, which compute variables on a grid, Lagrangian models follow air masses or parcels as they move through the atmosphere. The modeled motion is driven by a precomputed four-dimensional field of wind velocities and other meteorological variables. Because the field of driving meteorology cannot resolve the fine-scale turbulence, motion in the transport model has both a deterministic (advective) and a stochastic component. STILT separates the parcel velocity $\mathbf{u}$ accordingly as

$$\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}' \tag{1.3}$$

where $\bar{\mathbf{u}}$ is the mean wind velocity (here, the grid-scale wind velocity from the precomputed field) and $\mathbf{u}'$ is the stochastic part of the velocity. The stochastic velocity in each direction is modeled as a Markov process:

$$u_i'(t + \Delta t) = R(\Delta t)u_i'(t) + \left[1 - R^2(\Delta t)\right]^{1/2} \mathrm{N}(0, \sigma_i) \tag{1.4}$$

where $\Delta t$ is the model time step, $N(0, \sigma_i)$ is a Gaussian random velocity with standard deviation $\sigma_i$, and

$$R(\Delta t) = \exp(-\Delta t / T_{Li}) \tag{1.5}$$

describes the exponential decay in autocorrelation of the stochastic velocity over the Lagrangian time scale $T_{Li}$. The stochastic velocity varies randomly over times much longer than $T_{Li}$ ($R \approx 0$) and is constant over short times ($R \approx 1$).

The Lagrangian timescales $T_{Li}$ and turbulent velocity scales $\sigma_i$ cannot be predicted from first principles. Instead, empirical formulas are developed using a form of dimensional analysis called similarity theory. A brief description is given here; for a more complete overview, see *Stull* (1988). Variables expected to be relevant are identified and formed into dimensionless combinations, called dimensionless groups. Often, power laws or other simple relationships can be fit to the observed behavior of the dimensionless groups. The mixing depth $z_i$ itself is generally the most important length scale; since the mixing layer is defined by the presence of turbulent mixing, its extent is determined by the same turbulent processes that control parcel motion. Other variables relevant to turbulent motion depend on the dominant source of energy for turbulence, which may be either buoyant forces or (wind) shear.

Buoyancy drives vertical motion in the atmosphere when less dense air is present below more dense air. Because density is difficult to measure directly, buoyancy is generally assessed using a proxy variable called potential temperature, defined as

$$\theta = T \left( \frac{P}{P_0} \right)^{R/c_p} \tag{1.6}$$

where $T$ and $P$ are the temperature and pressure of an air parcel, $P_0$ is a reference pressure (usually surface pressure or 1 atmosphere), $R$ is the ideal gas constant, and $c_p$ is the specific heat of air. The potential temperature is the temperature the parcel would come to if transported adiabatically to pressure $P_0$. Correcting for the density

effect of the water content of humid air gives the virtual potential temperature $\theta_v$,

$$\theta_v = \theta(1 + 0.61r - r_L) \tag{1.7}$$

where $r$ is the mixing ratio of water vapor and $r_L$ is the mixing ratio of liquid water.

If virtual potential temperature decreases with height, the vertical stratification is out of equilibrium, which tends to cause convective mixing. In the mixing layer, solar heating of the land surface causes high (virtual) potential temperatures near the ground. The vertical velocity scale associated with the resulting buoyancy-driven motion is the convective velocity $w_*$,

$$w_* = \left[ (g/\overline{\theta_v})\overline{w'\theta'_{vs}}z_i \right]^{1/3} \tag{1.8}$$

where $g$ is acceleration due to gravity, $w$ is the vertical component of wind velocity. Here and in general, a prime denotes the stochastic component of a variable, an overbar indicates an average, and the subscript $s$ indicates the value at the surface. The quantity $\overline{w'\theta'}_{vs}$ is the upward flux of virtual potential temperature from the surface, which is generally the dominant source of energy for turbulence in the mixing layer during the day, so $w_*$ is a relevant variable in similarity theory under those conditions.

The second major source of turbulence in the mixing layer is wind shear, especially at the surface. The relevant velocity scale is the friction velocity $u_*$,

$$u_* = \left[ \overline{u'w'}^2_s + \overline{v'w'}^2_s \right]^{1/4} \tag{1.9}$$

where $u$ and $v$ are the east and north components of the wind velocity. The terms inside the square brackets express the surface flux of the horizontal components of momentum, caused by the frictional effect of the ground as air passes over it. If

both shear and buoyant effects are relevant, their relative importance can be included through the Obukhov length $L_{ob}$,

$$L_{ob} = -\frac{z_i u_*^3}{k w_*^3} \qquad (1.10)$$

where the dimensionless von Kármán constant $k$, relevant in other applications of the Obhukov length, is about 0.4. Empirical functions expressing Lagrangian timescales and turbulent velocity scales in terms of $w_*$, $u_*$, $L_{ob}$, and $z_i$ are presented in *Hanna* (1984).

### 1.3.2 Time-Inverted Transport

Development of large-scale (kilometers and up) Lagrangian transport modeling was originally motivated by the need to predict the dispersal of hazardous materials from point sources, especially in the wake of the 1986 Chernobyl nuclear disaster (*Thomson and Wilson*, 2013). Applications have also been made to the transport of volcanic ash (e.g. *Stohl et al.*, 2011; *Webster et al.*, 2012), insects (*Burgin et al.*, 2013), pathogens (*Gloster et al.*, 2010), chemical toxins (*Kinra et al.*, 2005), and so on.

In the above examples, the point of origin (source) is known in advance and the goal is to determine all downwind exposures. For our application, the opposite is true. We are interested only in air masses that ultimately pass over the measurement location, and the goal is to determine the upwind sources of those air masses. If we modeled the dispersal of emissions from every possible source point in our region of interest, most of the results would be irrelevant, because most of the modeled parcels would never be measured. Instead, we run the transport model in a time-reversed mode, tracking simulated air parcels backwards in time from the observation point. *Flesch et al.* (1995) showed that the flux-concentration relationship derived

from a time-reversed model is the same as that derived in forward time as long as the model satisfies the "well-mixed criterion," a consequence of the second law of thermodynamics.

In order to sample the stochastic part of the motion, a large number of air parcels are simulated backwards in time from each observation point. The sensitivity of a given observation $i$ to emissions fluxes is given (*Lin et al.*, 2003) by applying the reasoning of the earlier box model to each region $j$ of the surface:

$$\Delta C_i = \sum_j H_{ij} s_j, \qquad H_{ij} = \frac{m_{air}}{z_j \rho_j} \frac{1}{N_i} \sum_{p_i=1}^{N_i} \Delta t_{p_i,j} \qquad (1.11)$$

where $N_i$ is the total number of parcels and $\Delta t_{p_i,j}$ is the residence time of parcel $p_i$ within the mixing layer[1] above region $j$. Despite the complicated dynamics of the transport, the flux-concentration relationship is linear in the fluxes and can be encoded as a matrix $\mathbf{H}$.

Typically, the Lagrangian transport model itself does not predict the driving meteorology, which must be computed in advance by a separate model. Such models vary in horizontal, vertical, and temporal resolution and in the domain covered. Some are run operationally, on a regular schedule, with the results archived for public use; others must be configured for a particular application and run by the individual researcher or team. Most employ both a dynamical representation of the relevant physical processes and at least some weather observations, but some are better described as using the observations to nudge or initialize the dynamics while others use the dynamics to interpolate the observations. The choice of driving meteorology for transport is a main subject of chapter III.

---

[1]Or, as in STILT, within some specified fraction of the mixing layer, by default the lower half. So long as the mixing layer is truly well-mixed, the result is the same: half as many parcels are considered to be sensitive to the surface, but each sees the flux diluted into half the volume.

## 1.4 Flux Inversion

*Flux inversion* is the process of generating an optimal estimate of emissions flux by combining atmospheric observations with some additional information. The additional information may be either a prior flux estimate, usually from an inventory (in a Bayesian inversion), or one or more models of flux patterns (in a geostatistical inversion). In both cases, the goal is to maximize the probability of the fluxes given the observed concentrations. By Bayes' Theorem,

$$p(\mathbf{s}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{s})p(\mathbf{s}) \tag{1.12}$$

where $\mathbf{z}$ is a vector of $n$ observed concentrations (not heights[2]) and $\mathbf{s}$ is a vector of $m$ fluxes to be estimated. Both $p(\mathbf{z}|\mathbf{s})$ and $p(\mathbf{s})$ are generally assumed to be Gaussian[3].

The conditional probability of the observations takes the form:

$$p(\mathbf{z}|\mathbf{s}) \propto \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{H}\mathbf{s})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s})\right] \tag{1.13}$$

where the Jacobian $\mathbf{H}$ encodes the transformation from flux space to concentration space as computed by the transport model. It acts on the fluxes to produce the modeled observations $\mathbf{H}\mathbf{s}$, which are the concentrations that would be recorded if the emitted gas were carried forward to the observation site and measured without error. The conditional probability $p(\mathbf{z}|\mathbf{s})$ is maximized when the fluxes are such that the forward-modeled observations are close to the actual observations. Deviations between the actual observations and the modeled observations are weighted according to the model-data mismatch covariance matrix $\mathbf{R}$.

---

[2]The reuse of notation allows $z_i$ and $\mathbf{z}$ to retain their standard meanings in Chapters II and III, respectively.

[3]The assumption of Gaussian errors is not unproblematic, but it is common because it allows the probability to be maximized algebraically. Numerical methods, like the Monte Carlo technique used for the transdimensional hierarchical inversion in Chapter IV, can accommodate non-Gaussian error distributions.

The prior probability of the fluxes has the form:

$$p(\mathbf{s}) \propto \exp\left[-\frac{1}{2}(\mathbf{s} - \mathbf{s}_0)^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{s}_0)\right] \tag{1.14}$$

In a Bayesian inversion, $\mathbf{s}_0$ is a fixed prior estimate of the fluxes, which may include estimates from inventories of anthropogenic emissions and/or models of the biosphere. In a geostatistical inversion, we set $\mathbf{s}_0 = \mathbf{X}\beta$, a linear combination of $p$ variables with unknown "drift components" $\beta$.[4] Each column of $\mathbf{X}$ encodes some variable, called a model component or covariate, which is expected to be proportional to part of all of the emissions flux. A covariate may itself be an emissions estimate, or it may be some other explanatory factor such as population density, economic data, vegetation (for biogenic emissions), etc.

The drift components $\beta$ are taken to be unconstrained a priori, so in the geostatistical case we have, in place of $p(\mathbf{s})$,

$$p(\mathbf{s}, \beta) \propto \exp\left[-\frac{1}{2}(\mathbf{s} - \mathbf{X}\beta)^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\beta)\right] \tag{1.15}$$

and the goal is to optimize both the fluxes and the drift components. This difference is important. In the Bayesian case, $\mathbf{s}_0$ is fixed and fluxes are assigned a low prior probability if they deviate too much from that prior estimate. In the geostatistical case, the scale of the prior is given by $\beta$ and is allowed to vary, so fluxes are assigned a low probability only if their pattern is poorly explained by the provided model components $\mathbf{X}$. In either case, deviations are weighted according to the prior error covariance matrix $Q$.

---

[4]The terminology is carried over from applications in hydrology.

### 1.4.1 Covariance Specification

The true statistics of the model-data mismatch (represented by $\mathbf{R}$) and of the prior estimate or covariates (represented by $\mathbf{Q}$) are likely to be nontrivial. The model-data mismatch may include a variety of different kinds of errors in the observations $\mathbf{z}$:

- pure instrument noise, which is usually small except for low-cost sensors,

- correlated instrument error, such as drift over time,

- errors in the background concentration, which can affect all of the observations in the domain, and

- representativeness errors, due to the location of point measurements that may not be typical for the local area on the scale resolved by the inversion.

In addition, because the transport is uncertain, there may be errors in the modeled observations $\mathbf{Hs}$, which should also be encoded in $\mathbf{R}$. Transport errors at different observing sites may vary together (e.g., if the mixing depth is overestimated at a location upwind of several instruments) or in an anti-correlated way (e.g., if an emissions plume blows over one site in the real atmosphere but, due to a wind direction error, over a different site in the model).

Similarly, errors in the prior may be independent, such as when the reported emissions from a power plant or other point source are incorrect or when a particular source deviates from what is typical for its category. But if a parameter in the prior is incorrect on average, or if a widely distributed source is misspecified, then the flux will have a positive correlation between grid cells. The spatial and temporal patterns of such correlation can be very complex: as an extreme example, errors in motor vehicle emissions would be expected to follow the spatial distribution of the road network and to be more pronounced on weekdays than on weekends. On the other hand, if a

source is mislocated, the prior flux will be too low at the true location and too high at the erroneous location.

Unfortunately, these complex covariance structures are generally not known in any detail, so most authors use simple forms for the covariance matrices $\mathbf{R}$ and $\mathbf{Q}$: either diagonal (neglecting correlation) or with a correlation that decays exponentially in space and/or in time. Different variances may be assigned to different observations, such as between instrument sites or according to site characteristics (e.g., lower variance for background sites). A common choice in $\mathbf{Q}$ is an uncertainty proportional to the prior flux in each grid cell, often with some minimum value so that cells assigned zero flux in the prior can still be adjusted by the inversion.

Once the structure of the covariance matrices is decided, the magnitude of the variance and the decorrelation scales, if any, can be set using a combination of expert judgment, sensitivity tests, and any estimates that may be available in the literature to constrain some contributors to the covariance. Alternatively, as we do in chapters III and IV, the covariance parameters $\theta$ can be set so as to maximize the likelihood of the observed concentrations. In the Bayesian case, the likelihood is computed using the prior fluxes (*Michalak et al.*, 2005):

$$p(\mathbf{z}|\theta, \mathbf{s}_0) \propto |\Psi|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{H}\mathbf{s}_0)^T (\Psi)^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s}_0)\right] \qquad (1.16)$$

where

$$\Psi = \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R} \qquad (1.17)$$

is the expected *total* covariance due both to variations of the fluxes around $\mathbf{s}_0$ and to variations of the observations. Note that the normalizing determinant factor, which we have omitted elsewhere, is relevant here because $\Psi$ is a function of the covariance parameters, which are allowed to vary in the optimization.

Maximizing the likelihood is equivalent to minimizing the cost function:

$$\mathcal{L}_\theta = \ln|\Psi| + (\mathbf{z} - \mathbf{H}\mathbf{s}_0)^T(\Psi)^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s}_0) \tag{1.18}$$

The second term will be small if the variances are large compared to the errors. But if the variances are made too large, then $p(\mathbf{z})$ will be quite spread out, so the likelihood of the observed values will not be very high even if they are near the peak.

In the geostatistical case, since no prior values are assumed either for the fluxes or for the drift coefficients $\beta$, the likelihood is computed by integrating over all possible values (*Michalak et al.*, 2004):

$$p(\mathbf{z}|\theta) \propto \int_\beta \int_\mathbf{s} \mathrm{d}\mathbf{s}\,\mathrm{d}\beta\,p(\mathbf{z}|\mathbf{s},\theta)p(\mathbf{s},\beta|\theta) \tag{1.19}$$

which, after computing the Gaussian integrals, results in the rather messy cost function

$$\mathcal{L}_\theta = \ln|\Psi| + \ln|\mathbf{X}^T\mathbf{H}^T\Psi^{-1}\mathbf{H}\mathbf{X}|$$
$$+ \mathbf{z}^T\left(\Psi^{-1} - \Psi^{-1}\mathbf{H}\mathbf{X}\left(\mathbf{X}^T\mathbf{H}^T\Psi^{-1}\mathbf{H}\mathbf{X}\right)\mathbf{X}^T\mathbf{H}^T\Psi^{-1}\right)\mathbf{z} \tag{1.20}$$

In either case, the cost function must be minimized numerically, since it depends on the covariance parameters $\theta$ in a complicated way through $\Psi$.

## 1.4.2 Optimization

Given the observations $\mathbf{z}$, the sensitivity $\mathbf{H}$ according to the transport model, and the covariance matrices $\mathbf{R}$ and $\mathbf{Q}$, it remains to find the fluxes $\hat{\mathbf{s}}$ with the maximal

posterior probability. In the Bayesian case:

$$p(\mathbf{s}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{s})p(\mathbf{s}) \propto \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{Hs})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{z} - \mathbf{Hs}) - \frac{1}{2}(\mathbf{s} - \mathbf{s}_0)^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{s}_0)\right]$$

(1.21)

Equivalently, we seek to minimize the cost function

$$\mathcal{L} = \frac{1}{2}(\mathbf{z} - \mathbf{Hs})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{z} - \mathbf{Hs}) + \frac{1}{2}(\mathbf{s} - \mathbf{s}_0)^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{s}_0).$$

(1.22)

Since the cost function for Gaussian errors is quadratic, the optimization is straightforward:

$$0 = \frac{\partial\mathcal{L}}{\partial\mathbf{s}} = -\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H\hat{s}}) + \mathbf{Q}^{-1}(\mathbf{\hat{s}} - \mathbf{s}_0)$$

(1.23)

$$(\mathbf{Q}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{\hat{s}} = \mathbf{Q}^{-1}\mathbf{s}_0 + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}$$

(1.24)

$$= (\mathbf{Q}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{s}_0 + \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{Hs}_0)$$

(1.25)

which gives the *Best Linear Unbiased Estimator (BLUE)*:

$$\mathbf{\hat{s}} = \mathbf{s}_0 + \mathbf{K}(\mathbf{z} - \mathbf{Hs}_0), \quad \mathbf{K} = (\mathbf{Q}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$$

(1.26)

The matrix $\mathbf{K}$ is called the *gain*; it controls the degree to which the observations influence the posterior estimate. When the model-data mismatch variance is large, indicating low confidence in the observations, $\mathbf{K}$ approaches zero, and the posterior estimate reverts to the prior.

Although the algebra is more complicated, the geostatistical case is similar; the only conceptual difference is that the cost function must be minimized with respect to both the fluxes and the drift coefficients $\beta$. In either case, if the dimension of the problem is not too large, or the covariance matrices are simple (e.g., diagonal), or we have a lot of computing resources, we can simply perform the necessary algebra and

17

calculate $\hat{\mathbf{s}}$. If the exact calculation is not tenable, we can resort again to numerical optimization.

## 1.5   Local Context

The work presented here includes case studies of mixing depth observations and methane flux estimates in and around Los Angeles, California. These investigations incorporate, build on, and contribute to the activities of the Megacities Carbon Project (MCP), a multi-institution collaboration that aims to use Los Angeles as a testbed for developing urban greenhouse gas monitoring and quantification methods that can be shared to other global megacities. Chapters III and IV, in particular, rely on collaboration resources: the network of in situ mole fraction observing sites described in *Verhulst et al.* (2017), the locale-specific WRF configuration of *Feng et al.* (2016) (also used for comparison in Chapter II), and the geostatistical inversion methodology on which my flux inversions are based (*Yadav et al.*, 2018). Other MCP-affiliated observations not employed here but potentially relevant to related future work include remote sensing and flask samples.

In October 2015, large quantities of natural gas began to leak from an underground storage facility at Aliso Canyon, near the Porter Ranch neighborhood northwest of downtown Los Angeles. The leak was discovered on October 23, although it may have begun earlier, and was declared sealed on February 18, 2016. Based on aircraft-based $CH_4$ mixing ratio observations taken just downwind of the facility on 13 days over the lifetime of the leak, *Conley et al.* (2016) reconstructed the approximate timeline of leak magnitude, estimating the total loss at 97,100 Mg $CH_4$. The maximum leak rate of 60 Mg/hr as estimated by *Conley et al.* (2016) is comparable to the total rate of $CH_4$ emissions from the Los Angeles area. The leak's position near a large urban area with an existing measurement network, together with the independent quantification using aircraft observations, creates a rare opportunity to test the urban observing

system against a known anomalous methane release from a point source. In Chapters III and IV, we evaluate the impacts of driving meteorology, observation density, and inversion method on our ability to detect the Aliso Canyon leak without relying on prior knowledge.

## 1.6 Overview of Dissertation

In the next three chapters, I report on several studies focusing on different elements of observation-based urban greenhouse gas flux estimation. Chapter II presents an analysis of a long (two years) time series of measurements of the mixing depth in and near Pasadena, California, using lidar instruments. A novel automated method is presented for extracting the mixing depth from the lidar observations. We show that the day-to-day variability in the mixing depth is quite large, dominating over the seasonal cycle, and that two meteorological products used to drive transport for flux estimation consistently overestimate the mixing depth. We discuss the likely resulting bias in flux estimates.

In Chapter III, we explicitly test the effects of the driving meteorology on $CH_4$ flux inversions. We show that changes in the flux, including the seasonal cycle and the effects of large emissions events, can be detected in inversions using less highly-tuned meteorology – which could open the way to near-real-time emissions monitoring. We also show that most of the differences in magnitude between flux estimates using different meteorological drivers can be attributed to overall biases in sensitivity between those drivers, and that the fluxes can be calibrated to correct for the bias. However, a trusted model must be chosen as a standard, because the sensitivity differences cannot be predicted using mean values of the relevant meteorological variables. In particular, it appears that the model mixing depth bias observed in Chapter II, though persistent, does not produce the expected magnitude of flux bias.

Chapter IV focuses on the methodology of flux estimation, and on the interac-

tions between methodology, driving meteorology, and data coverage. I find that the change detection demonstrated in Chapter III is possible with either geostatistical or Bayesian inversions, or with simpler methods incorporating no prior information at all about the pattern of emissions, so long as all of the available data is incorporated. When the observational constraint is reduced, the choice of method becomes more important: using a reduced observing network, only the geostatistical inversion can reliably detect the Aliso Canyon leak event. I also find that flux estimates supported by the full dataset and those using the most highly-tuned driving meteorology (WRF) are less sensitive to the choice of method than those using less data and/or generic meteorology.

Finally, Chapter V lays out the many other choices involved in flux estimation and briefly surveys their status in the urban context. Some general conclusions are drawn.

# CHAPTER II

# Aerosol Lidar Observations of Atmospheric Mixing in Los Angeles: Climatology and Implications for Greenhouse Gas Observations

This chapter has been reproduced from *Ware et al.* (2016) under Creative Commons license CC BY-NC-ND 4.0. The graphical design of Table 2.1 has been updated from the published version, and the references have been merged.

## 2.1   Introduction

Improved understanding of sources, sinks, and controlling processes of $CO_2$ and other greenhouse gases (GHGs) will require robust methods for estimating surface fluxes. Observations of GHG concentrations capture the influence of known and unknown sources and sinks alike, making these observations an important complement to models and inventories. Top-down GHG inversions have been used for some time to estimate fluxes on global (*Tans et al.*, 1990), continental (*Bousquet et al.*, 2000), and regional (*Lauvaux et al.*, 2013; *Peters et al.*, 2007; *Schuh et al.*, 2010) scales, and there is increasing focus on bringing a similar approach to individual cities (*McKain et al.*, 2012; *Lauvaux et al.*, 2013; *Breon et al.*, 2014; *Turnbull et al.*, 2015). However, relating observed concentrations to surface fluxes requires a representation of atmospheric

transport. On the regional and urban scales, the extent and variablity of vertical mixing is a dominant source of uncertainty (*McKain et al.*, 2012) that can easily overwhelm the effects of instrument error. It is therefore critical to represent vertical mixing accurately.

The spatiotemporal structure of vertical mixing and diffusion can be complex. However, it can be useful to approximate gases recently emitted from the surface as being confined to and uniformly distributed throughout a near-surface layer. A cluster of related concepts – atmospheric or planetary boundary layer (PBL), convective boundary layer, mixed layer – are commonly used to describe the part of the atmosphere which "responds to surface forcings with a timescale of about an hour or less." (*Stull*, 1988) Various specific definitions of these layers are in use (*Seibert et al.*, 2000), some referring to thermodynamic variables and others directly to mixing or turbulence. Layers identified by different definitions can be conceptually distinct and therefore need to be considered differently.

The layer relevant to the dilution of GHGs is that within which substantial vertical mixing takes place. The time scale of mixing under turbulent conditions has been estimated at tens of minutes (*Stull*, 1988; *van Stratum et al.*, 2012; *Pozzer and Janssen*, 2015). Nonetheless, a fully well-mixed equilibrium may not exist; we therefore follow *Seibert et al.* (2000) in referring to the *mixing layer*. When we refer to the *mixing height* or *mixing depth*, we mean the altitude of the top of the mixing layer. In addition to GHGs, the mixing depth also controls the dilution of aerosols and of other trace gases produced primarily within the mixing layer, including those that contribute to poor air quality. It is well known that shallow mixing contributes to air quality exceedances as these species are trapped near the surface, and observations such as those presented here can help in defining the presence of these conditions.

While it is difficult to measure the vertical distribution of GHGs directly, especially on an ongoing basis, we can measure the mixing height by observing the distribution

of aerosol. Lidar systems measure the backscatter of a laser from particulate matter in the atmosphere, providing a vertical (or skew) profile of the concentration of scattering particles. We make use of such an instrument, the Sigma Space Mini-Micropulse LiDAR (MiniMPL), as well as a Vaisala CL51 ceilometer. These and other remote sensing instruments benefit from continuous operation, making observations at a rate of once per minute or more. New models like the MiniMPL are smaller and more portable than earlier research lidars and have better signal-to-noise performance than ceilometers.

Note that the mixing layer may not always coincide with the boundary layer commonly diagnosed by applying thermodynamic criteria to data from radiosondes (e.g. using the parcel method). In Pasadena, comparison to results from a series of sonde launches suggests that the mixing depth is related to, though not identical with, the depth of the boundary layer as defined using thermodynamic criteria. This finding is consistent with past results. Working in Indiana and the Amazon basin, respectively, *Coulter* (1979) and *Martin et al.* (1988) found that mixing depths determined using lidar observations were similar to and well-correlated with, though generally somewhat higher than, those determined from temperature profiles. *Marsik et al.* (1995) found that mixing depths from lidar in Atlanta were slightly *lower* that those measured using sondes. We discuss the comparison to sonde data in greater detail in section 2.3.3.

Given the high frequency of observations, operational use of lidar to measure the mixing height benefits from an at least partially automated method of analysis. A variety of schemes have been used. The simplest, the gradient method (*Endlich et al.*, 1979), searches for the minimum (most negative) vertical gradient of the backscatter signal, indicating a sudden decrease in density of scatterers. Related is the inflection point method (*Menut et al.*, 1999), which searches for zeros of the second spatial derivative of the backscatter. The wavelet method (*Ehret et al.*, 1996; *Davis et al.*,

1997, 2000; *Baars et al.*, 2008), which we use, is a refinement of the gradient method that takes into account the typical spatial scale of the boundary region at the top of the mixing layer. The variance method (*Hooper and Eloranta*, 1986; *Menut et al.*, 1999) identifies the entrainment zone at the top of the mixing layer by detecting a maximum in the temporal variance of backscatter, indicating the presence of turbulent vertical mixing. The idealized-profile method (*Steyn et al.*, 1999; *Eresmaa et al.*, 2006; *Münkel et al.*, 2006) attempts to fit the vertical backscatter profile to an ideal representation of aerosol density in and above the mixing layer – typically an error function. Some studies have applied a combination of methods: for example, using the gradient to refine a spatially (*Lammert and Bösenberg*, 2006) or temporally (*Hennemuth and Lammert*, 2006) coarse estimate generated by the variance method, or using gradient methods to select a number of candidate heights, then selecting between them by minimizing disagreement with a physical model (*Di Giuseppe et al.*, 2012).

In any method, the most serious challenge in automated mixing layer detection is to distinguish between the mixing layer top and other similar boundaries in the atmosphere, such as fog, low clouds, or residual layers of scatterers remaining aloft from previous days (*Haeffelin et al.*, 2012; *Lewis et al.*, 2013). One approach to this challenge is to use the automated system only to generate a set of candidate heights and then rely on a human expert to distinguish between them. A person with some knowledge of atmospheric physics can often, though not always, identify the top of the mixing layer by visual inspection of a whole day's backscatter data. We take a different approach, aiming to automate the entire process in order to allow for long-term continuous operation. Following recent work (*Gan et al.*, 2011; *Lewis et al.*, 2013), we apply criteria that constrain the detected boundary to behavior that is physically reasonable, and we automatically detect and exclude conditions in which the instrument beam is blocked by fog or clouds. Finally, modifying a method introduced by *Lewis et al.* (2013), we implement a voting scheme, processing the day's

data in several different ways and interpreting the degree of concurrence as a measure of confidence that our algorithm has selected the correct boundary.

In section 2.2, we describe the backscatter data, the instrument used to obtain them, and our automated method for extracting the mixing height. Section 2.3 presents our findings as to the climatological mixing state in the LA area and its temporal and spatial variation. We compare the results obtained with the MiniMPL to mixing depth estimates from a ceilometer, to a one-day sonde intensive, and to PBL heights from models and reanalysis. Finally, in section 2.4, we discuss the implications of our work for GHG flux estimation and suggest possible future applications.

## 2.2 Method

### 2.2.1 Instrumentation

We collected aerosol backscatter data using a Sigma Space Mini-Micropulse Li-DAR (MiniMPL) operating at the Caltech campus in Pasadena, California. The MiniMPL is a compact version of the standard MPL, also manufactured by Sigma Space, that populates the NASA MPLNET lidar network. The MiniMPL inherits many of the design features of the MPL, such as a fiber coupled detector and robust optical train. Compared to the MPL, the MiniMPL reduces the power-aperture product to minimize cost, size, weight, and power requirements. As a result, detection range is limited to the troposphere while the MPL measures into the stratosphere. For tropospheric applications such as GHG flux estimation and air quality monitoring, however, the MiniMPL is designed to match the data quality of a standard MPL.

The MiniMPL transceiver shown in Fig. 2.1 weighs 13 kg and measures 380 x 305 x 480 mm in width, depth and height. The system consists of a laptop and the lidar transceiver, which are connected by a USB cable and consume 100W during normal operation. The whole system fits in a storm case with a telescopic handle and

Figure 2.1: A complete MiniMPL lidar system consists of an optical transceiver (shown) and a laptop running data acquisition and post-processing software.

wheels that can be checked in as regular luggage during a domestic or international flight. The system's portability allows for applications that would not be possible with the standard MPL. In section 2.3.6, we demonstrate the feasibility of operating the MiniMPL out of a moving car, enabling us to observe the spatial structure of the mixing layer without the use of aircraft.

The MiniMPL's Nd:YAG laser emits polarized 532 nm light at a 4 KHz repetition rate and 3.5 uJ nominal pulse energy. The laser beam is expanded to the size of the telescope aperture (80 mm) to satisfy the eye safe requirements in ANSI Z136.1.2000 and IEC 60825 standards. Laser light is scattered back toward the instrument by particles and molecules in the atmosphere and collected by an 80mm diameter receiver. Distance to the scattering event is calculated from the time of flight. The instrument reports the number of scattering events recorded during a user-defined accumulation time (in our case, 30 s) originating in each vertical bin. We use a vertical range resolution of 30 m. Although this study does not make use of it, the MiniMPL also measures the depolarization (*Flynn et al.*, 2007) of the scattered light with a contrast ratio greater than 100:1.

The receiver uses a pair of narrowband filters with bandwidth less than 180 pm to reject the majority of solar background noise. The filtered light is then collected by a 100 um multimode fiber and fed into a Silicon Avalanche Photodetector (Si APD) operating in photon-counting mode (Geiger mode). Photon-counting detection enables the MiniMPL design to be lightweight and compact with high signal-to-noise ratio (SNR) throughout the troposphere.

To further maximize the SNR, MiniMPL uses a coaxial design; the transmitter and receiver Field of View (FOV) overlap with each other from range zero. This design eliminates the need for a wide FOV in order to minimize the overlap distance as in some biaxial lidar systems (*Kuze et al.*, 1998). A wide FOV can result in measuring multiple scattering from aerosol (*Spinhirne*, 1982) and can distort depolarization

| | | MPL | MiniMPL |
|---|---|---|---|
| **PERFORMANCE** | Range resolution | 5/15/30/75 m (software programmable) | 5/15/30/75 m (software programmable) |
| | Minimum range | 250 m | 100 m |
| | Accumulation time | 1 sec - 15 mins | 1 sec - 15 mins |
| | Detection range* | Typically to 25 km | Typically to 15 km |
| | Polarization | standard | standard |
| | Scanning | Not available | optional |
| **OPTICS** | Laser wavelength | 532 nm | 532 nm |
| | Laser pulse energy | 6 - 8 µJ @ 2500 Hz | 3 - 4 µJ @ 2500 Hz |
| | Eye-safety | ANSI Z136.1 2000, IEC 60825 | ANSI Z136.1 2000, IEC 60825 |
| | Receiver diameter | 178 mm | 80 mm |
| | Pump laser diode | Guaranteed to 10,000 hours, user replaceable | Guaranteed to 10,000 hours |
| | Detector | Fiber coupled | Fiber coupled, user replaceable |
| **DIMENSIONS** | Size | 300 x 350 x 850 mm | 240 x 305 x 480 mm |
| | Weight (portability) | 27 kg | 13 kg |
| **DATA** | Operating system | Windows 7/10 | Windows 7/10 |
| | Computer interface | USB | USB |
| | Data transfer | LAN Ethernet | LAN Ethernet |
| **ENVIRONMENT** | Temperature | Operating +10°C to 35°C | Operating +10°C to 35°C |
| | Humidity | 0 to 80% | 0 to 80% |
| **POWER** | Supply | 110-240 VAC 50-60 Hz | 110-240 VAC 50-60 Hz |
| | Consumption | 500 W | 100 W |

\* Choosing a coarser resolution results in a longer detection range. Sample data is based on a 30s/30m setting.

Table 2.1: Technical specifications for the MiniMPL (used in this study) and the standard MPL.

measurements (*Tatarov et al.*, 2000). On the other hand, a very narrow receiver FOV could make the lidar system sensitive to external factors like shock, vibration and temperature, making the system unsuitable for field deployment. The design of MiniMPL balances the above requirements and constraints, with an FOV of 240 urad.

Additional technical specifications of the MiniMPL, along with those of the standard MPL for comparison, are given in table 2.1.

### 2.2.2 Calibration

The raw event count reported by the MiniMPL must be calibrated and normalized in order to arrive at the quantity of interest, Normalized Relative Backscatter (NRB),

which is approximately proportional to the concentration of scatterers at a given distance above the instrument. First, the event count is corrected for the deadtime of the detector, a period after each photon incidence during which no additional photons can be detected. The likely number of missed incidences can be extrapolated from the rate of detected photons. After the deadtime correction, the background (no laser light) value is subtracted. The event rate is then scaled by the laser pulse energy, which prevents changes in pulse energy from appearing as variation of the measured backscatter. Next, a correction is applied to account for laser light, called afterpulse, that strikes the inside of the instrument and returns to the detector without interacting with the atmosphere.

Finally, two corrections account for the fraction of scattered photons that are intercepted by the detector. The solid angle subtended by the collecting lens is inversely proportional to the square of the distance to the scattering event, so the event rate is multiplied by $r^2$. Since the MiniMPL laser beam overlaps with the receiver field of view from range zero, there is no need for an overlap correction in the sense required by a biaxial instrument. However, because not all of the light incident on the collecting lens is focused onto the photon counter, a geometric factor calibration is still required. For historical reasons, this factor is also referred to as an overlap correction.

These steps are summarized in the following calibration equation:

$$B_{nr} = \left[ \frac{x(z)C(x(z)) - bC(b)}{E} - \frac{x_{ap}(z)C(x_{ap}(z)) - b_{ap}C(b_{ap})}{E_{ap}} \right] \frac{z^2}{O(z)} \qquad (2.1)$$

where $x(z)$ is the raw event rate signal at distance $z$ from the instrument, $C(x)$ is the deadtime correction factor for event rate $x$, $b$ is the background, $E$ is the laser pulse energy, $x_{ap}(z)$ is the afterpulse signal at the time corresponding to distance $z$, $b_{ap}$ is the background of the afterpulse signal, $E_{ap}$ is the energy of the afterpulse, $O(z)$

Table 2.2: Number of days of MiniMPL data collection without gaps longer than one hour, by month and by concurrence score of the mixing depth estimation algorithm.

| Month | Total | 1/5 | 2/5 | 3/5 | 4/5 | 5/5 |
|---|---|---|---|---|---|---|
| January | 52 | 3 | 4 | 11 | 10 | 24 |
| February | 26 | 4 | 7 | 4 | 5 | 6 |
| March | 29 | 1 | 5 | 8 | 7 | 8 |
| April | 31 | 1 | 3 | 12 | 10 | 5 |
| May | 26 | 0 | 7 | 5 | 8 | 6 |
| June | 14 | 1 | 2 | 1 | 3 | 7 |
| July | 0 | 0 | 0 | 0 | 0 | 0 |
| August | 49 | 1 | 5 | 15 | 14 | 14 |
| September | 85 | 1 | 18 | 24 | 16 | 26 |
| October | 64 | 5 | 13 | 16 | 13 | 17 |
| November | 48 | 4 | 13 | 15 | 11 | 5 |
| December | 52 | 4 | 14 | 16 | 7 | 11 |
| All Months | 476 | 25 | 91 | 127 | 104 | 129 |

is the overlap correction factor at distance $z$, and $B_{nr}$ is the Normalized Relative Backscatter (NRB). An example vertical profile of NRB can be seen in Figure 2.3. In order to reduce the impact of short time scale fluctuations on our mixing depth retrieval, we apply a additional two-minute sliding average to the NRB values already aggregated to a thirty-second accumulation time by the instrument.

### 2.2.3 Observations

The MiniMPL collected backscatter data at Caltech on 530 days between August 1, 2012 and October 23, 2014, operating between dawn and dusk. Of those, 54 included data gaps of longer than one hour, including late starts to data collection, persistent midday rain or fog, or obstruction of the beam by obstacles. We exclude those days from the analysis. The remaining 476 days are distributed across all months other than July. Table 2.2 shows the number of days of data by month as well as the concurrence scores of the mixing depth estimates (see section 2.2.4).

We analyze backscatter data in daily increments. Over the course of the day, the changing backscatter profile gives a picture of the distribution of scatterers in

the lower atmosphere (Figure 2.2a). In the Los Angeles area, the scattering signal is typically quite strong due to the high levels of anthropogenic aerosols. Since they are produced primarily within the mixing layer, aerosols are concentrated near the surface. During the day, solar heating of the surface drives vertical mixing, causing the mixing layer to deepen and carrying aerosols to higher altitudes. As surface heating decreases in the late afternoon, the region of active vertical mixing shrinks, but the aerosols may remain aloft for some time. Frequently, aerosols carried aloft by one day's mixing can still be observed the following day in a residual layer disconnected from the surface. In the coastal mountain environment of Los Angeles, aerosols can also be carried above the mixing layer by the dominant circulation pattern, resulting in a sometimes complex stratification structure with thin, lofted aerosol layers (*Lu and Turco*, 1994, 1995).

### 2.2.4 Analysis

We use a Haar wavelet covariance method to identify boundaries between layers with high and low aerosol density. At a given height $z$, the wavelet covariance $w$ is given by integrating the product of the backscatter profile with a Haar wavelet $H$ centered at $z$:

$$w(z) = \frac{1}{d} \int dz' \, B_{nr}(z') H(z, z', d) \tag{2.2}$$

$$H(z, z', d) = \begin{cases} 1 & z - d/2 < z' < z \\ -1 & z < z' < z + d/2 \\ 0 & |z' - z| > d/2 \end{cases} \tag{2.3}$$

The width $d$, or dilation, of the wavelet is chosen to correspond to the typical size of the transition zone at the top of the mixing layer, 200 m. As illustrated in Figure 2.3, the covariance is highest where the backscatter decreases rapidly with height.
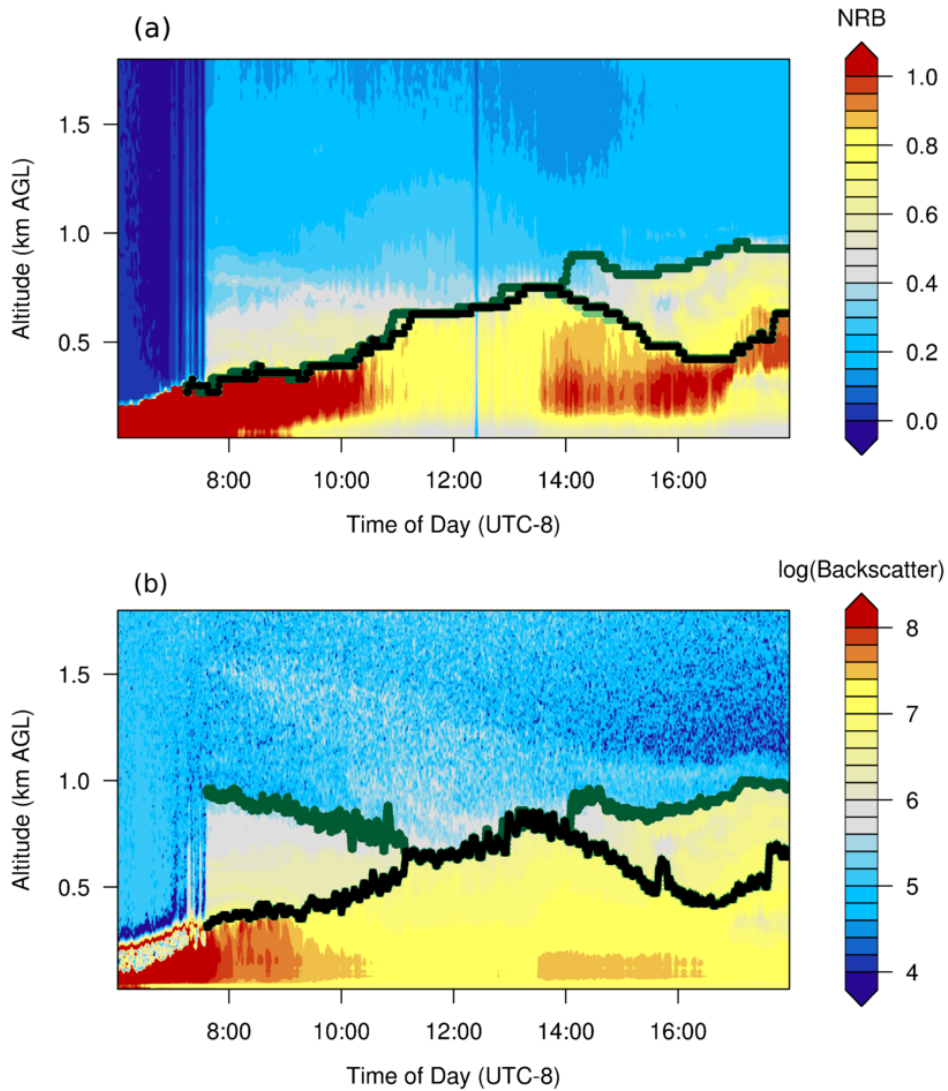
Figure 2.2: A sample day of backscatter data (heatmap) from the MiniMPL (panel a) and ceilometer (panel b, see section 2.3.2) with mixing heights as estimated by our algorithm (black symbols: majority opinion; green symbols: estimates initialized at other times of day). Prior to 8am, both instrument beams are completely extinguished near the surface; the algorithm recognizes the presence of fog and does not attempt to make an estimate. In the late afternoon – and in the morning in the case of the ceilometer – the various estimates disagree as to the mixing height, identifying two different boundaries. We report the majority opinion together with the degree of concurrence (4/5 for the MiniMPL, 2/5 for the ceilometer). Note that MiniMPL NRB values and ceilometer backscatter values do not use comparable scales.
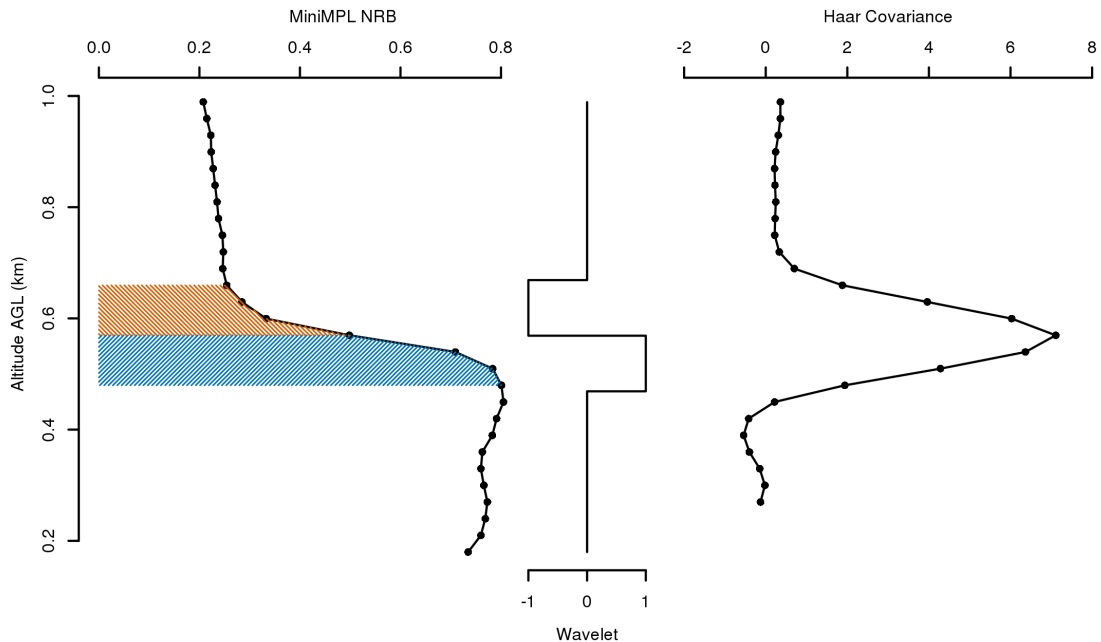
Figure 2.3: An illustration of the wavelet method. The instrument returns a vertical profile of normalized relative backscatter (NRB, left). To compute the wavelet covariance at a given altitude $z$, the backscatter profile is integrated against a Haar wavelet centered at $z$ (middle). In this example, the covariance is given by the difference in area between the orange (upper) shaded region and the blue (lower) region, which indicates the decrease in backscatter over the scale of the wavelet. The resulting Haar wavelet covariance is shown at right.

Because aerosols are concentrated within the mixing layer, such a rapid decrease in backscatter occurs at the top of the mixing layer. We therefore use high wavelet covariance values to identify the mixing layer top.

In order to increase the likelihood of detecting the mixing layer top rather than some other boundary – for example, a structure within the mixing layer or a residual layer of aerosols further aloft – each day's data is considered as a whole. Call the set of times during a single day at which backscatter data is available $t_1, \ldots, t_{max}$. First, designate a single time $t_k$ and compute the altitude $z_k(t_k)$ at which the Haar covariance is maximum. Any later timepoints are then considered in order, beginning

33

with $t_{k+1}$. The altitude $z_k$ of the detected boundary is constrained to vary at a rate no faster than $v$. For the MiniMPL, we set $v = 100 \, \text{m/min}$, a conservative upper bound on typical rates of change of the mixing layer height (*Stull*, 1988). This is equivalent to setting the Haar covariance to zero outside the range $(z_k(t_{i-1}) - v(t_i - t_{i-1}), z_k(t_{i-1}) + v(t_i - t_{i-1}))$. In addition, a multiplicative bias factor is applied to suppress the Haar covariance for unlikely but possible rates of change, decreasing linearly from one at $(2/3)v$ to zero at $v$. Similarly, any timepoints earlier than $t_k$ are considered in reverse order, beginning with $t_{k-1}$.

For $t_k = t_1$, an additional physical constraint is applied: the mixing layer top must begin each day within 500 m of the ground. This aids in selecting a boundary that is continuous with the top of the nocturnal boundary layer, as the mixing layer should be.

This process is repeated for five values of $t_k$ distributed evenly throughout the day, including the earliest time $t_1$ and the latest time $t_{max}$. The result is a set of estimates $z_{k_1}(t), \ldots, z_{k_n}(t)$ of the mixing height as a function of time.

A voting procedure is then used to select one estimate from the set. First, estimates are checked for pairwise agreement according to one of several criteria. In this study, we consider two estimates to be in agreement if they differ by no more than one unit of instrument vertical resolution (30 m for the MiniMPL as we operate it) as to the maximum depth of the mixing layer during the midday period. This criterion is optimized for determining that maximum; other criteria, such as agreement to within a tolerance over a specified fraction of the data period, might be better suited for other purposes. Next, this pairwise agreement is used to calculate a concurrence score for each estimate. An estimate $E$ has a concurrence score equal to the fraction of all estimates that agree with $E$ according to the selected criterion – see Figure 2.2 for an example. A 3/10 penalty is applied to the concurrence score of any estimate that violates the start-of-day condition, i.e. that puts the mixing layer top above

34

500 m at the start of the day. This was already forbidden during processing for the estimate beginning at $t_1$, but it may occur in other cases, and it generally indicates that the estimate has been fooled by a residual layer. After applying the penalty, the estimate with the highest score is selected for reporting, and the concurrence score can be used as a measure of confidence. Concurrence ties are broken by selecting the estimate with the earliest start time $t_k$; note that for concurrence scores of better than one-half, the tied estimates necessarily agree as to the chosen criterion. We recommend excluding estimates with scores less than one-half.

Because fog or clouds can completely extinguish the instrument beam, preventing any information from being returned from higher altitudes, it is important that our processing algorithm be able to detect this circumstance. Under foggy conditions, the altitude of highest Haar wavelet covariance does not represent the top of the mixing layer – in fact, there likely is no mixing layer – but only the maximum altitude to which the beam was able to pierce the fog before being extinguished. This situation is common in Pasadena in the early morning. We detect fog by checking directly for beam extinction, i.e. a layer of very high backscatter values with close to zero signal from above, and do not report any mixing height while fog is present.

Although the altitude of maximum Haar covariance on a foggy morning does not represent the mixing layer top, it remains useful, since that altitude transitions smoothly into the mixing layer top as the fog burns off. Fog, clouds, or rain that occur in the middle of the day are more problematic, since they often produce discontinuous changes in signal. We treat such occurrences as data gaps, and we exclude days on which gaps, including fog or rain, persist for too much of the total data period. In any case, we report for each day the maximum length of any gap in data, including instrument malfunction, a late start to data gathering, or beam extinction. It is important to check the maximum gap length before making use of the data, and to establish a standard for maximum allowable gap length, since long gaps can produce

nonsensical results.

## 2.3 Results

### 2.3.1 Climatology and Variation

On the basis of our estimates, we emphasize the very large daily variability in the mixing height in the LA basin. The maximum depth of the mixing layer in afternoon may differ by a factor of two from one day to the next. On average, the greater degree of insolation does produce deeper afternoon mixing layers in summer than in winter. Using backscatter data from the MiniMPL on days with concurrence scores of 4/5 or higher and without gaps longer than 1 hr, we find the mean afternoon maximum mixing depth to be 770 m AGL in summer (June and August) and 670 m AGL in winter (December-February).

However, as illustrated in Figure 2.4, this seasonal difference is overwhelmed by the very large day- to week-scale variability. Within-season standard deviations in afternoon maximum mixing height are about 220 m in both summer and winter, representing 29% and 32% of the means, respectively. Similarly, a given day's mixing height cannot reliably be extrapolated from measurements made on previous days. Across 105 cases across all seasons in which we achieve concurrence scores of 4/5 or higher on both of two consecutive days, the root-mean-square difference in afternoon maximum mixing depth at Caltech is 230 m.

The high variability reinforces that applications of climatological mixing depth values are subject to large uncertainties; sustained observations like those we present here can quantify those uncertainties. Such observations can also be used to calibrate models or to choose between parameterization schemes in meteorological models, as we discuss in sections 2.3.4 and 2.3.5. Comprehensive comparisons to a model and/or to other meteorological observations over a long period could also provide a more

36
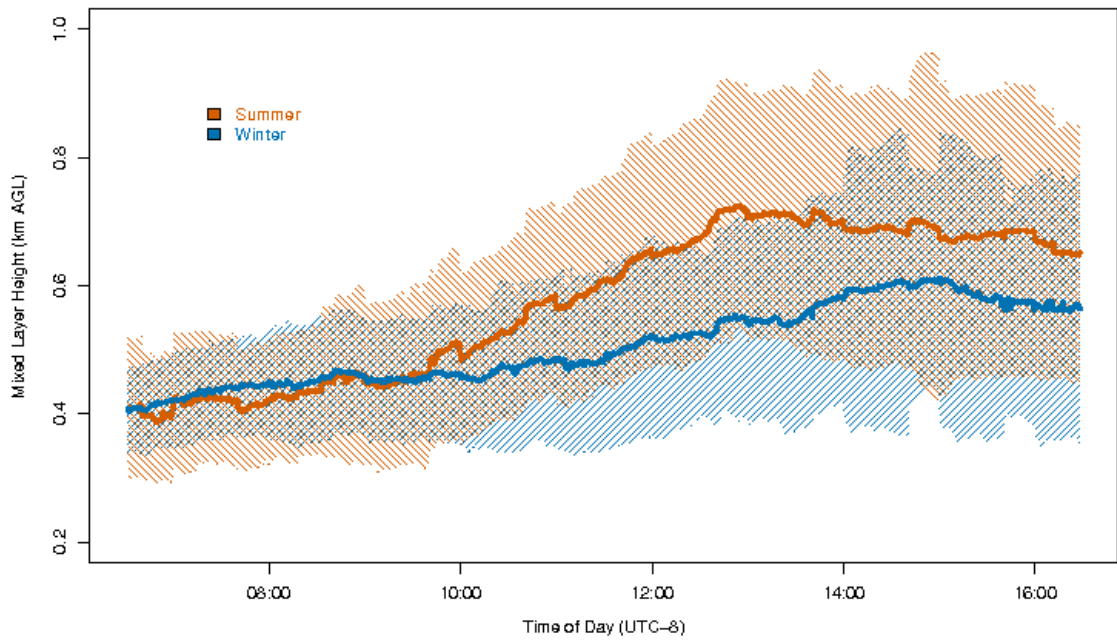
Figure 2.4: Solid curves: average diurnal cycles of mixing height in June-August (orange) and December-February (blue). Shaded regions: one standard deviation of between-days variability. Estimates according to the MiniMPL, retaining only days on which the concurrence score was at least 4/5.

granular understanding of the mixing dynamics. A robust explanation is needed for the variation we observe, which takes place too consistently and on too short a time scale to be attributed solely to unusual events such as forest fires or the LA basin's periodic Santa Ana winds.

### 2.3.2 Ceilometer

Alongside the MiniMPL, we also operated a Vaisala CL51 ceilometer (a successor instrument to the CL31, for details see e.g. *Munkel and Rasanen* (2004); *McKendry et al.* (2009); *Münkel et al.* (2011)) at the same site. The measurement principle of the ceilometer is similar to that of the lidar, but the overlap correction and other calibration steps are performed by proprietary software not visible to or modifiable by the user (*Wiegner et al.*, 2014). The resulting quantity is referred to simply as the backscatter profile. The CL51 operates at 910 nm, in the near-infrared; it uses a 16 s temporal bin and a 10 m vertical range resolution.

We apply a version of the same algorithm to estimate mixing depths based on ceilometer backscatter data as we use with the MiniMPL . An example is shown in Figure 2.2b for comparison to the MiniMPL results on the same day. As is visible in the figure, especially in regions of low backscatter signal, the ceilometer's signal-to-noise performance is not as good as that of the MiniMPL. As a result, some adjustments are necessary. First, the maximum allowed rate of change $v$ in the mixing layer height must be relaxed; for the ceilometer we set it to 150 m/min. This change is necessary because noise can temporarily disguise a change in the boundary location; the algorithm must be able to "snap back" to the true location of the boundary even after it has moved some distance away.

Second, the ceilometer tends to show an unrealistically large signal in the near field. *He et al.* (2006) note a similar artifact, which they attribute to an imperfectly corrected overlap error (see also *Wiegner et al.*, 2014). Such errors are caused by

differences in the optical geometry of the outgoing beam aperture and the detector that collects scattered photons. Because the erroneous backscatter signal associated with the artifact decays very rapidly, it has a high wavelet covariance. The algorithm therefore tends to detect the artifact in place of the real boundary, estimating the mixing depth at the lowest possible altitude.

To solve this problem, and to dampen noise in general, it is standard to take the logarithm of the ceilometer backscatter data prior to processing. Although physically unmotivated, this preprocessing step flattens out large signals, decreasing the influence of the low-altitude artifact. Figure 2.5 shows an example of a case in which the log transform allows the algorithm to detect the correct boundary. However, the log transform also introduces an offset. It suppresses the magnitude of the gradient of the backscatter more where backscatter values are higher:

$$\frac{\mathrm{d}}{\mathrm{d}z} \log(b(z)) = \frac{1}{b(z)} \frac{\mathrm{d}b(z)}{\mathrm{d}z} \tag{2.4}$$

Backscatter decreases with height in the transition from the mixing layer to the free troposphere above, so the strongest gradient in $\log(b(z))$ generally occurs at a higher altitude than the strongest gradient in $b(z)$. This effect carries over to the wavelet covariance method, causing a positive offset of about 50 m. The offset is due to a methodological choice to identify the altitude of greatest relative change in scattering, not a difference in physical reality. It should therefore be noted and compensated for in comparisons with estimates that identify the altitude of greatest absolute change, i.e. those that do not employ a log transform.

The effect of the log transform on the whole dataset is shown in Figure 2.5. There are two distinct populations. On some days, the low-altitude artifact traps the maximum mixing depth at the bottom of the instrument range. Applying the log transform removes the effect of the artifact, allowing the true mixing depth (which is
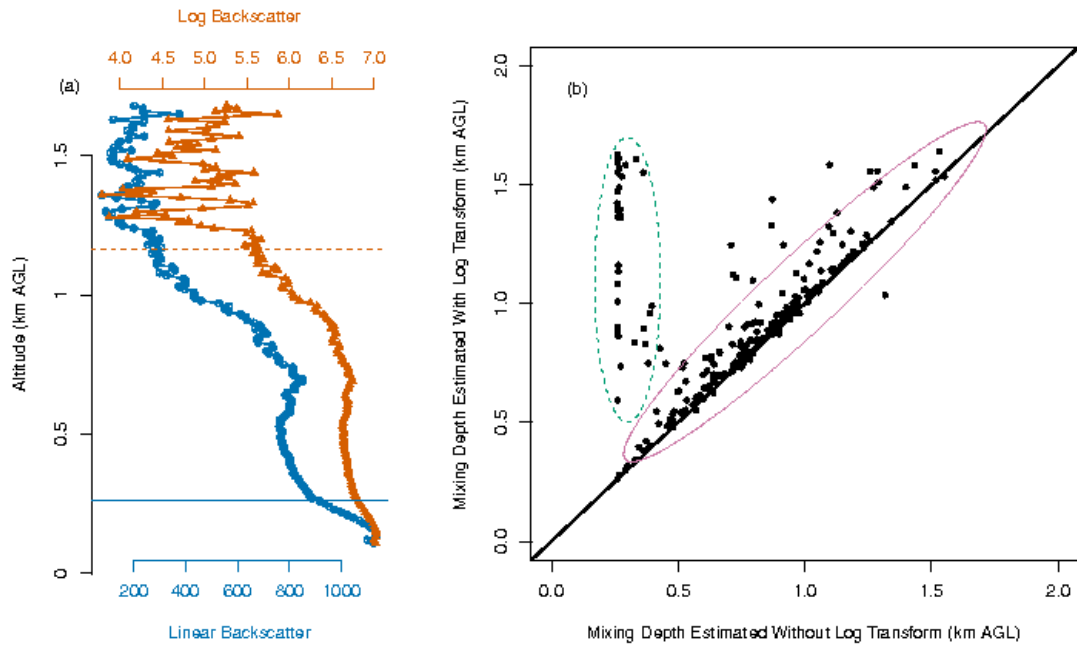
Figure 2.5: (a) Representative backscatter profiles from the ceilometer, with (orange, triangular symbols) and without (blue, round symbols) applying a log transform. The horizontal lines show the corresponding mixing height as estimated by the algorithms: solid blue line, without transform; dashed orange line, with transform. Note the very high backscatter values at low altitudes in the untransformed data, which fool the algorithm into selecting an unrealistically low mixing height. (b) Maximum afternoon mixing depths as estimated using ceilometer data with (vertical axis) or without (horizontal axis) applying the log transform. Days on which the untransformed data is affected by the low-altitude artifact are indicated by the dashed green ellipse; taking the log transform removes the effect of the artifact. On other days (indicated by the solid pink ellipse), the bias introduced by the transform is visible. The solid black line is the 1-1 line. Only days with concurrence scores of at least 3/5 are shown.

variable) to be detected. On days on which the algorithm is not fooled by the artifact, the offset introduced by the log transform is visible: applying the transform results in an average increase of about 50 m in the estimated mixing depth.

Even with adjustments, our confidence in mixing height estimates derived from the ceilometer is not as high as in those derived from the MiniMPL. One proxy for confidence in a given day's results is the degree of concurrence among estimates in the voting procedure (see section 2.2). As can be seen in Figure 2.6, the MiniMPL achieves unanimity (concurrence score of 5/5) or near-unanimity (score of 4/5) on 51% days for which data is available. By contrast, the ceilometer achieves a score of 4/5 or better on only 36% of days. It is for this reason that we focus our results on estimates derived from MiniMPL observations.

### 2.3.3   Sonde Comparison

In September 2012, a one-day intensive campaign of sonde launches was conducted for comparison to mixing layer information from the MiniMPL. Sondes were launched every three hours between 7:00am and 7:00pm local time. The results are displayed in Figure 2.7. In each case, the PBL height is extracted from the sonde using the method of *Heffter* (1980). At 7:00am, morning fog is still present and the mixing layer has not yet developed. At 10:00am, 1:00pm, and 4:00pm, the mixing height identified using the backscatter data coincides with the sonde-derived PBL height to within 150 m. Since the top of the mixing layer is in fact a transitional zone of 100 to 200 m thickness, it should not be considered to have a well-defined exact location. Some discrepancy should therefore be expected even between methods that detect substantially the same layer. In this one-day comparison, the backscatter method displays no identifiable systematic bias with respect to the sonde method; of course, the comparison presented is too limited to conclude that no bias exists.

We emphasize that, while the sonde comparison provides some confidence that,
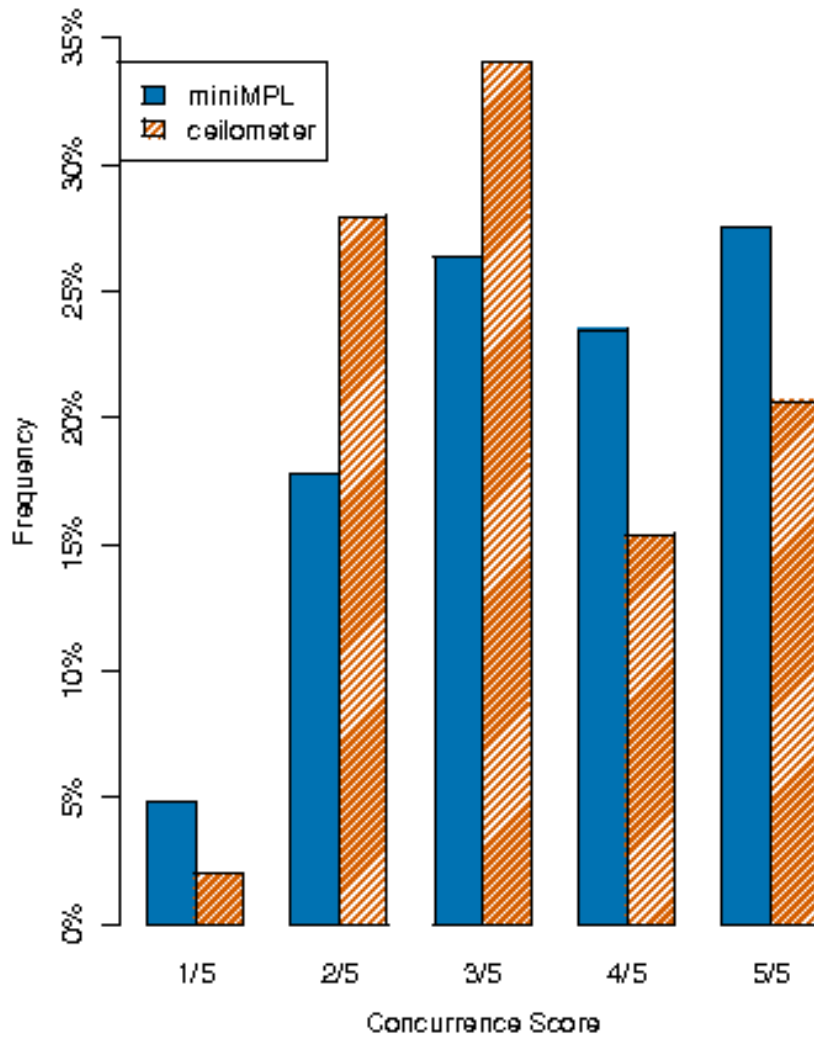
Figure 2.6: Degree of concurrence achieved by the algorithm using backscatter data from the MiniMPL (blue, solid) or from the ceilometer (orange, dashed), shown as a fraction of days on which both instruments were operating.
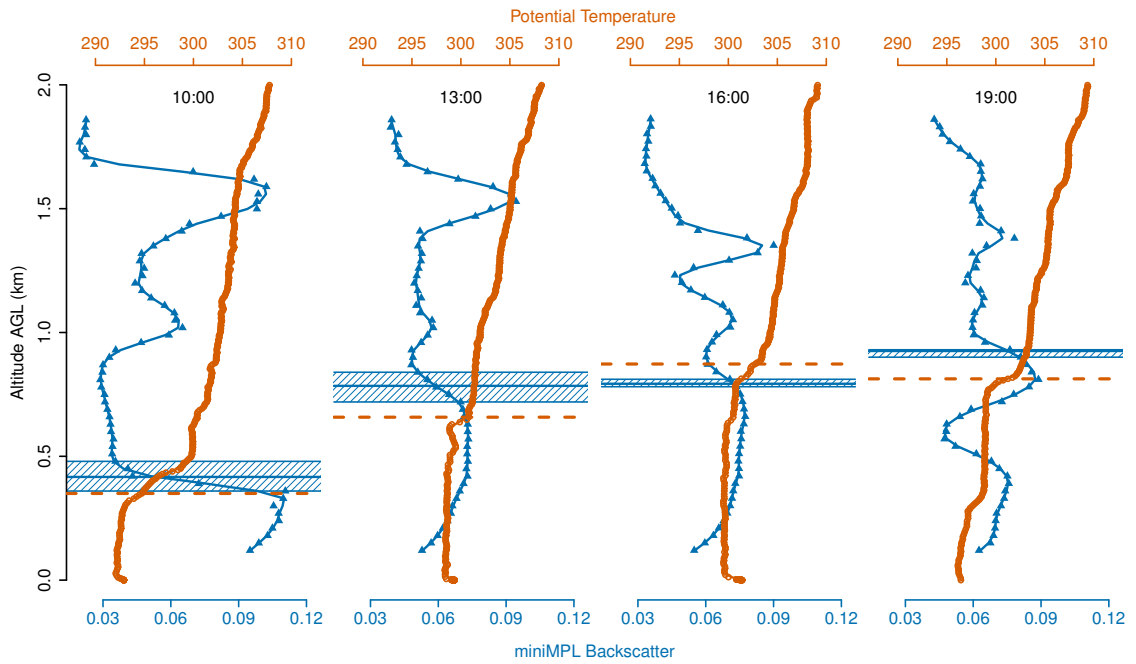
Figure 2.7: Orange: potential temperature profiles from sonde launches, with the corresponding PBL height as calculated using Heffter's method (horizontal dashed line). Blue with triangles: contemporaneous MiniMPL backscatter profiles, with the mean (center horizontal line) and range (shaded area) of the algorithmically-estimated mixing height over the 30-minute period surrounding the sonde launch.

at least during the day, the layer in which elevated aerosol levels are present does correspond to the thermodynamic boundary layer, it is in any case the former that is of most interest for interpreting atmospheric concentrations of trace gases. For the purpose of linking atmospheric measurements to emissions rates, the important question is what part of the atmosphere should be considered in contact with the surface. In other words, through what volume are species emitted from the surface dispersed?

By 7:00pm, the mixing layer has begun to collapse and the structure is becoming more complicated. Two distinct boundaries are visible in both the potential temperature profile and the backscatter distribution, and both methods select the higher of these. Indeed, the day's aerosol emissions are distributed up to the higher boundary at 920 m. However, with the decrease in solar heating to drive vertical motion, the upper part of the identified layer (above about 500 m) is probably no longer interacting with the surface. Our method has therefore failed to detect a region of substantial, active vertical mixing. This case serves as a reminder that the mixing layer concept is not always straightforwardly applicable, particularly in the evening as vertical mixing tapers off. Care should be taken in interpreting and applying our or any other mixing depth estimates around sundown, even on days – like this one – with otherwise robust retrievals.

### 2.3.4 North American Regional Reanalysis (NARR) Comparison

GHG flux inversion studies typically make use of PBL heights derived from meteorological models or reanalysis products. We compare afternoon maximum mixing depth estimates based on MiniMPL data to PBL height estimates from the Weather Research and Forecasting model (WRF) and the North American Regional Reanalysis (NARR). NARR is a reanalysis product providing a variety of atmospheric and surface variables over North America at 32 km spatial resolution and at 3 hour intervals

Table 2.3: Mean, median, 1st and 3rd quartiles, and standard deviation of afternoon maximum PBL height (NARR) or mixing depth (MiniMPL), in km AGL, over 227 days with concurrence score at least 4/5 and without data gaps longer than on hour.

| Method | $Q_1$ | Median | Mean | $Q_3$ | $\sigma$ | $\sigma$/Mean |
|--------|-------|--------|------|-------|----------|---------------|
| MiniMPL | 0.63 | 0.75 | 0.84 | 0.98 | 0.27 | 32% |
| NARR | 1.46 | 1.84 | 1.92 | 2.20 | 0.62 | 32% |

(*Mesinger et al.*, 2006). We find a large and persistent difference between afternoon maxima of MiniMPL-derived mixing depths at Caltech and PBL height estimates at the nearest NARR grid location. Figure 2.8 shows the distributions of these quantities over 227 days on which the MiniMPL estimate achieves a concurrence score of at least 4/5 and without data gaps longer than one hour. The maximum NARR PBL height exceeds the maximum MiniMPL-derived mixing depth on all but one day, differing by a factor of two or more on 63% of days. Summary statistics are in Table 2.3.

Interestingly, although maximum NARR PBL heights are an average of 2.5 times MiniMPL derived mixing depths, the two quantities are similarly distributed. Both show substantial variability, with standard deviations about 32% of the respective means, and both are skewed toward high values, with skewness 0.84 (NARR) and 0.88 (MiniMPL). However, NARR does not reproduce the detailed timing of this variability. Even after scaling maximum NARR PBL heights down by a factor of 2.5 to account for the mean difference, a root-mean-square difference of 360 m remains between scaled NARR estimates and MiniMPL estimates on the same days. This is almost as large as the RMS difference of 370 m in a sample of $10^6$ random pairs of MiniMPL estimates and scaled NARR estimates.

We can attribute NARR's failure to accurately represent the boundary layer in Pasadena at least in part to its coarse spatial grid. The meteorology of the Los Angeles basin is strongly influenced by the coastal mountain topography (see Figure 2.9), resulting in a complex pattern of circulation (*Lu and Turco*, 1994, 1995). It comes as no surprise that a product unable to resolve the rapid changes in elevation
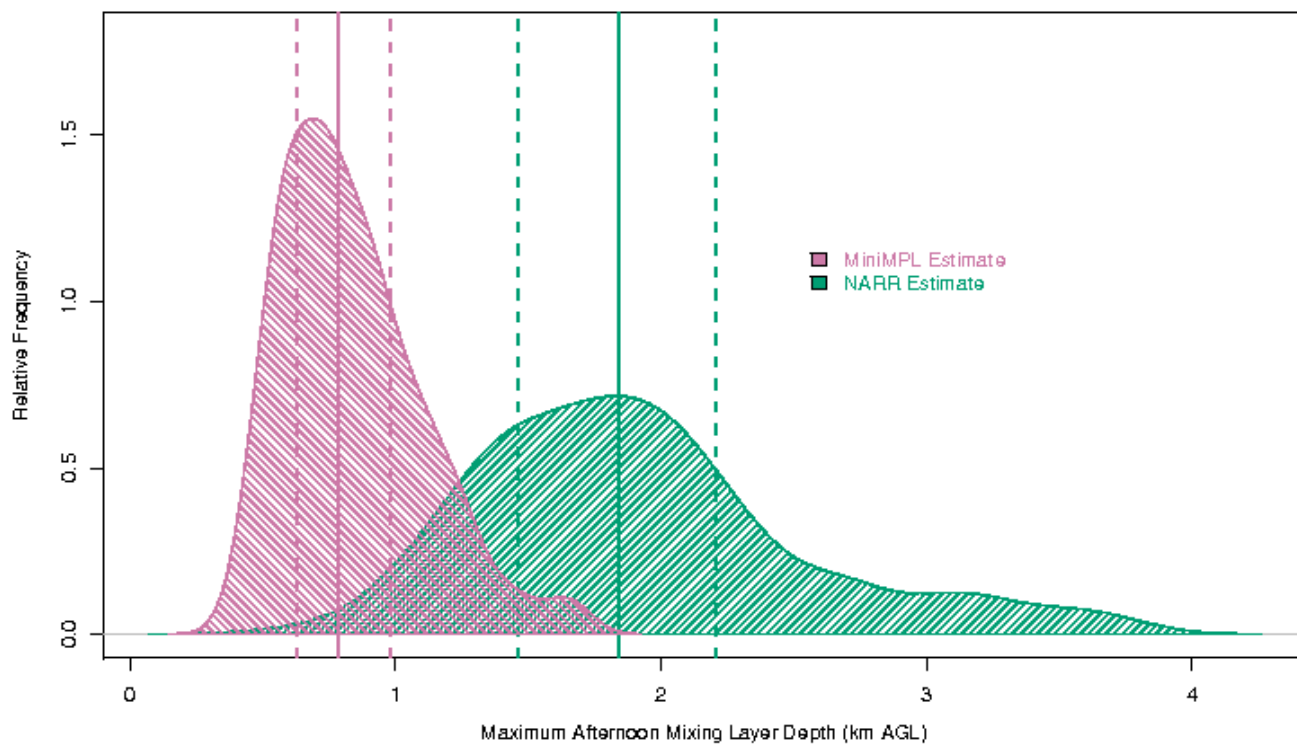
Figure 2.8: Gaussian kernel density (smoothed relative frequency) of maximum afternoon mixing depth according to the MiniMPL (pink, left peak) and according to NARR (green, right peak) over 227 days with MiniMPL concurrence score at least 4/5. Solid vertical lines: median; dashed vertical lines: quartiles.
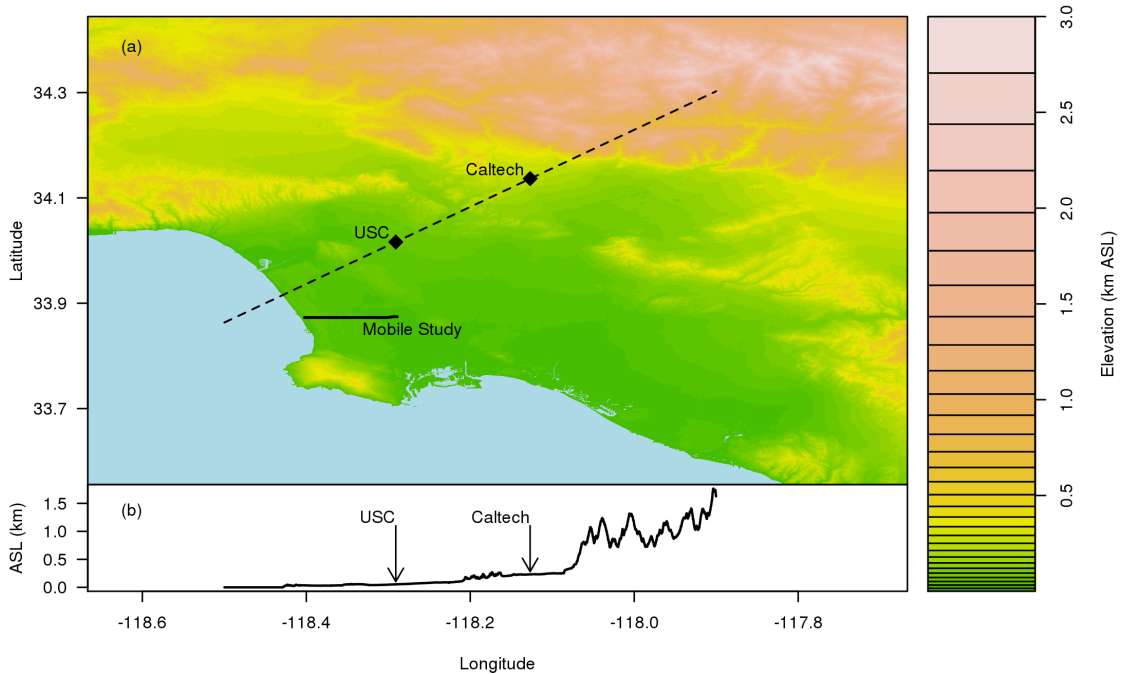
Figure 2.9: (a) Elevation map of the Los Angeles Basin (*Witt*, 2015). The labeled diamonds indicates the location of the measurement site at Caltech (in Pasadena). The solid line shows the route taken in the mobile study; the dashed line corresponds to the cross section in panel (b). (b) Elevation cross section along the dashed line in panel (a); the longitude scale is the same for both panels.

will struggle to predict PBL heights in this environment. If NARR is used to drive a transport model for GHG flux estimation in Los Angeles or in other areas with meteorology strongly influenced by the detailed topography, careful evaluation and correction of mixing depth biases will be critical for avoiding large errors. Since a biased mixing depth results in a proportional bias in flux estimates (see the general argument in section 2.4), we would expect a 250% bias in an LA flux inversion using NARR.

### 2.3.5 Weather Research and Forecasting (WRF) Comparison

Given the difficulty posed by the rapidly-varying topography of the LA basin, one might expect a high-resolution model to better represent the mixing dynamics. We

47

compare mixing depth estimates from MiniMPL data taken during a deployment of the instrument in October-November 2015 to PBL heights from such a high-resolution model, a WRF setup developed specifically for the Los Angeles environment by *Feng et al.* (2016) to simulate $CO_2$ concentrations. The model is initialized with NARR and with sea surface temperatures from NCEP and uses three nested domains, with the innermost domain covering the LA basin at a resolution of 1.3 km. Using observations from the intensive Calnex campaign in 2010, including aircraft and ceilometer PBL measurements, *Feng et al.* (2016) tested a variety of WRF configurations. We employ only the MYNN_UCM_d03 configuration, which they found to minimize errors.

We redeployed the MiniMPL to Caltech for the three-week period of October 21 to November 9, 2015. Of these twenty days of observations, the mixing depth estimation algorithm achieves a concurrence score of 4/5 or better on six days and a score of 3/5 on another nine days. Although this comparison period is too short to allow robust statistical conclusions, we make some preliminary observations. Like NARR, WRF PBL heights show variability that is similar in relative terms to that of Mini-MPL derived mixing heights. Over the three-week comparison period, the standard deviation of maximum afternoon WRF PBL heights is 540 m, about 37% of the mean. However, WRF estimates PBL heights that are greater than MiniMPL-derived mixing depths on all but one afternoon. On average, afternoon maximum WRF PBL height exceeds afternoon maximum MiniMPL mixing depth by 730 m. Considering only days with high concurrence scores reduces the discrepancy considerably. The mean difference on days with scores of 4/5 or better is 380 m, suggesting that the concurrence voting scheme effectively identifies days that are easier to analyze.

The discrepancy we find between modeled PBL height and MiniMPL-derived mixing depth is surprising given the excellent agreement reported by *Feng et al.* (2016). During the 2010 Calnex campaign period, they report a mean WRF-derived daytime PBL height (using the same MYNN_UCM_d03 configuration we use here) of 828.8

m, in good agreement with a mean mixing depth of 835.7 m obtained from ceilometer measurements using the gradient method. They also report substantially less variability in modeled PBL height than in measured mixing depth. Further work, including a model-data comparison covering a longer period, is clearly needed to resolve this perplexing difference. While such a comparison is beyond the scope of this study, we do note that NARR PBL height estimates for May-June 2010 are generally similar to those from our comparison period in October-November 2015, with a mean daily maximum of 2.1 km.

Our analysis here cannot distinguish between differences due to errors in mixing depth estimation, errors in modeled PBL depth, or conditions under which the mixing layer fails to correspond to the thermodynamic PBL. Still, it is prudent to expect that the same complex stratification which can cause the mixing depth estimation algorithm to fail might also indicate challenging conditions for the model. By selecting days with high concurrence scores, MiniMPL observations can be used to choose "golden days" for model analysis. Alternately, if a model is run over a long period, days with good agreement between the model and lidar estimates can be selected for flux estimation. For example, Figure 2.10 shows a pair of days which would not be readily distinguished on the basis of model results alone. The additional information provided by the lidar estimates lets us assign greater confidence to modeling on the day with good agreement (panel a) than that with poor agreement (panel b).

Sustained lidar can also inform the choice of model configurations or parameters, as *Feng et al.* (2016) and others (e.g., *Nehrkorn et al.*, 2013a) have done with PBL observations from limited campaigns. In addition to increasing confidence in that choice simply by virtue of a larger volume of data, long-term observations can provide more detailed information about how model errors depend on season or on other meteorological conditions. For example, *Lewis et al.* (2013) found that PBL height as estimated by the general circulation model GEOS-5 differs most from that measured
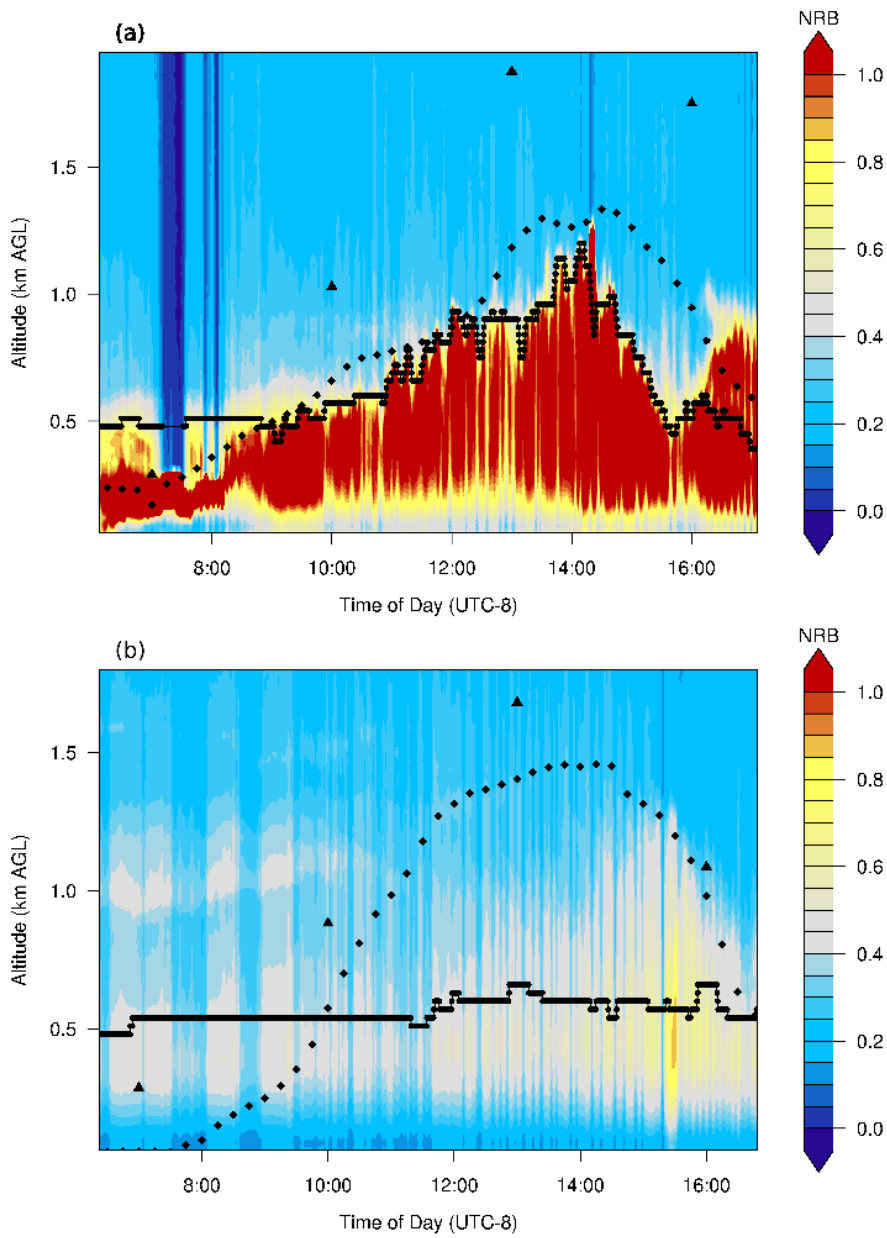
Figure 2.10: Examples of days with good (panel a) and poor (panel b) agreement between MiniMPL-derived mixing depths (small circles) and PBL heights as estimated by WRF (large diamonds). NARR PBL heights (large triangles) show large discrepancies in both cases.

by the lidar network MPLNET in winter. Unlike sondes, lidar data can validate not only the depth of the mixing layer but also the timing of its development and collapse. That timing can be critical; for example, in an urban setting, a difference of one hour may determine whether the mixing layer begins to develop before, during, or after the emissions peak associated with the morning rush hour.

An alternate method for integrating mixing depth observations into flux estimation is to characterize a known model bias and correct for it after the modeling stage. *Zhao et al.* (2009) use three months of wind profiler measurements to derive an empirical relationship between observed and modeled PBL heights. They apply that relationship to scale down modeled PBL height before computing fluxes, reducing the residual error by a factor of 1.5. Among the advantages of postprocessing corrections of this kind are that they are simple to apply, allowing accuracy to be improved even in less detailed inversions, and that they can be combined with the strategies discussed above to further control any errors remaining after tuning model parameterization and/or selecting out "golden days."

### 2.3.6 Spatial Variation

Taking advantage of the MiniMPL's portability, we also conducted a one-time pilot mobile study in which backscatter data was collected over a period of about twenty minutes as the MiniMPL was transported due west toward the Pacific coast in the back of a passenger car. This observing strategy, which could not have been implemented with a full-size research lidar, is made possible by the compact size and low power requirements of the MiniMPL. Now that we have demonstrated its feasibility, we hope that this new approach will allow for both more regular mapping of the spatial structure of the mixing layer and more nimble mobile deployment of lidar in response to irregular events like fires and gas leaks.

The spatial profile of aerosol backscatter near the Pacific coast is shown in Figure
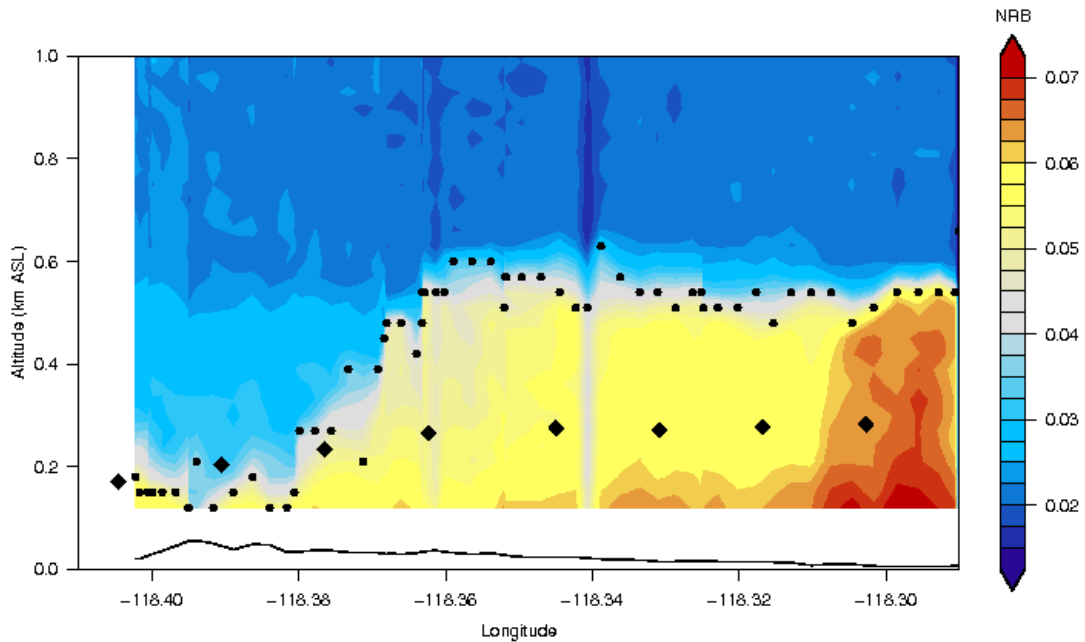
Figure 2.11: Heatmap: MiniMPL backscatter intensity near the Pacific coast (located at longitude -118.41). Small black circles: mixing depth as estimated by the gradient method using MiniMPL data. Large black diamonds: PBL height as estimated by WRF. Black curve at bottom: topography (same vertical scale).

2.11. The transition between the shallow marine layer, which extends two to three kilometers onto land, and the convective regime that dominates further inland is clearly visible. The vertical structure in this case is simple, with a well-defined mixing layer of high backscatter adjacent to the ground and a sharp decrease in backscatter at the top of that layer. The mixing depth as estimated by the minimum backscatter gradient is indicated in the figure by the black circles (our retrieval algorithm is not suitable, since it relies on the temporal evolution of the boundary at a fixed location).

Figure 2.11 also shows the PBL height as predicted by WRF. The WRF prediction agrees well with the MiniMPL-derived mixing depth near the coast, but does not increase as sharply further inland. Unlike at Caltech, in this case the WRF PBL height is lower than the observed mixing depth. Repeated measurements of this kind

could reveal whether the difference is consistent with time and at locations elsewhere along the coast, both in the immediate Los Angeles area and elsewhere, which could contribute to model development in the challenging coastal environment. Further work characterizing the coastal transition could also aid in understanding the fate of GHG emissions from sources like ports and marine industry.

## 2.4   Conclusions

Researchers have recognized that the representation of mixing dynamics is both critical for the interpretation of top-down emissions estimates and also a major source of uncertainty (e.g., *Newman et al.*, 2013; *Zimnoch et al.*, 2010). *McKain et al.* (2012) advocate the use of column-integrated concentration measurements in urban studies, among other reasons in order to avoid the impact of mixing height errors. A common strategy (*Breon et al.*, 2014) is to rely only on observations made during midafternoon, when the mixing layer is at or near its maximum depth and the detailed timing of its dynamics are less important. But we observe even afternoon maximum mixing height in Los Angeles to vary substantially from day to day, typically differing from the seasonal mean by 30%.

A simple dimensional argument demonstrates the impact of such variations. Suppose that an instrument measures the in situ concentration of a trace gas at a particular location. This concentration is expressed as a molar fraction, or, equivalently given the local density of dry air, as a volume concentration $C$ in moles of gas per unit volume, i.e. in $n/L^3$. The goal is to use the measurement to infer a surface flux $F$, expressed in moles of gas emitted or absorbed per unit area per unit time, i.e. as $n/(L^2T)$. On dimensional grounds, any method for relating the concentration to the flux must incorporate some temporal information, such as the time $\tau$ during which the sampled air mass was exposed to the flux, and also some vertical length scale.

The relevant vertical length scale is the mixing height $h$, which controls the height

of the space into which the emitted gas is diluted. We therefore expect

$$F \propto hC/\tau \tag{2.5}$$

in which case an error in the mixing height $h$ will result in a proportional error in the flux estimate. In detailed models, this picture is complicated to some degree by higher-order effects, e.g. the coupling between vertical motion and horizontal wind shear, but the essential proportionality remains. Applying sustained observations to control mixing depth errors, whether by validating models, choosing suitable periods for analysis, or characterizing and correcting for errors in postprocessing, is critical for accurate GHG flux estimation.

We have focused above on determining the depth of the mixing layer, especially at its afternoon maximum. But the mixing layer concept is not always applicable. Even when the mixing height is applicable, it does not fully describe the complex structure of the lower troposphere. The potential exists to extract much more information about that structure from lidar backscatter data. Among other applications, a more complete picture of the mixing state could contribute to our understanding of the transport of species emitted from the surface. Here we suggest one direction in particular for future work.

The mixing layer itself may exhibit internal structure. For example, in Los Angeles, the sea breeze circulation pushes near-surface air inland during the day. As a result, the air mass within the mixing layer over Pasadena in the afternoon has traveled over downtown in the preceding hours. The time scale of this horizontal motion, and the varying emissions rates and compositions from the traversed areas, may create a stratification, in which fresh emissions from Pasadena are concentrated in the lowest part of the mixing layer while those from downtown are more thoroughly mixed throughout. If we were able to observe and understand within-layer dynamics

of this kind, we could much more precisely link trace gases observed in the atmosphere to their points of emission, allowing us to answer more specific questions about the sources and composition of emissions in the urban environment.

Since the lidar is primarily sensitive to aerosols and not to trace gases, the distribution of aerosol would need to be used as a proxy for the distribution of co-emitted trace gases, assuming that the two are transported within the mixing layer in a similar way, at least on short time scales and over small distances. That assumption would need to be tested before it could form the basis of any future work. Challenges notwithstanding, this is an exciting possibility for future applications, including more detailed validation of transport models and finer-scale attribution of emissions sources within complex urban environments like that of Los Angeles.

## 2.5   Acknowledgments

# CHAPTER III

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

This chapter is awaiting submission to *Journal of Geophysical Research: Atmospheres*. It was coauthored with Eric A. Kort, Riley Duren, Kristal Verhulst, and Vineet Yadav.

## 3.1    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 spatial 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 secular 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, observing 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 in particular. 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. 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.*, 2013b; *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.

Rather than focus on validating transport models or driving meteorology or quantifying errors therein, we ask: *with respect to what purposes or questions, if any, and to what degree, does density in an observing network reduce the importance of the choice of driving meteorology?* Following the method of *Yadav et al.* (2018), we perform geostatistical inversions of methane emissions using in situ mole fraction observations in the South Coast Air Basin (SoCAB) in and around Los Angeles, California. We compare inversions driven by the locally-validated WRF modeling system developed by *Feng et al.* (2016) for the SoCAB to those driven by three broadly-available models or reanalysis products: HRRR, NARR, and GDAS. The leak of large quantities of natural gas from the Aliso Canyon storage facility beginning in October 2015 serves

as a known-release experiment (*Conley et al.*, 2016), allowing us to test the inversion systems' capacity to detect changes or unusual events.

## 3.2  Methods

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), which estimates the sensitivity of in situ $CH_4$ mole fraction measurements to emissions fluxes by simulating the transport of 800 particles 60 hours back in time from each observation. We estimate fluxes at a spatial resolution of 0.03 degrees within the SoCAB – a total of 1826 grid cells – 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 3.1 along with an example posterior $CH_4$ flux map (from the inversion using WRF will the full observing network).

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. Those comparisons were used to select the PBL and urban surface schemes with the best performance in the LA environment. The inner WRF domain, which includes the region considered here, has a spatial resolution of 1.3 km.

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 with data assimilation, 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.

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, but cover larger areas (North America and the whole globe). Given the complex topography and sea breeze circulation (*Lu and Turco*, 1994, 1995), we would not expect these coarse products to accurately represent conditions on fine scales within our estimation domain, which spans only about 200 km from east to west. However, HRRR, NARR, and GDAS are all run in a routine operational mode, and output can be downloaded from the NOAA READY archive in a format immediately suitable for transport modeling. For low cost, low latency monitoring in diverse urban environments, these products are available off-the-shelf.

Our inversions process data from the surface monitoring network maintained by the LA Megacities Carbon Project (*Verhulst et al.*, 2017), which measures $CH_4$ mole fractions at nine locations within our domain. We use only observations taken between 12:00 and 16:00 local time, since the representation of vertical mixing is believed to be most reliable in the afternoon when the mixing layer is most fully developed. Smoothed observations from an additional site on San Clemente Island, in the Pacific ocean off the coast of Los Angeles, are used to estimate background concentrations (see *Verhulst et al.*, 2017). Any excess above that background is attributed to local sources within the SoCAB.

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
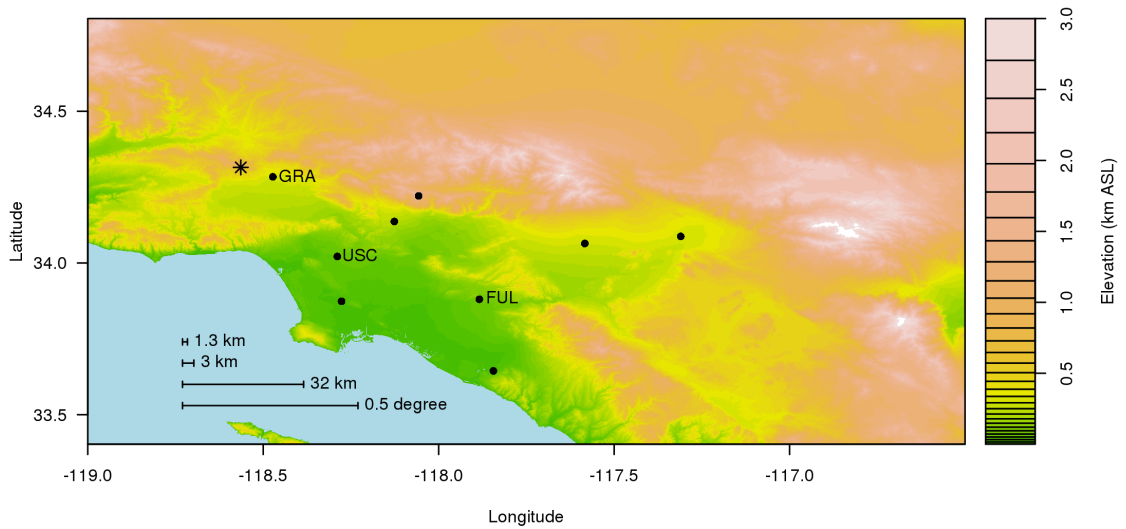
Figure 3.1: Colors: Elevation map of the study domain. Circles: locations of observing sites. The three sites included in the reduced network are indicated by their three-letter codes. The star in the western part of the domain indicates the location of the Aliso Canyon facility. Scale bars indicate 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.

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 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. A complete description of the observing network is available in *Verhulst et al.* (2017).

In all inversions, we employ a geostatistical inversion methodology similar to that used by *Yadav et al.* (2018). For each four-day period, the inversion solves for emissions fluxes $\mathbf{s}$ in each of the 1826 grid cells as well as for the magnitudes $\beta$ of model components (columns of $\mathbf{X}$). The cost function to be minimized is:

$$L(\mathbf{s}, \beta) = \frac{1}{2}(\mathbf{z} - \mathbf{Hs})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{z} - \mathbf{Hs}) + \frac{1}{2}(\mathbf{s} - \mathbf{X}\beta)^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\beta)$$

where $\mathbf{z}$ are the observed enhancements (mole fraction minus background), $\mathbf{H}$ is the sensitivity matrix derived from the transport model, $\mathbf{R}$ is the model-data mismatch covariance matrix, and $\mathbf{Q}$ is the flux-space covariance matrix.

We use two model components in $\mathbf{X}$: a spatially flat model of mean emissions, and a model proportional to the distribution of emissions in the CALGEM inventory (*Zhao et al.*, 2009; *Jeong et al.*, 2012). Note that CALGEM is not treated as a prior: the total magnitude of fluxes from CALGEM is not an input to the inversion and, in particular, solutions are not penalized for departing from that total. In addition, 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.

Both the model-data mismatch $\mathbf{R}$ and the flux covariance $\mathbf{Q}$ are taken to be diagonal, i.e. both errors and deviations from the model components are taken to be uncorrelated. In $\mathbf{R}$, each measurement location is assigned a separate variance $\sigma_i^2$. $\mathbf{Q}$ is the sum of two parts, one proportional to each of the model components,

with proportionality factors $\theta_1$ and $\theta_2$. All of the covariance parameters $\{\sigma_i^2, \theta_1, \theta_2\}$ are estimated on the basis of the mole fraction data using the Restricted Maximum Likelihood (RML) method (*Michalak et al.*, 2005).

Also following *Yadav et al.* (2018), we estimate fluxes for two four-day periods at a time, then discard the first of the two periods, treating it as a spin-up window. Covariance parameters, model component magnitudes, and fluxes are therefore all recomputed for each new time period. Since we expect no significant methane sinks within the domain, we constrain the flux in each grid cell to be nonnegative. We use the L-BFGS-G optimization method (*Byrd et al.*, 1995) to minimize the cost function subject to that constraint.

The imposition of nonnegativity breaks the Gaussianity of the posterior emissions probability, which prevents us from calculating posterior uncertainties analytically. In test inversions run without the nonnegativity constraint, the calculated total-basin flux uncertainty at the 95% level was generally between 18% and 35%. However, that uncertainty estimate does not include bias, which is substantial in some cases (see section 3.3.1). In addition, because it is calculated over an eight-day window only, the posterior uncertainty does not fully account for variations in the actual or modeled sensitivity of the observations to surface fluxes. We therefore rely on the spread of flux estimates across a number of consecutive four-day periods, rather than the calculated uncertainty for any given period, as a more realistic estimate of variance when testing for flux changes (see section 3.3.2).

## 3.3   Results

### 3.3.1   Basin Total Flux

Estimated whole-basin methane fluxes from each of the four inversions are shown in Figure 3.2. When the full observing network is included, estimates using trans-

port driven by WRF and NARR are in broad agreement with the 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) (see shaded region, Figure 3.2). Emissions 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 constant difference in overall mean total sensitivity assigned by each model to the measurement network. We compute the mean total sensitivity $H_{mean}$ 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 3.2. Treating WRF as a transfer standard, we perform an empirical calibration, scaling the posterior fluxes from the NARR, HRRR, and GDAS-driven inversions by the ratios of the sensitivities computed using those models relative to those using WRF:

$$\mathbf{s}_{cal} = \frac{H_{mean}}{H_{mean,WRF}} \times \mathbf{s} \qquad (3.1)$$

After calibration, the mean posterior emissions come into much closer alignment. The difference in mean flux 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 3.3.

If the sensitivity bias could be corrected using direct observations, our results suggest that accurate flux estimates might be possible 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

64
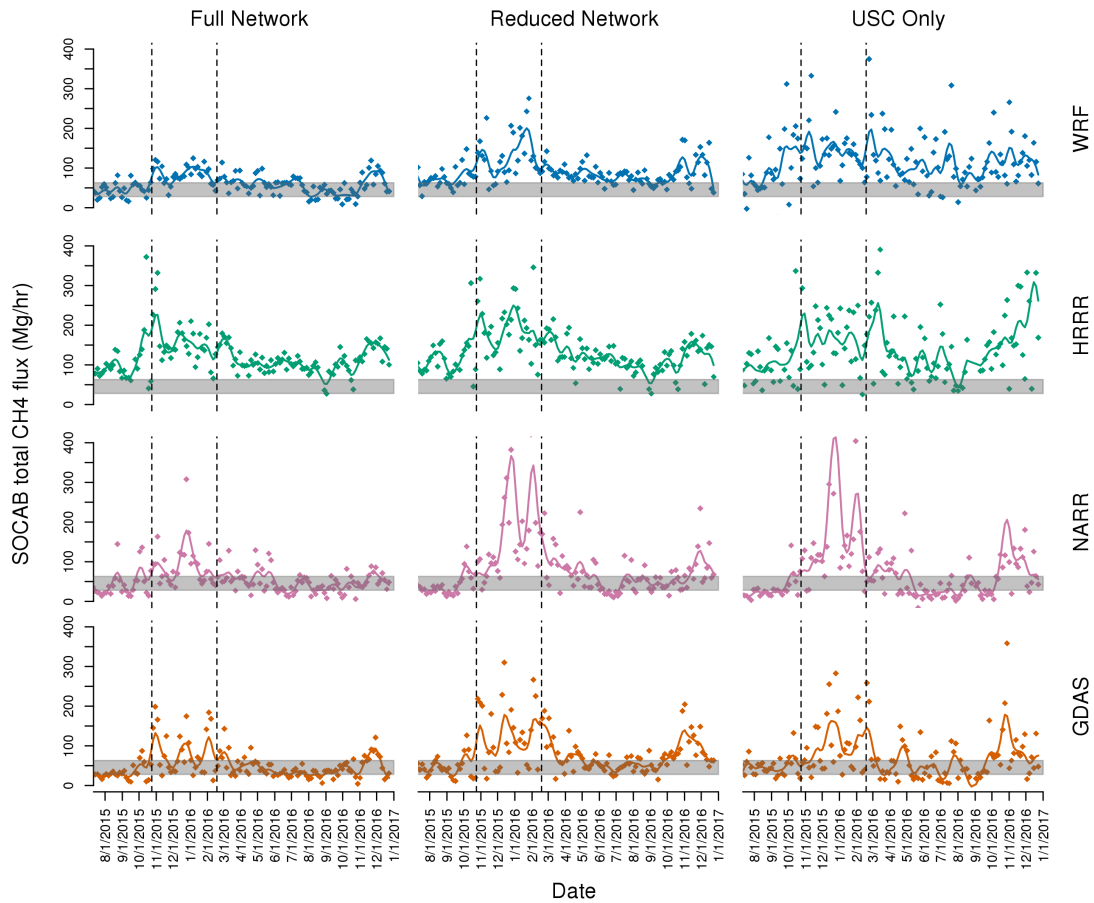
Figure 3.2: Points: estimated total CH$_4$ flux time series for the South Coast Air Basin (SoCAB), at four-day time intervals, according to inversions using transport driven by each of four meteorological models and using the full observing network (9 sites), a reduced network (3 sites), or a single observing site. Curves: spline fits to 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.
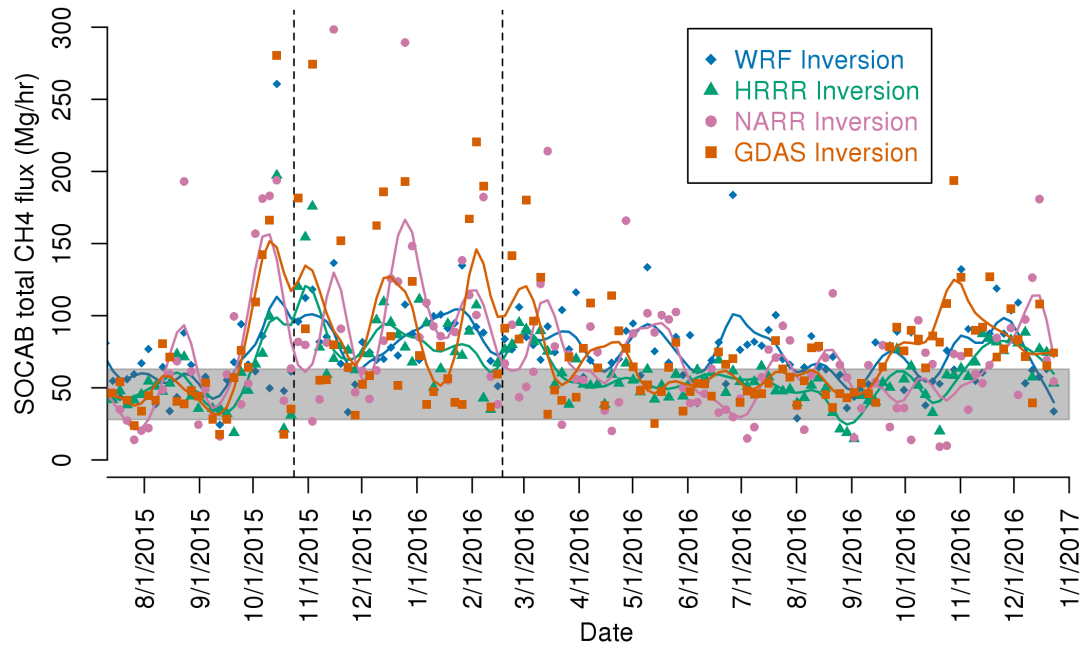
Figure 3.3: Estimated SoCAB total CH$_4$ flux time series in inversions using the full observing network after calibration by scaling the fluxes by the relative total sensitivity assigned to the observing network by each driver of the transport model. The calibration brings the estimates into close agreement overall. 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.

sensitivity to surface fluxes by tracking the amount of time simulated air parcels spend in contact with the surface. The sensitivity of the $i$th observation to the $j$th flux region is given by (*Lin et al.*, 2003)

$$ H_{ij} = \frac{m_{air}}{\rho_j} \frac{\tau}{z_j}; \qquad \tau = \frac{1}{N_i} \sum_{p_i=1}^{N_i} \Delta t_{p_i,j} \qquad (3.2) $$

where $z_j$ is the mixing depth, accounting for the effect of dilution, and $\tau$ is the average time spent by the parcels within the bottom one-half of the mixing layer above the flux region. We might therefore 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 3.1, 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 insensitive to the surface either because they have exited the domain horizontally or because they are above the bottom half of the mixing layer, we also computed the fraction of their residence time that measured parcels spent near the surface. As shown in Table 3.1, 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

|                                | 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%        |
| Total Average Sensitivity      | 100  | 53 / 53%    | 96 / 96%     | 119 / 120%   |

Table 3.1: 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 3.3.1, and percentages are relative to the same variables in WRF. Fourth row: for HRRR, NARR, and GDAS, expected sensitivity relative to that in WRF given the above variables. Fifth row: actual total mean sensitivity. The actual relative sensitivities are not accurately predicted on the basis of the mean meteorological variables.

fourth line of Table 3.1, in which we compute the relative sensitivity predicted by those mean variables according to

$$\frac{\mathrm{H_{mean}}}{\mathrm{H_{mean,WRF}}} = \frac{z_{WRF}}{z} \times \frac{\tau}{\tau_{\mathrm{WRF}}} \times \frac{f}{f_{\mathrm{WRF}}} \qquad \text{(predicted)} \qquad (3.3)$$

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 3.1.

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 3.2 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), we showed in Chapter II 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 3.1 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.

### 3.3.2 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 3.2.

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 our test of whole-basin flux totals is less sensitive than that performed by *Yadav et al.* (2018), which focused only on fluxes in the area around the Aliso Canyon facility. The tradeoff is that the whole-basin test does not require advance knowledge of the location of a flux anomaly.

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 3.2, this intermittency results 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 complementary value of the two approaches, particularly for less-optimized meteorology.

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

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 3.2: 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.
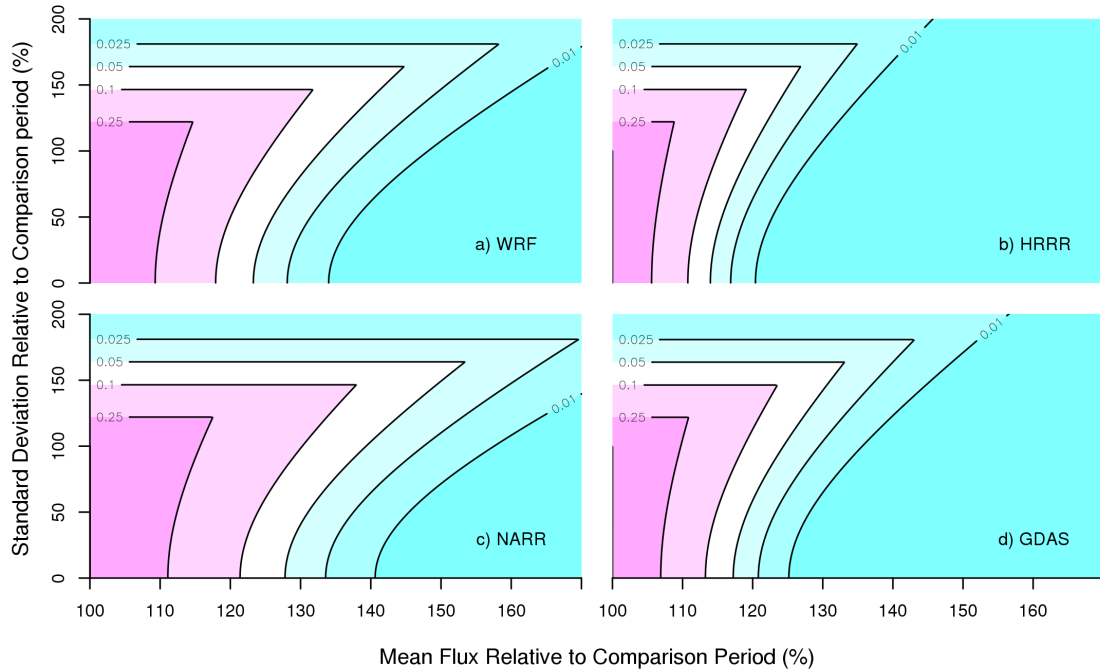
Figure 3.4: Sensitivity (p-values) of inversions using each meteorological driver to hypothetical flux events occurring between September 4 and October 26, 2017, as a function of the change in mean flux and variance relative to the same period in 2016. The inversions shown here use the full observing 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.

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 flux increases and variances. The results are shown in figure 3.4 for the inversions driven by each of the four meteorological products. In this example, the inversions using WRF and NARR could detect a 30-40% increase above the baseline if the variance were approximately unchanged, the inversion using GDAS could detect a 20-30% increase, and the inversion using HRRR could detect an increase of about 20% – but the same thresholds do not persist at other times.

### 3.3.3 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 3.5 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.

## 3.4 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
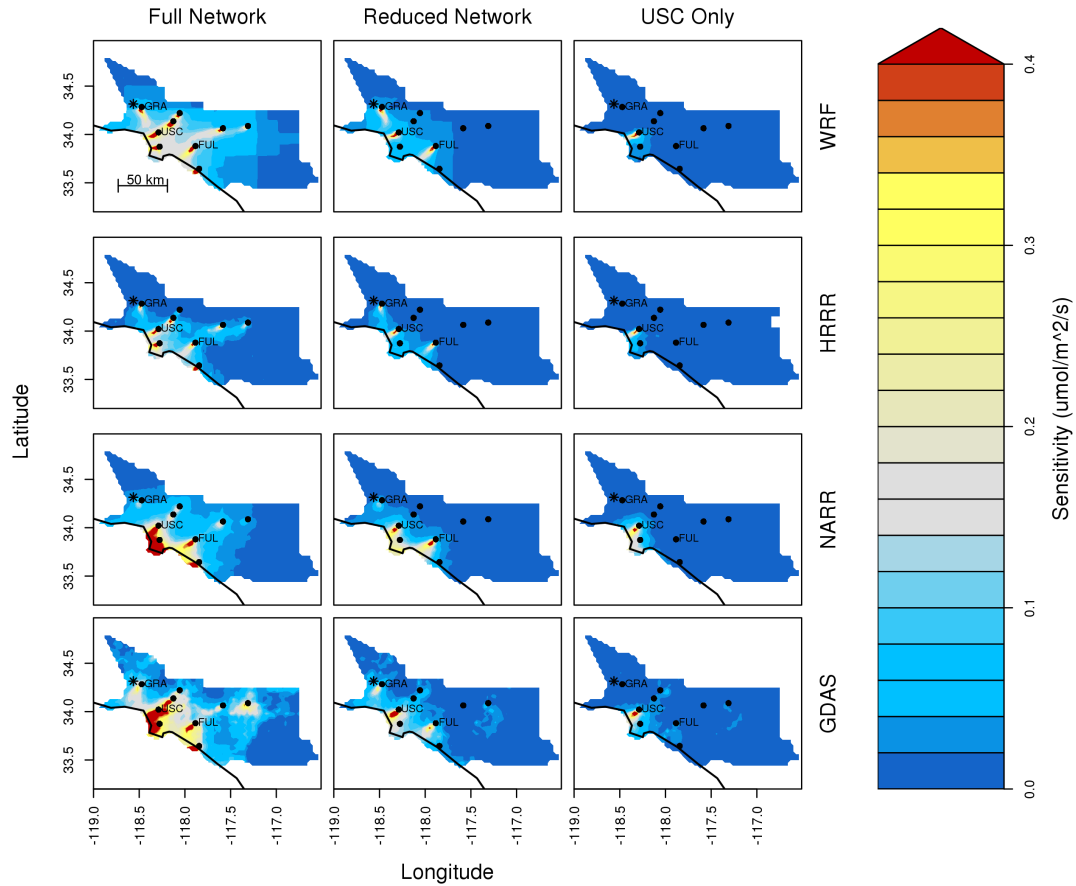
Figure 3.5: Heat map: sensitivity of the full observing network (9 sites), a reduced network (3 sites), and the USC site alone to fluxes within the SoCAB during the first four days of the Aliso Canyon natural gas leak, October 24-27, 2015, as computed by STILT driven by each of four meteorological products. 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.

chosen as a transfer standard. Debiasing with weather observations would not be successful as the sensitivity bias is not predicted by the mean values of the relevant 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. And although our study period is not long enough to observe them, trends over the course of years could likely be characterized in the same way.

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. (One exception was the near-real-time monitoring performed by *Lauvaux et al.* (2013) in Davos, Switzerland in 2011-2012.) While several factors contribute to delays, one major source of latency is the time, expense, and computational resources involved in meteorological modeling for transport. We have demonstrated that at least some operational monitoring goals 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. With these rapidly-available meteorology products, and given the modest computation expense involved in the geostatistical inversion itself, emissions fluxes can be estimated as quickly as concentration data can be collected and quality-controlled.

## 3.5 Acknowledgments

# CHAPTER IV

# Impact of Methodology Choice in an Urban Methane Flux Inversion

## 4.1 Introduction

As inverse methods have become a standard tool for estimating greenhouse gas emissions fluxes, practitioners have attempted to characterize the major sources of error and to reduce the role of expert judgment. Early global inversions (e.g. *Tans et al.*, 1990; *Enting et al.*, 1995) estimated fluxes in a few large regions, but the use of large regions introduced aggregation error (*Kaminski et al.*, 2001). Since then, grid-scale inversions have become the norm, especially on the urban scale (e.g. *Breon et al.*, 2014; *Lauvaux et al.*, 2016; *Yadav et al.*, 2018), often with many more unknown fluxes than observations are available to constrain them. The flux estimates are then sensitive to the choice of information used to regularize the problem.

One response to this sensitivity has been to improve the quality of the inputs. For $CO_2$, flux priors based on disaggregation of national total emissions using proxy variables (e.g. nightlights, population density) have been produced at 1.1 km scale globally (*Oda and Maksyutov*, 2011), with refinements to as low as 30.3 m under study (*Oda et al.*, 2017). Truly bottom-up products, constructed from process-level information about sources, cover a few cities (*Gurney et al.*, 2012; *Feng et al.*, 2016;

*Patarasuk et al.*, 2016) at building scale and the contiguous United States at 100 km or finer (*Gurney et al.*, 2009). Methane inventories are not yet as detailed, but are a subject of active research (e.g. *Maasakkers et al.*, 2016), especially since it has become clear that atmospheric measurements imply higher total emissions in North America than have been accounted for (*Brandt et al.*, 2014).

An alternative to producing the best possible inputs is to reduce the method's dependence on inputs, relying on the mole fraction measurements to the greatest possible extent. That philosophy has guided the adaptation and development of the geostatistical method in the context of flux estimation (*Michalak et al.*, 2004). Although one or more plausible spatial flux patterns must still be provided, the observations are permitted to guide the relative importance of those patterns as well as the overall flux magnitude. Advantages of the geostatistical method include reduced sensitivity to prior error, especially bias, and that the magnitude of an inventory or other prior estimate can be reserved as an independent point of comparison. The main disadvantage to be expected is increased sensitivity to model-data mismatch error, including transport error.

So far, the geostatistical and traditional Bayesian inversion methods have not been compared side by side in the urban context. Here, we make that comparison, estimating $CH_4$ emissions fluxes in the South Coast Air Basin, California using both methods. We also make estimates with two simpler regularization schemes, Tikhonov and lasso regularization, which involve no prior information about either the magnitude or the pattern of fluxes. We evaluate the general features of the estimates, their time series behavior and spatial flux patterns, and the ability to detect flux changes and events.

## 4.2 Methods

We estimate methane emissions fluxes in and around Los Angeles, California using several methods (enumerated by *Turner et al.* (2018)): a Bayesian inversion using a prior, a geostatistical inversion, and simple regressions either with Tikhonov regularization or with lasso (L1 norm) regularization. All of the analyses are informed by hourly average in situ $CH_4$ mole fraction measurements taken at observing sites associated with the LA Megacities Carbon Project. We perform inversions using both the full observing network (nine sites within the domain) as well as with a reduced network of three sites. A smoothed time series of mole fractions at the additional San Clemente Island site, off the Pacific coast just west of Los Angeles, is subtracted as an estimate of background. Both the observing network and the background method are described in *Verhulst et al.* (2017). We restrict the observations to the afternoon hours (16:00 to 20:00 local time), when the mixing layer is generally well-developed and model representations of vertical mixing are considered to be most reliable.

The sensitivity of the mole fraction observations to surface fluxes is computed with the Stochastic Time-Inverted Lagrangian Transport model (STILT) (*Lin et al.*, 2003; *Nehrkorn et al.*, 2010). We drive STILT with meteorological fields from the Weather Research and Forecasting model (WRF) as configured for the Los Angeles area by *Feng et al.* (2016) and with three other models or reanalysis products: the NOAA High Resolution Rapid Refresh model (HRRR) (*Benjamin et al.*, 2016), the North American Regional Reanalysis (NARR) (*Mesinger et al.*, 2006), and the Global Data Assimilation System (GDAS). The characteristics of these meteorological fields are discussed in greater detail in Chapter III.

The estimation domain is the South Coast Air Basin (SoCAB), encompassing Orange County and parts of Los Angeles, Riverside, and San Bernardino counties. As in Chapter III, we estimate fluxes at a spatial resolution of 0.03 degrees and at a four-day time scale. Each inversion solves two consecutive four-day periods at a time,

then discards the first, treating it as a spin-up period. The fluxes are constrained to be nonnegative, since we expect no significant surface methane sinks within the domain.

All four of our methods seek to minimize a cost function $\mathcal{L}$, the first term of which depends on the deviation of the observed mole fractions $\mathbf{z}$ from those expected on the basis of the estimated fluxes $\mathbf{s}$:

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2; \qquad \mathcal{L}_1 = (\mathbf{z} - \mathbf{Hs})^T \mathbf{R}^{-1}(\mathbf{z} - \mathbf{Hs}) \tag{4.1}$$

where $\mathbf{R}$ is the model-data mismatch covariance matrix. We specify a diagonal $\mathbf{R}$, assuming uncorrelated errors in the observations, but allow a different variance $\sigma_i^2$ at each observation site to reflect the potentially differing magnitudes of (especially transport) error.

Minimizing $\mathcal{L}_1$ alone is not sufficient to determine the fluxes uniquely because the number of fluxes (1826 per four-day period) exceeds the number of observations (no more than 124 per four-day period). As a result, the Jacobian matrix $\mathbf{H}$ of the transport must have a nontrivial null space, so $\mathcal{L}_1$ is insensitive to some variations in the fluxes. Some additional cost $\mathcal{L}_2$ is necessary to regularize the minimization.

A simple and widely-used method for regularizing ill-posed optimization problems is to set $\mathcal{L}_2 = \gamma \|\mathbf{s}\|$, where $\gamma$ is a positive constant and $\| \cdot \|$ is a norm on the vector to be optimized. For greenhouse gas flux inversion, this choice has the disadvantage that it will cause any fluxes unconstrained by the observations to be set to zero – but it can give us an idea of what information is contained purely in the observations, since no other inputs are involved. We use two methods of this form. If $\| \cdot \|$ is the Euclidean (L2) norm, the result is Tikhonov regularization:

$$\mathcal{L}_2 = \gamma \mathbf{s}^T \mathbf{s} \qquad \text{(Tikhonov)}. \tag{4.2}$$

The Tikhonov cost term will tend to promote a broad distribution in space, since the same total flux has a lower cost if spread over several grid cells. By contrast, if $\| \cdot \|$ is the L1 norm, we have lasso regularization (*Tibshirani*, 1996),

$$\mathcal{L}_2 = \gamma \sum_i s_i \qquad \text{(Lasso)}, \tag{4.3}$$

which can promote a sparse solution with few nonzero flux values. Although unrealistic for many emissions cases, we might expect sparseness to make the lasso regression more sensitive to point sources, which could be desirable for event detection.

In the Bayesian approach, we suppose that we have a prior estimate $\mathbf{s}_0$ of the fluxes, which we believe to be reasonable. The role of the observations is to identify deviations from the prior estimate. The regularizing term is set by the prior probability of the fluxes:

$$\mathcal{L}_2 = -2 \ln p(\mathbf{s}) = (\mathbf{s} - \mathbf{s}_0)^T \mathbf{Q}^{-1} (\mathbf{s} - \mathbf{s}_0) \qquad \text{(Bayesian)} \tag{4.4}$$

where $\mathbf{Q}$ is the prior flux covariance. Here, we set $\mathbf{s}_0$ equal to the gridded methane flux inventory from the California Greenhouse Gas Emissions Measurement (CALGEM) Project (*Zhao et al.*, 2009; *Jeong et al.*, 2012). We specify a diagonal prior covariance as a linear combination of an uncertainty proportional to the prior estimate in each grid cell, plus an additional constant variance, i.e.

$$\mathbf{Q} = \theta_1 \operatorname{diag}(\mathbf{s}_0^2) + \theta_2 \mathbf{I}_m. \tag{4.5}$$

Note that a Bayesian interpretation of Tikhonov regularization is possible, if the prior probability of the fluxes is

$$p(\mathbf{s}) = \exp \left\{ -\frac{1}{2} (\mathbf{s} - \mathbf{0})^T (\gamma \mathbf{I}_m)(\mathbf{s} - \mathbf{0}) \right\} \tag{4.6}$$

but this is obviously not a very sensible prior in our case, since it implies zero emissions

everywhere.

The geostatistical method (*Michalak et al.*, 2004) differs from the Bayesian in that the prior is replaced by a model $\mathbf{X}$ of the emissions, the components of which are to be scaled by one or more initially unknown parameters $\beta$ (called "drift coefficients"). The cost function is specified as

$$\mathcal{L}(\mathbf{s}, \beta) = \mathcal{L}_1 + \mathcal{L}_2; \qquad \mathcal{L}_2 = (\mathbf{s} - \mathbf{X}\beta)^T \mathbf{Q}^{-1} (\mathbf{s} - \mathbf{X}\beta) \qquad \text{(Geostatistical)} \qquad (4.7)$$

and the drift coefficients are optimized along with the fluxes. A crucial conceptual difference from the Bayesian approach is that the model contains no information about the overall scale of the fluxes – only their distribution in space and, if the inversion covers multiple time periods, in time. The overall flux magnitude is left to be determined by the observations. Here, we implement a near-minimal geostatistical model with only two components: the same CALGEM flux inventory as in the Bayesian case, and a uniform flux throughout the SoCAB domain. We specify the prior covariance $\mathbf{Q}$ with the same form as in the Bayesian case.

In the Bayesian and geostatistical inversions, we estimate the observation variances $\sigma_i^2$ for each measurement site, along with the prior covariance parameters $\theta_1$ and $\theta_2$, using the Restricted Maximum Likelihood (RML) method (*Michalak et al.*, 2004, 2005). The parameter values are chosen so as to maximize the likelihood of the actually existing observations, and are computed separately for the Bayesian and geostatistical cases and for each inversion period. We also apply RML to estimate the observation variances and the regularization parameter $\gamma$ in Tikhonov regularization; the relevant RML cost function is derived in Appendix A. For the regression using lasso regularization, the RML calculation is prohibitive because the probability distribution implied by the cost function is not Gaussian. In that case we set the observation variances to $(40 \text{ ppb})^2$, a typical value from the Tikhonov case, and the

regularization parameter to 1 ppb$^{-1}$, which gives the lasso cost term approximately the same magnitude as the Tikhonov term.

## 4.3    Results

Figure 4.1 shows the estimated flux time series using each of the four methods (Bayesian, geostatistical, Tikhonov, and lasso) and with transport driven by each of the four meteorological products (WRF, HRRR, NARR, and GDAS), including mole fraction data from all nine observing sites. The characteristic features of each method are visible.    The geostatistical inversion is the only method without an absolute flux magnitude in the regularizing cost term, which allows it to show the greatest variation in the flux magnitude both across time and depending on the meteorological driver. In the other three methods, the flux is more tightly constrained, trending towards the prior in the Bayesian inversion and towards zero in the Tikhonov and lasso regressions. With our choice of parameters, lasso regularization allows for larger deviations from zero than does Tikhonov regularization, especially during the period of greatly increased emissions from the Aliso Canyon natural gas leak.

The differences between methods become more pronounced as the observational constraint is reduced. The estimated fluxes using the reduced set of three observing sites are shown in Figure 4.2. The fluxes estimated by the geostatistical method become more variable as the importance of any given observation and its sensitivity increase. In the Bayesian inversion, less information is available to pull the estimates away from the prior, so the time series becomes less feature-rich, deviating substantially only at times of greatly increased flux: during the Aliso Canyon leak and, in some inversions, in late 2016. For the same reason, the regularized estimates collapse to near zero much of the time.

As discussed in Chapter III, most of the difference in absolute flux magnitude between the geostatistical inversions using different meteorological drivers of transport
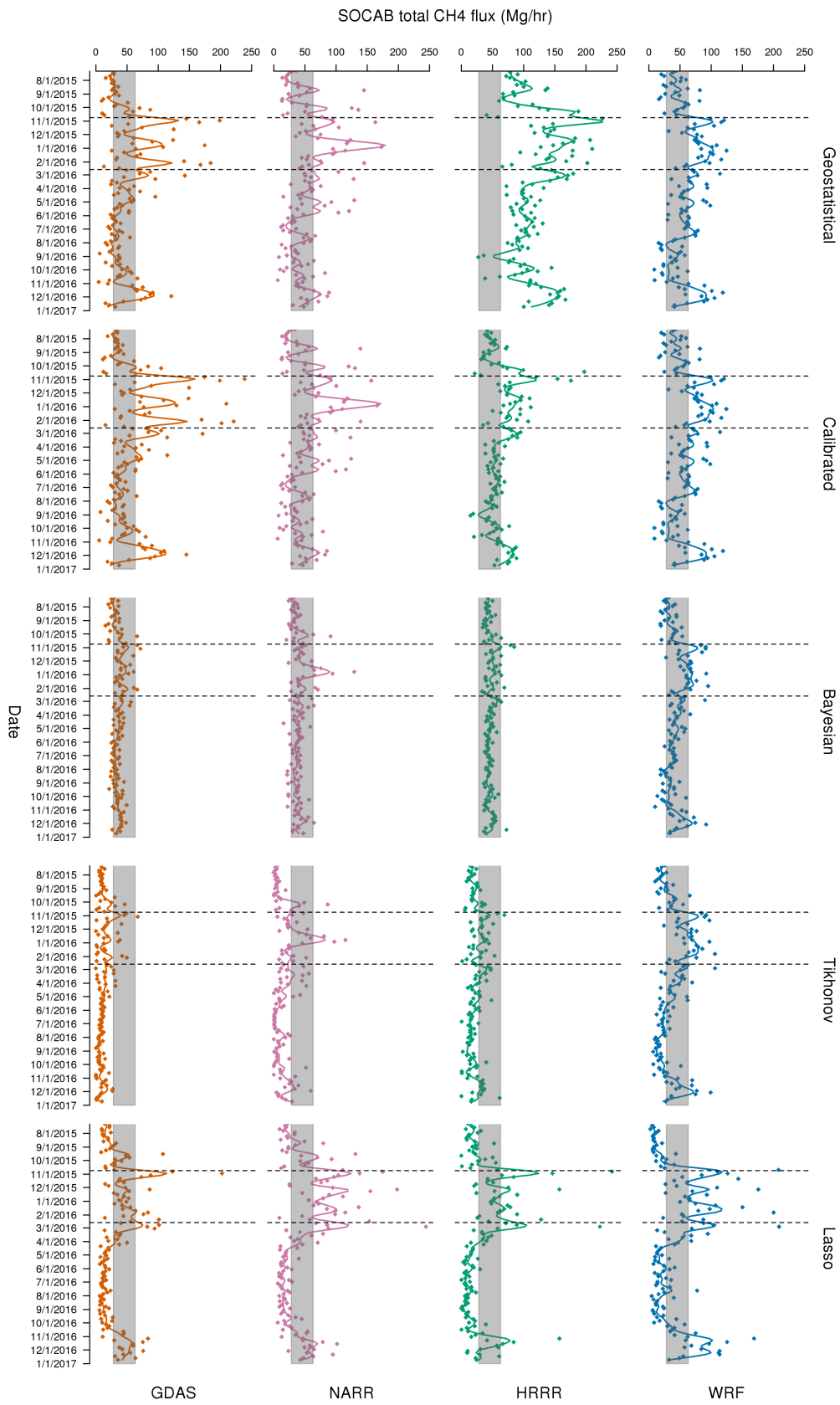
Figure 4.1: Estimated total CH4 flux time series for the South Coast Air Basin (SoCAB) according to inversions or regressions using four methods, plus the geostatistical inversion with sensitivity calibration, and with transport driven by each of four meteorological models. The shaded band indicates 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.
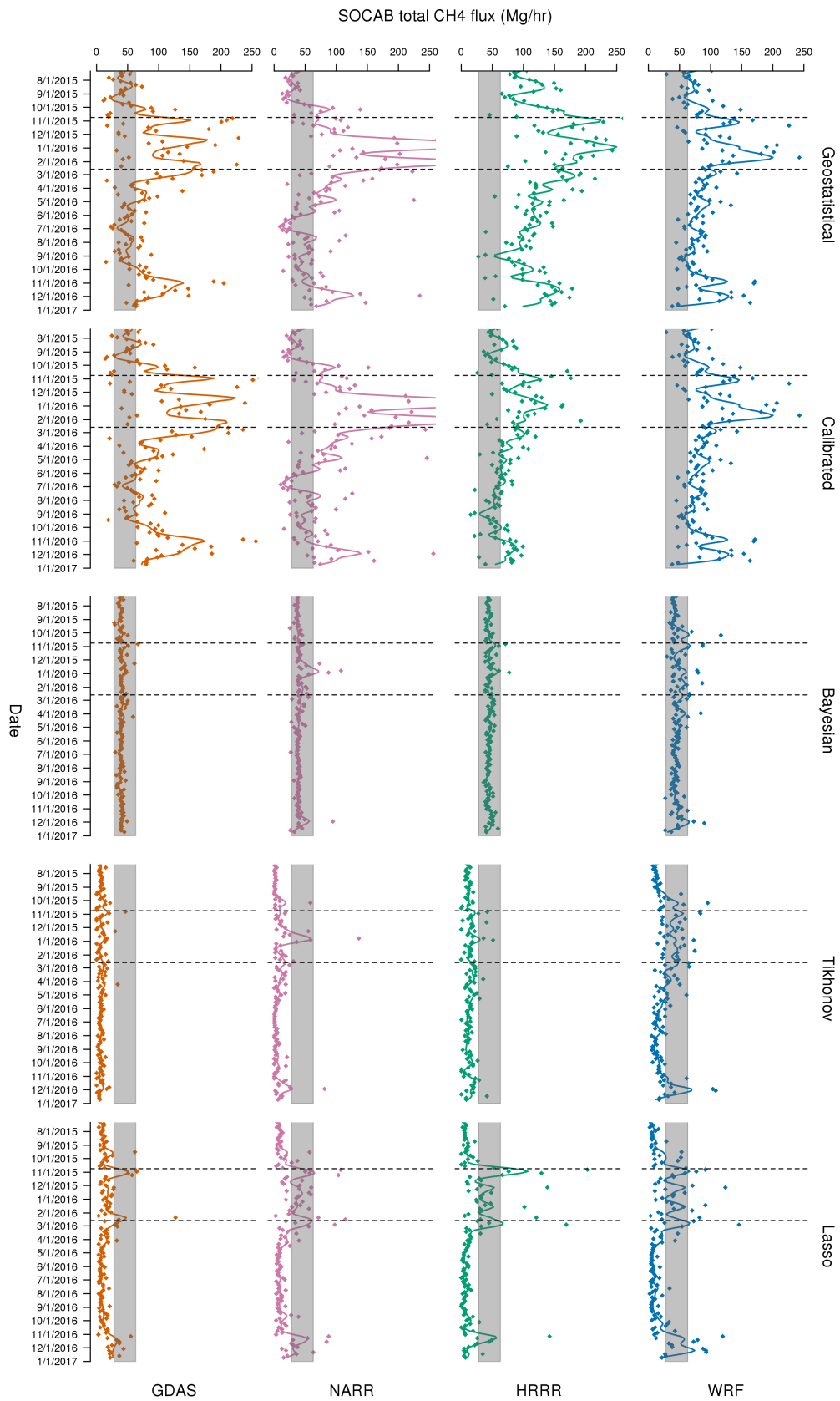
Figure 4.2: As in Figure 4.1, but using only three observing sites: USC, Fullerton, and Granada Hills. The geostatistical inversion becomes noisier, and the Bayesian inversion collapses towards the prior.

85

can be explained by the differences in total mean sensitivity assigned to the observing network. The flux estimates can be calibrated after the fact by scaling them according to those average differences in sensitivity. That calibration is neither necessary nor effective in the Bayesian inversions, in which the proportionality of flux estimates to sensitivity is diluted by the influence of the prior.

A direct comparison illuminates the impact of method choice on the flux estimate time series. Figure 4.3 shows linear RMA regression analyses of the Bayesian and calibrated geostatistical estimates. Each point represents the estimated fluxes during one four-day period. The coefficient of determination $r^2$ measures the similarity in the shape of the time series between the two methods, irrespective of any overall offset bias or difference of scale. Including observational data from the full LA Megacities network results in higher $r^2$ values, confirming our expectation that improving the data constraint should reduce the importance of the method choice. Among inversions making use of the full network, the highest $r^2$ is observed when transport is driven by WRF, the most highly-tuned and finest-resolution meteorological product.

### 4.3.1  Change Detection

We evaluate each method's ability to detect emissions flux changes and events by testing the significance of the increase due to the Aliso Canyon natural gas leak. As in Chapter III, we compare posterior flux estimates from October 24 to December 27, 2015 to the corresponding period in 2016. We use two complementary tests: Welch's t-test for a difference in the mean flux between the two periods, and an F-test for a difference in the variance of the flux estimates. Generally, the F-test is the more effective of the two in inversions using coarse driving meteorology (NARR or GDAS), which produce noisier flux estimates, whereas the t-test is the more effective in inversions using high-resolution meteorology (WRF or HRRR). Significance statistics for both tests are given in table 4.1.
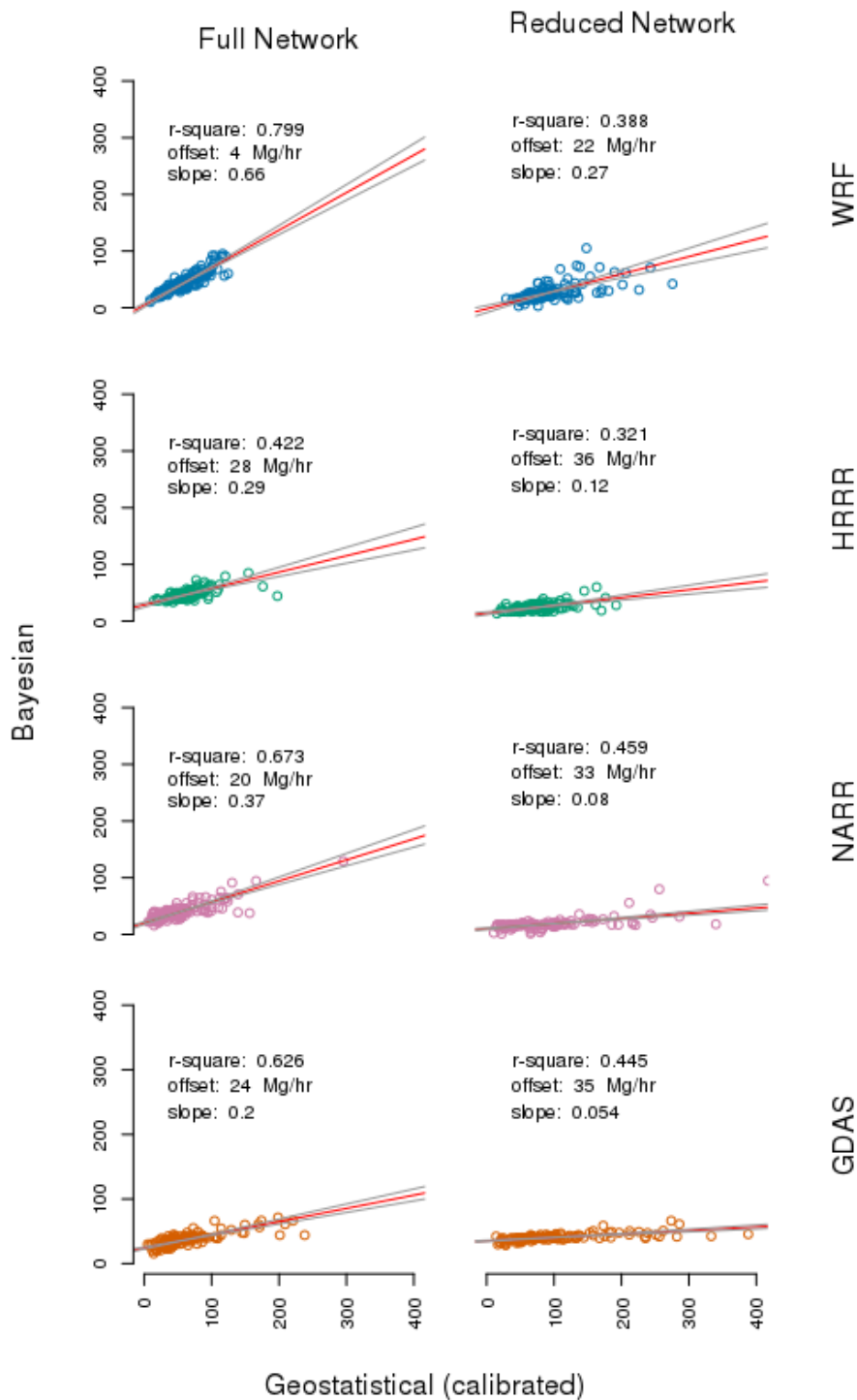
Figure 4.3: Scatter plots of four-day flux estimates using the calibrated geostatistical method (horizontal axis) and the Bayesian method (vertical axis), with the slope, offset, and r-square statistics of the RMA regression between the two. The timeseries are most similar (highest $r^2$) when more data is included (left column), especially when WRF is used to drive transport.

87

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

|  | WRF | HRRR | NARR | GDAS | With reduced network WRF | HRRR | NARR | GDAS |
|---|---|---|---|---|---|---|---|---|
| Geostatistical | 0.047* | <0.001* | 0.048* | 0.012* | 0.024* | <0.001* | 0.075 | <0.001* |
| Bayesian | 0.011* | 0.012* | 0.11 | 0.013* | 0.10 | 0.10 | 0.37 | 0.17 |
| Tikhonov | <0.001* | 0.027* | 0.0057* | 0.27 | 0.0051* | 0.12 | 0.087 | 0.93 |
| Lasso | <0.001* | 0.0017* | <0.001* | <0.001* | <0.001* | 0.034* | 0.0017* | <0.001* |

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

|  | WRF | HRRR | NARR | GDAS | With reduced network WRF | HRRR | NARR | GDAS |
|---|---|---|---|---|---|---|---|---|
| Geostatistical | 0.17 | 0.025* | 0.016* | 0.039* | 0.63 | 0.004* | 0.051 | 0.3 |
| Bayesian | 0.035* | 0.0082* | 0.081 | 0.027* | 0.67 | 0.31 | 0.39 | 0.13 |
| Tikhonov | 0.034* | 0.005* | 0.040* | 0.011* | 0.58 | 0.40 | 0.29 | 0.11 |
| Lasso | 0.92 | 0.049* | 0.004* | 0.34 | 0.77 | 0.082 | 0.19 | 0.91 |

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

|  | WRF | HRRR | NARR | GDAS | With reduced network WRF | HRRR | NARR | GDAS |
|---|---|---|---|---|---|---|---|---|
| Geostatistical | 0.32 | <0.001* | <0.001* | 0.044* | 0.6 | 0.056 | 0.016* | 0.021* |
| Bayesian | 0.63 | 0.062 | 0.0027* | 0.0057* | 0.52 | 0.17 | 0.42 | <0.001* |
| Tikhonov | 0.52 | 0.39 | 0.023* | 0.0031* | 0.20 | 0.88 | 0.059 | 0.0053* |
| Lasso | 0.55 | 0.055 | 0.0061* | 0.0016* | 0.71 | 0.075 | 0.76 | 0.087 |

Table 4.1: 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.

In analyses using the full LA Megacities observing network (nine sites within the domain), the Aliso Canyon event is significant ($p < 0.05$) in at least one of our two tests in almost every case, regardless of the inversion or regularization methodology and using any of HRRR, NARR, or GDAS (though not WRF) to drive transport. With only three observing sites, however, only the geostatistical method generally retains the ability to detect the leak: by the mean flux increase when driven by HRRR and by the increase in variance when driven by NARR or GDAS. In the other methods, the reduced observational constraint is not sufficient to overcome the influence of the prior (in the Bayesian inversion) or the downward pressure of the regularization.

Contrary to what we might have expected for a sparseness-promoting method, estimates with lasso regularization are not consistently more sensitive to the large, localized emissions from Aliso Canyon than are those with Tikhonov regularization. A likely reason is that the observations are too localized to ensure retrieval of the most important sparse components of the flux (*Candès and Wakin*, 2008).

### 4.3.2    Spatial Flux Localization

Ideally, an operational $CH_4$ monitoring system would not only detect emissions changes at the city scale but also locate them in space, allowing for detailed follow-up investigation. Without foreknowledge of the location of the Aliso Canyon natural gas leak, none of the methods tested here have that capability. Figure 4.4 shows maps of mean flux difference between the initial period of the leak, October 24 to December 27, 2015, and the corresponding comparison period in 2016. Figure 4.5 shows the pixel-by-pixel significance of those differences in Welch's t-test.

Several distinct patterns of flux differences are visible. Lasso regularization does produce large, localized changes in the flux, but they are of both signs, often with increases adjacent to decreases. Because the variance is very large, almost none of
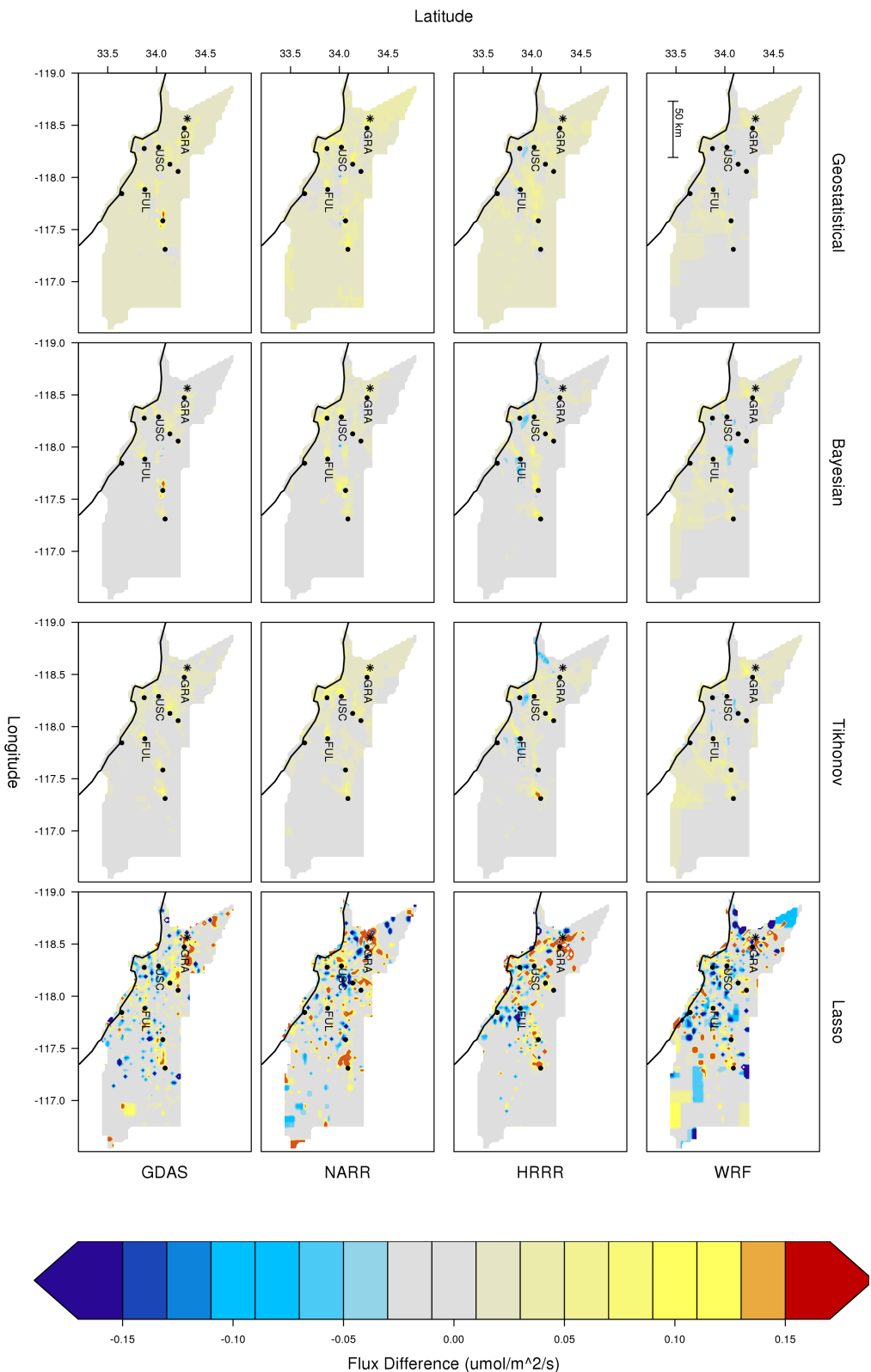
Figure 4.4: Differences in spatially explicit estimated CH$_4$ emissions fluxes between the first 48 days of the Aliso Canyon natural gas leak (October 24 to December 27, 2015) and the corresponding period in 2016, according to each of four inversion or regularization methods and using transport driven by each of four meteorological models or reanalysis products. Warm colors indicate higher fluxes in 2015 than in 2016; cool colors indicate lower fluxes in 2015.
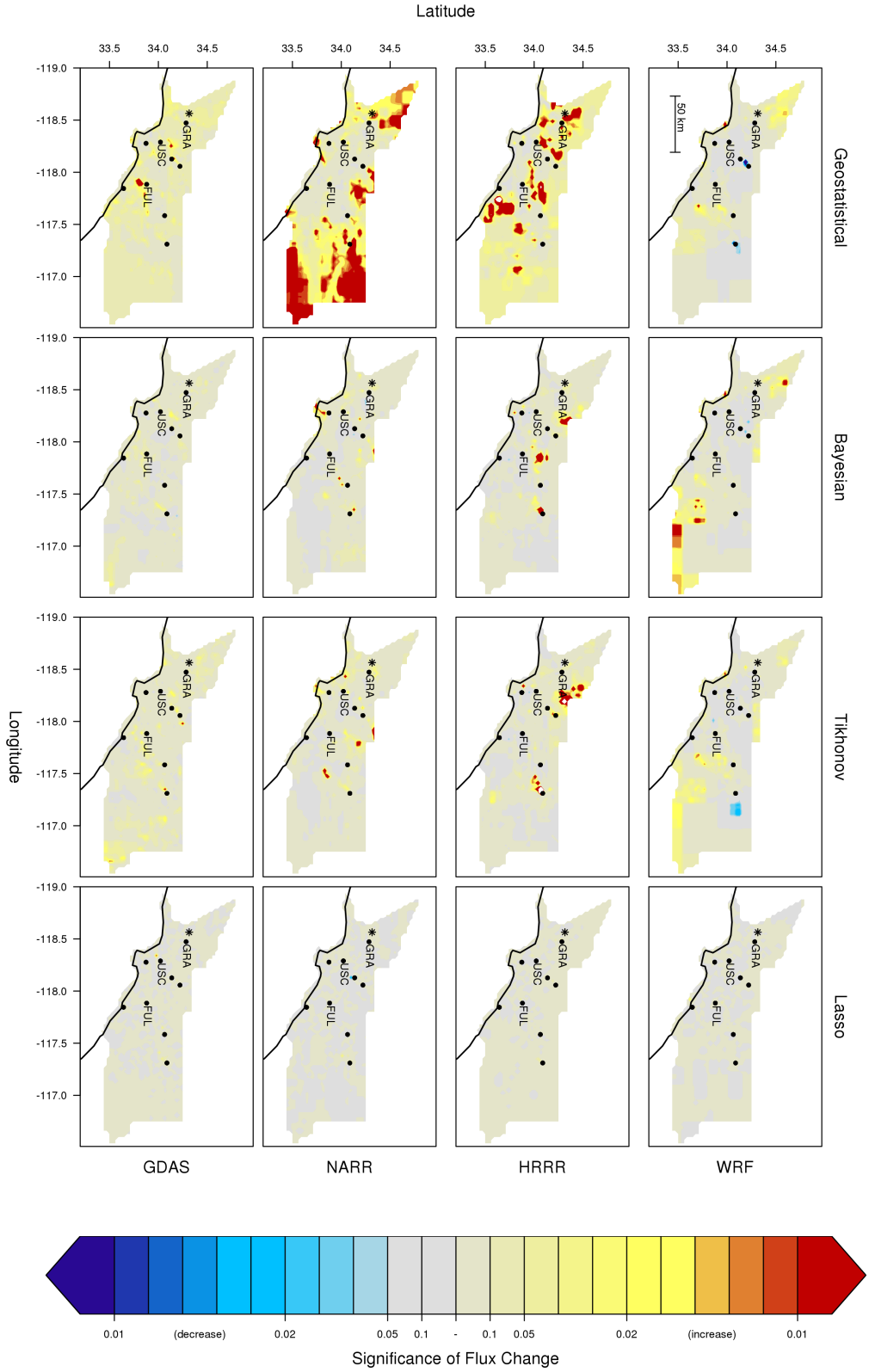
Figure 4.5: Significance (p-values) of the CH$_4$ emissions differences shown in Figure 4.4, according to Welch's t-test as applied individually to each 0.03 degree grid cell. Warm colors indicate significantly higher fluxes in 2015; cool colors, significantly higher fluxes in 2016; and neutral colors, no significant difference.

the changes are significant. The Bayesian inversion and the estimate using Tikhonov regularization both concentrate the flux increase near the measurement sites, where the sensitivity of the observations is highest. That accommodates increased observed mole fractions with the minimum increase in flux, so as to depart as little as possible from the prior (Bayesian) or from zero (Tikhonov). Some of the changes are significant, but not in the Aliso Canyon area.

In most of the geostatistical inversions, the flux increase is distributed broadly over the whole domain. This is a consequence of our model $\mathbf{X}$, which included a constant flux term; by increasing the drift coefficient corresponding to the constant term, the inversion can improve the fit to observations without incurring any other penalty. One result is that the flux estimates increase even in regions with little observational constraint, such as in the far east of the domain. Many of the changes are significant, but because they result mainly from an across-the-board increase, their spatial pattern is not informative. Interestingly, the geostatistical inversion driven by WRF shows a flux difference pattern more similar to the Bayesian inversions, lacking the broad increase.

## 4.4  Discussion

Using the same observations and geostatistical methodology we employ here and with transport driven by the same WRF configuration, *Yadav et al.* (2018) were able to identify a significant flux increase localized around the Aliso Canyon site by making two additional interventions. First, a moderate-size region encompassing the site was selected. Increasing the size of the region makes the test more robust to transport error, especially wind direction error, which can cause the inversion to place the increased flux incorrectly or inconsistently. However, in an operational monitoring context in which the location of a flux event is not known in advance, aggregated regions to be tested must be selected in some objective manner. A promising solution

is transdimensional Bayesian inversion (*Lunt et al.*, 2016), in which the flux estimates are made in aggregated regions the number and locations of which are treated as additional variables to be optimized.

In addition to testing an expert-selection region, *Yadav et al.* (2018) filtered the time series of flux estimates to remove periods during which the Granada Hills observing site, nearest to the leak location, had reduced sensitivity to that region. The necessity of subselecting for adequate sensitivity suggests that the observation coverage is insufficient. If a denser set of mole fraction observations were available, or if an event occurred in the interior of the domain, we might be more successful in locating it.

# CHAPTER V

# Conclusions

The effort to quantify greenhouse gas emissions at the urban scale is still new, and there are so far no generally agreed-upon standard methods. A variety of approaches are being explored and tested. This thesis is a contribution to that exploration, focusing on a few of the relevant factors: what can be learned independent of the meteorology used to drive transport, mixing-depth observations that might be applied to evaluate that driving meteorology, and the interactions between the observation density, inversion or regularization methodology, and choice of meteorology. But there is a many-directional space of other possible methods and observations and of choices we have left fixed in the work presented here.

## 5.1 Further Directions

Our discussion has focused on methane, but $CO_2$ emissions at the urban scale are also uncertain (*Gately and Hutyra*, 2017), and $CO_2$ flux estimation poses a different set of challenges. Some studies (*Hutyra et al.*, 2015; *Gurney et al.*, 2017) have recently shown that the $CO_2$ flux contribution from the urban biosphere can be substantial. For the purpose of estimating fluxes, the biosphere contribution introduces at least four complications:

1. An additional, non-anthropogenic component must be included in prior flux

estimates (in a Bayesian approach) or as an explanatory variable (in a geostatistical approach). In flux inversions on larger scales, this role has generally been filled by a process-based model. The selection and tuning of the biosphere model then become new axes of choice in the estimation method.

2. If the goal is to quantify fossil fuel emissions, those must be separated from other sources on the basis of their distribution in space and time, isotopic composition, or the presence of other co-emitted gases.

3. Because plants both take up and give off $CO_2$, the surface can be a net sink, so negative fluxes must be allowed. The inversion may then overfit the observations by introducing spurious dipoles – large, nearly canceling positive and negative fluxes in adjacent regions.

4. In our inversions, we have estimated methane emissions at a temporal cadence of four days or longer, implicitly assuming that the flux is more or less constant on that time scale. That assumption fails for biosphere $CO_2$ exchange in an especially pernicious manner. Photosynthetic uptake occurs during the day, when the mixing layer is generally deeper than at night, so the resulting decrease in mixing ratio near the surface is small. At night, plants release $CO_2$ into a much shallower nocturnal surface layer, causing a correspondingly larger increase in surface mixing ratio. On average over the course of 24 hours, the mixing ratio may be enhanced even if the net flux is zero or negative, an effect known as the diurnal rectifier (*Randall et al.*, 1996).

Carbon in fossil fuels is depleted of $^{14}C$, so measurements of isotopic composition can distinguish between carbon-containing compounds originating from fossil sources and from contemporary biogenic sources. Many studies have used isotopic analysis (sometimes of $^{13}C$ and/or isotopes of hydrogen or oxygen in addition to $^{14}C$) to identify the dominant urban source types of $CO_2$ (e.g., *Kuc and Zimnoch*, 1997;

*Djuricin et al.*, 2010) and $CH_4$ (e.g., *Kuc et al.*, 2003; *Townsend-Small et al.*, 2012) as well as CO (e.g., *Klouda et al.*, 1986; *Klouda and Connolly*, 1995; *Vimont et al.*, 2017). Generally, these studies have apportioned carbon in the atmosphere without linking it explicitly to surface fluxes, but exploratory studies (*Wu et al.*, 2018) have investigated the potential value of including isotopic measurements in an inversion system. However, expense and low observation frequency are significant barriers, since air must be captured in flasks for subsequent isotopic analysis off-site.

Apart from isotopic analysis, the presence of other co-emitted species may be used to apportion GHG emissions. The most commonly measured tracer for fossil fuel $CO_2$ is CO. Because it is preferentially emitted in incomplete combustion processes, especially in motor vehicles (*Turnbull et al.*, 2011, 2015; *Vogel et al.*, 2010), CO can also be used to distinguish between mobile and stationary sources of fossil fuel $CO_2$. CO is considered a less accurate tracer for fossil sources overall than $^{14}C$ (*Levin and Karstens*, 2007), but has the advantage that it can be measured in situ at high frequency. For methane source identification, a promising tracer is $C_2H_6$ (e.g., *Wennberg et al.*, 2012; *Smith et al.*, 2015), which is present in leaked natural gas but is not produced by microbial sources such as ruminants, landfills, and wastewater treatment. *Nathan et al.* (2018) provide a comprehensive review and assessment of other tracers in the urban context. So far, though, no urban flux inversion study using real data has incorporated multi-species observations.

Urban flux inversions so far, including those presented here, have relied on in situ mole fraction observations using a relatively small number (no more than 12) of high-precision instruments. However, a significant body of non-inversion work on urban greenhouse gas emissions has incorporated observations of the total abundance of $CO_2$ and/or $CH_4$ in a column of the atmosphere (e.g., *Kort et al.*, 2012), and flux inversions on larger scales have used total column observations from satellites (e.g. *Turner et al.*, 2015). Because the mixing layer in which the local emissions are concentrated makes

up typically only about 10% of the atmosphere, column observations must be able to detect much smaller enhancements in order to be useful, so instrument precision is more likely to be a significant concern than with in situ measurements. As argued by *McKain et al.* (2012), though, the relationship between column abundance and fluxes is less sensitive to the mixing depth, so errors in modeled transport may be less impactful. Recent prospective studies have looked at what precision, field of view, and measurement frequency future satellite observations would need in order to spatially constrain urban $CO_2$ (*Broquet et al.*, 2018) or concentrated $CH_4$ sources (*Turner et al.*, 2018) in an inversion framework.

High-density networks of low-cost sensors have also been receiving attention as an alternative or supplement to expensive, high-precision spectroscopic analyzers for in situ observations. The BErkeley Atmospheric $CO_2$ Observing Network (BEACON) (*Shusterman et al.*, 2016) takes this approach, deploying low-cost sensors at schools in the San Francisco Bay Area, California. In a study of simulated observations from variations on the BEACON network, *Turner et al.* (2016) concluded that increasing the density of observations is more important than ensuring better than moderate precision.

In order to attribute enhanced mixing ratios or column abundances to local sources, the background air entering the domain of interest must be characterized. In the work presented in Chapters III and IV, we estimated background using the method of *Verhulst et al.* (2017) to filter and smooth observations at an offshore site. That approach is generally reasonable in Los Angeles, where the dominant wind pattern brings in clean air from over the Pacific Ocean. Still, *Verhulst et al.* (2017) found that the uncertainty of the in-basin enhancements is dominated by the background estimate, rather than by the measurement precision at the primary observing sites.

A more sophisticated background estimation method might identify the predominant wind direction at any given time and select the most appropriate upwind site,

as was done in Indianapolis by *Lauvaux et al.* (2016). In some larger-scale flux inversions (e.g. *Lunt et al.*, 2016), the background has been specified in a spatiotemporally varying way over a whole vertical curtain at the edge of the study domain, according to the output of a global circulation model. Then each observation can be assigned a unique background value according to when and where the sampled air masses cross the boundary curtain according to the transport model.

As shown in Chapters III and IV, whatever observations are employed, choices made in the course of the inversion analysis may also have a substantial effect on the outcome. Our discussions of transport have focused on the driving meteorology, keeping the Lagrangian model STILT fixed and running it always in its default configuration. Especially on the urban scale where there may be strong sources in the very near field of observing sites, though, the details of the transport simulation may have more significant effects than have been recognized. So far, literature is sparse on the choice of transport models, their configuration and parameterization, and the effects on flux inversion. Outside the urban context, *Hegarty et al.* (2013) compared several Lagrangian models by releasing a known inert tracer, and several studies (*Lauvaux et al.*, 2012; *Pillai et al.*, 2012) have compared Lagrangian to Eulerian transport models in an inversion context.

A parallel effort has sought to characterize sources of error in detail so as to specify covariance structures as realistically as possible. Much of that work has focused on transport error. For example, *Lin and Gerbig* (2005) devised a method for estimating the additional model-data mismatch variance due to errors in the modeled wind speed, as follows: the wind field in the driving meteorological model is compared to observations and the statistics of the deviation, including decorrelation scales in space and time, are computed. Those wind error statistics are then implemented as an additional stochastic component of the wind in the transport model. The increase in the spread of forward-modeled mole fractions $\mathbf{Hs}_0$ when the wind errors are included

in **H** is taken as an estimate of the transport error variance and incorporated into the model-data mismatch covariance matrix **R**. *Gerbig et al.* (2007) proposed to handle mixing depth error in a similar way, by stochastically varying the mixing depth in the transport model according to the observed error statistics.

## 5.2   Concluding Thoughts

A great deal of work remains to be done before we will fully understand the importance of the many choices involved in observational urban greenhouse gas flux estimation. The results presented here, together with other recent investigations in a variety of directions, suggest that the situation is likely to be complex. Continuous observations of the mixing depth can reveal bias in meteorological models, but that bias may not propagate transparently into flux estimates. A high-resolution, locally validated WRF configuration can reduce the impact of method choice, but it may not be necessary or even beneficial in monitoring for emissions changes.

If there is a lesson to be learned, it may be that no single method is optimal for all purposes. We will need to consider our goals and the relevant conditions and make judicious choices in each case – including whether to prioritize the best possible methods or to give weight to operational requirements and the establishment of a standard for direct comparison.

# APPENDIX

# APPENDIX A

# RML Cost Function for Tikhonov Regularization

In Tikhonov regularization, the likelihood of the observations is

$$p(\mathbf{z}|\theta) = \int d\mathbf{s}\ p(\mathbf{z}|\mathbf{s})p(\mathbf{s}) \tag{A.1}$$

$$\propto |\mathbf{R}|^{-1/2}\gamma^{n/2}\int d\mathbf{s}\ \exp\left\{-\frac{1}{2}\Big[(\mathbf{z}-\mathbf{Hs})^T\mathbf{R}^{-1}(\mathbf{z}-\mathbf{Hs})+\gamma\mathbf{s}^T\mathbf{s}\Big]\right\} \tag{A.2}$$

where $n$ is the dimension of $\mathbf{s}$, i.e., the number of fluxes to be estimated. The integral is Gaussian, so

$$p(\mathbf{z}|\theta) \propto |\mathbf{R}|^{-1/2}\gamma^{n/2}|\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}+\gamma\mathbf{I}_n|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{z}^T\mathbf{R}^{-1}\mathbf{z}\right\}$$

$$\times \exp\left\{\frac{1}{2}\mathbf{z}^T\mathbf{R}^{-1}\mathbf{H}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}+\gamma\mathbf{I}_n)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}\right\} \tag{A.3}$$

$$= \Big[\gamma^{-m}|\gamma\mathbf{R}||\mathbf{H}^T(\gamma\mathbf{R})^{-1}\mathbf{H}+\mathbf{I}_n|\Big]^{-1/2}$$

$$\times \exp\left\{\frac{1}{2}\mathbf{z}^T\mathbf{R}^{-1}\Big[\mathbf{H}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}+\gamma\mathbf{I}_n)^{-1}\mathbf{H}^T\mathbf{R}^{-1}-\mathbf{I}_m\Big]\mathbf{z}\right\} \tag{A.4}$$

where $m$ is the number of observations (so $\mathbf{R}$ is $m\times m$). Applying the matrix identity

$$(\mathbf{V}^T\mathbf{B}^{-1}\mathbf{V}+\mathbf{A}^{-1})^{-1}\mathbf{V}^T\mathbf{B}^{-1} = \mathbf{A}\mathbf{V}^T(\mathbf{V}\mathbf{A}\mathbf{V}^T+\mathbf{B})^{-1} \tag{A.5}$$

gives

$$p(\mathbf{z}|\theta) \propto \left[\gamma^{-m}|\gamma\mathbf{R}||\mathbf{H}^T(\gamma\mathbf{R})^{-1}\mathbf{H} + \mathbf{I}_n|\right]^{-1/2}$$

$$\times \exp\left\{\frac{1}{2}\mathbf{z}^T\mathbf{R}^{-1}\left[\mathbf{H}(\gamma^{-1}\mathbf{I}_n)\mathbf{H}^T(\mathbf{H}(\gamma^{-1}\mathbf{I}_n)\mathbf{H}^T + \mathbf{R})^{-1} - \mathbf{I}_m\right]\mathbf{z}\right\} \quad (A.6)$$

$$= \left[\gamma^{-m}|\gamma\mathbf{R}||\mathbf{H}^T(\gamma\mathbf{R})^{-1}\mathbf{H} + \mathbf{I}_n|\right]^{-1/2}$$

$$\times \exp\left\{\frac{1}{2}\mathbf{z}^T\mathbf{R}^{-1}\left[\mathbf{I}_m - \mathbf{R}(\mathbf{H}(\gamma^{-1}\mathbf{I}_n)\mathbf{H}^T + \mathbf{R})^{-1} - \mathbf{I}_m\right]\mathbf{z}\right\} \quad (A.7)$$

$$= \left[\gamma^{-m}|\gamma\mathbf{R}||\mathbf{H}^T(\gamma\mathbf{R})^{-1}\mathbf{H} + \mathbf{I}_n|\right]^{-1/2}$$

$$\times \exp\left\{-\frac{1}{2}\mathbf{z}^T(\gamma^{-1}\mathbf{H}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{z}\right\}. \quad (A.8)$$

The first factor can be simplified using the Matrix Determinant Lemma

$$|\mathbf{A}^{-1} + \mathbf{U}\mathbf{V}^T| = |\mathbf{A}^{-1}||\mathbf{V}^T\mathbf{A}\mathbf{U} + \mathbf{I}|, \quad (A.9)$$

such that

$$p(\mathbf{z}|\theta) \propto \left[\gamma^{-m}|\gamma\mathbf{R} + \mathbf{H}\mathbf{H}^T|\right]^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{z}^T(\gamma^{-1}\mathbf{H}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{z}\right\} \quad (A.10)$$

$$= |\Phi|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{z}^T\Phi^{-1}\mathbf{z}\right\}, \qquad \Phi = \gamma^{-1}\mathbf{H}\mathbf{H}^T + \mathbf{R}. \quad (A.11)$$

The cost function is therefore

$$\boxed{\mathcal{L}_\theta = \ln|\Phi| + \mathbf{z}^T\Phi^{-1}\mathbf{z}} \quad (A.12)$$

# BIBLIOGRAPHY

# BIBLIOGRAPHY

Aznar, A., M. Day, E. Doris, S. Mathur, and P. Donohoo-Vallett (2015), City-level energy decision making: Data use in energy planning, implementation, and evaluation in US cities, *Tech. rep.*, National Renewable Energy Laboratory (NREL), Golden, CO (United States).

Baars, H., A. Ansmann, R. Engelmann, and D. Althausen (2008), Continuous monitoring of the boundary-layer top with lidar, *Atmospheric Chemistry and Physics*, *8*(23), 7281, doi:10.5194/acp-8-7281-2008.

Ballantyne, A., R. Andres, R. Houghton, B. Stocker, R. Wanninkhof, W. Anderegg, L. Cooper, M. DeGrandpre, P. Tans, J. Miller, et al. (2015), Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty, *Biogeosciences*, *12*(8).

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.

Bousquet, P., P. Peylin, P. Ciais, C. Le Quéré, P. Friedlingstein, and P. P. Tans (2000), Regional changes in carbon dioxide fluxes of land and oceans since 1980, *Science*, *290*, 1342–1346.

Brandt, A. R., G. Heath, E. Kort, F. O'sullivan, G. Pétron, S. Jordaan, P. Tans, J. Wilcox, A. Gopstein, D. Arent, et al. (2014), Methane leaks from North American natural gas systems, *Science*, *343*(6172), 733–735.

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.

Broquet, G., F.-M. Bréon, E. Renault, M. Buchwitz, M. Reuter, H. Bovensmann, F. Chevallier, L. Wu, and P. Ciais (2018), The potential of satellite spectro-imagery for monitoring $CO_2$ emissions from large cities, *Atmospheric Measurement Techniques*, *11*(2), 681.

Burgin, L., J. Gloster, C. Sanders, P. Mellor, S. Gubbins, and S. Carpenter (2013), Investigating incursions of bluetongue virus using a model of long-distance Culicoides biting midge dispersal, *Transboundary and Emerging Diseases*, *60*(3), 263–272.

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.

Cambaliza, M., P. Shepson, J. Bogner, D. Caulton, B. Stirm, C. Sweeney, S. Montzka, K. Gurney, K. Spokas, O. Salmon, et al. (2015), Quantification and source apportionment of the methane emission flux from the city of Indianapolis, *Elementa: Science of the Anthropocene*, *3*.

Candès, E. J., and M. B. Wakin (2008), An introduction to compressive sampling, *IEEE signal processing magazine*, *25*(2), 21–30.

Clark-Thorne, S. T., and C. J. Yapp (2003), Stable carbon isotope constraints on mixing and mass balance of $CO_2$ in an urban atmosphere: Dallas metropolitan area, Texas, USA, *Applied Geochemistry*, *18*(1), 75–95.

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.

Coulter, R. L. (1979), A comparison of three methods for measuring mixing-layer height, *Journal of Applied Meteorology*, *18*, 1495–1499, doi:10.1175/1520-0450(1979)018<1495:ACOTMF>2.0.CO;2.

Crippa, M., G. Janssens-Maenhout, F. Dentener, D. Guizzardi, K. Sindelarova, M. Muntean, R. Van Dingenen, and C. Granier (2016), Forty years of improvements in European air quality: Regional policy-industry interactions with global impacts, *Atmospheric Chemistry and Physics*, *16*(6), 3825–3841.

Davis, K. J., D. H. Lenschow, S. P. Oncley, C. Kiemle, G. Ehret, A. Giez, and J. Mann (1997), Role of entrainment in surface-atmosphere interactions over the boreal forest, *Journal of Geophysical Research*, *102*, 29,219–29,230.

Davis, K. J., N. Gamage, C. R. Hagelberg, C. Kiemle, D. H. Lenschow, and P. P. Sullivan (2000), An objective method for deriving atmospheric structure from airborne lidar observations, *Journal of Atmospheric and Oceanic Technology*, *17*(11), 1455, doi:10.1175/1520-0426(2000)017.

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*.

Di Giuseppe, F., A. Riccio, L. Caporaso, G. Bonafé, G. P. Gobbi, and F. Angelini (2012), Automatic detection of atmospheric boundary layer height using ceilometer backscatter data assisted by a boundary layer model, *Quarterly Journal of the Royal Meteorological Society*, *138*, 649–663, doi:10.1002/qj.964.

Djuricin, S., D. E. Pataki, and X. Xu (2010), A comparison of tracer methods for quantifying $CO_2$ sources in an urban region, *Journal of Geophysical Research: Atmospheres*, *115*(D11).

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

Ehret, G., A. Giez, C. Kiemle, K. J. Davis, D. H. Lenschow, S. P. Oncley, and R. D. Kelly (1996), Airborne water vapor DIAL and in situ observations of a sea-land interface, *Beiträge zur Physik der Atmosphäre*, *69*, 215–228.

Endlich, R. M., F. L. Ludwig, and E. E. Uthe (1979), An automatic method for determining the mixing depth from lidar observations, *Atmospheric Environment*, *13*(7), 1051–1056.

Enting, I., C. Trudinger, and R. Francey (1995), A synthesis inversion of the concentration and $\delta^{13}c$ of atmospheric $co_2$, *Tellus B*, *47*(1-2), 35–52.

Eresmaa, N., A. Karppinen, S. M. Joffre, J. Räsänen, and H. Talvitie (2006), Mixing height determination by ceilometer, *Atmospheric Chemistry and Physics*, *6*(6), 1485–1493, doi:10.5194/acp-6-1485-2006.

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.

Flesch, T. K., J. D. Wilson, and E. Yee (1995), Backward-time Lagrangian stochastic dispersion models and their application to estimate gaseous emissions, *Journal of Applied Meteorology*, *34*(6), 1320–1332.

Flynn, C. J., A. Mendoza, Y. Zheng, and S. Mathur (2007), Novel polarization-sensitive micropulse lidar measurement technique, *Optics Express*, *15*(6), 2785–2790, doi:10.1364/oe.15.002785.

Gan, C. M., Y. Wu, B. L. Madhavan, B. Gross, and F. Moshary (2011), Application of active optical sensors to probe the vertical structure of the urban boundary layer and assess anomalies in air quality model PM 2.5 forecasts, *Atmospheric Environment*, *45*, 6613–6621, doi:10.1016/j.atmosenv.2011.09.013.

Gately, C., and L. Hutyra (2017), Large uncertainties in urban-scale carbon emissions, *Journal of Geophysical Research: Atmospheres*, *122*(20).

Gerbig, C., S. Körner, and J. Lin (2007), Vertical mixing in atmospheric tracer transport models: Error characterization and propagation, *Atmospheric Chemistry and Physics Discussions*, *7*(5), 13,121–13,150.

Gloster, J., A. Jones, A. Redington, L. Burgin, J. H. Sørensen, R. Turner, M. Dillon, P. Hullinger, M. Simpson, P. Astrup, et al. (2010), Airborne spread of foot-and-mouth disease–Model intercomparison, *The Veterinary Journal*, *183*(3), 278–286.

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., D. L. Mendoza, Y. Zhou, M. L. Fischer, C. C. Miller, S. Geethakumar, and S. de la Rue du Can (2009), High resolution fossil fuel combustion $CO_2$ emission fluxes for the United States, *Environmental science & technology*, *43*(14), 5535–5541.

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.

Gurney, K. R., J. Liang, R. Patarasuk, D. O'Keeffe, J. Huang, M. Hutchins, T. Lauvaux, J. C. Turnbull, and P. B. Shepson (2017), Reconciling the differences between a bottom-up and inverse-estimated FFCO$_2$ emissions estimate in a large US urban area, *Elem Sci Anth*, *5*.

Haeffelin, M., F. Angelini, G. P. Gobbi, Y. Morille, G. Martucci, C. D. O'Dowd, S. Frey, S. Lolli, L. Sauvage, I. Xueref-Rémy, B. Wastine, and D. G. Feist (2012), Evaluation of mixing-height retrievals from automatic profiling lidars and ceilometers in view of future integrated networks in Europe, *Boundary-Layer Meteorology*, *143*, 49–75, doi:10.1007/s10546-011-9643-z.

Hanna, S. (1984), Applications in air pollution modeling, in *Atmospheric turbulence and air pollution modelling*, pp. 275–310, Springer.

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.

He, Q. S., J. T. Mao, J. Y. Chen, and Y. Y. Hu (2006), Observational and modeling studies of urban atmospheric boundary-layer height and its evolution mechanisms, *Atmospheric Environment*, *40*, 1064–1077, doi:10.1016/j.atmosenv.2005.11.016.

Heffter, J. L. (1980), Air resources laboratories atmospheric transport and dispersion model, *Tech. Rep. ERL ARL-81*, NOAA.

Hegarty, J., R. R. Draxler, A. F. Stein, J. Brioude, M. Mountain, J. Eluszkiewicz, T. Nehrkorn, F. Ngan, and A. Andrews (2013), Evaluation of Lagrangian particle dispersion models with measurements from controlled tracer releases, *Journal of Applied Meteorology and Climatology*, *52*(12), 2623–2637.

Hennemuth, B., and A. Lammert (2006), Determination of the atmospheric boundary layer height from radiosonde and lidar backscatter, *Boundary-Layer Meteorology*, *120*, 181–200, doi:10.1007/s10546-005-9035-3.

Hooper, W. P., and E. W. Eloranta (1986), Lidar measurements of wind in the planetary boundary layer: the method, accuracy and results from joint measurements with radiosonde and kytoon, *Journal of Climate and Applied Meteorology*, *25*, 990–1001.

Hutyra, L., C. Gately, S. Decina, A. Reinmann, P. Templer, T. Nehrkorn, and S. Wofsy (2015), Attribution of urban greenhouse gas fluxes: Does the biosphere in cities matter?, in *AGU Fall Meeting Abstracts*.

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).

Kaminski, T., P. J. Rayner, M. Heimann, and I. G. Enting (2001), On aggregation errors in atmospheric transport inversions, *Journal of Geophysical Research: Atmospheres*, *106*(D5), 4703–4715.

Kinra, S., G. Lewendon, R. Nelder, N. Herriott, R. Mohan, M. Hort, S. Harrison, and V. Murray (2005), Evacuation decisions in a chemical air pollution incident: Cross sectional survey, *British Medical Journal*, *330*(7506), 1471.

Klouda, G., L. Currie, D. Donahue, A. Jull, and M. Naylor (1986), Urban Atmospheric $^{14}CO$ and $^{14}CH_4$ Measurements by Accelerator Mass Spectrometry, *Radiocarbon*, *28*(2A), 625–633.

Klouda, G. A., and M. V. Connolly (1995), Radiocarbon ($^{14}C$) measurements to quantify sources of atmospheric carbon monoxide in urban air, *Atmospheric Environment*, *29*(22), 3309–3318.

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.

Kuc, T., and M. Zimnoch (1997), Changes of the $CO_2$ sources and sinks in a polluted urban area (southern Poland) over the last decade, derived from the carbon isotope composition, *Radiocarbon*, *40*(1), 417–423.

Kuc, T., K. Rozanski, M. Zimnoch, J. M. Necki, and A. Korus (2003), Anthropogenic emissions of $CO_2$ and $CH_4$ in an urban environment, *Applied Energy*, *75*(3-4), 193–203.

Kuze, H., H. Kinjo, Y. Sakurada, and N. Takeuchi (1998), Field-of-view dependence of lidar signals by use of Newtonian and Cassegrainian telescopes, *Applied Optics*, *37*(15), 3128–3132, doi:10.1364/ao.37.003128.

Lammert, A., and J. Bösenberg (2006), Determination of the convective boundary-layer height with laser remote sensing, *Boundary-Layer Meteorology*, *119*(1), 159–170, doi:10.1007/s10546-005-9020-x.

Lauvaux, T., A. Schuh, M. Uliasz, S. Richardson, N. Miles, A. Andrews, C. Sweeney, L. Diaz, D. Martins, P. Shepson, et al. (2012), Constraining the $CO_2$ budget of the corn belt: Exploring uncertainties from the assumptions in a mesoscale inverse system, *Atmospheric Chemistry and Physics*, *12*(1), 337.

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.

Lauvaux, T., N. L. Miles, A. Deng, S. J. Richardson, M. O. Cambaliza, K. J. Davis, B. Gaudet, K. R. Gurney, J. Huang, D. O'Keefe, et al. (2016), High-resolution atmospheric inversion of urban $CO_2$ emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX), *Journal of Geophysical Research: Atmospheres*, *121*(10), 5213–5236.

Levin, I., and U. Karstens (2007), Inferring high-resolution fossil fuel $CO_2$ records at continental sites from combined $^{14}CO_2$ and CO observations, *Tellus B*, *59*(2), 245–250.

Lewis, J. R., E. J. Welton, A. M. Molod, and E. Joseph (2013), Improved boundary layer depth retrievals from MPLNET, *Journal of Geophysical Research: Atmospheres*, *118*, 9870–9879, doi:10.1002/jgrd.50570.

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).

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.

Lunt, M. F., M. Rigby, A. L. Ganesan, and A. J. Manning (2016), Estimation of trace gas fluxes with objectively determined basis functions using reversible-jump Markov Chain Monte Carlo, *Geoscientific Model Development*, *9*(9), 3213–3229.

Maasakkers, J. D., D. J. Jacob, M. P. Sulprizio, A. J. Turner, M. Weitz, T. Wirth, C. Hight, M. DeFigueiredo, M. Desai, R. Schmeltz, et al. (2016), Gridded national inventory of US methane emissions, *Environmental Science & Technology*, *50*(23), 13,123–13,133.

Marsik, F. J., K. W. Fischer, T. D. McDonald, and P. J. Samson (1995), Comparison of methods for estimating mixing height used during the 1992 Atlanta Field Intensive, *Journal of Applied Meteorology*, *34*, 1802–1814.

Martin, C. L., D. Fitzjarrald, M. Garstang, A. P. Oliveira, S. Greco, and E. Browell (1988), Structure and growth of the mixing layer over the Amazonian rain forest, *Journal of Geophysical Research: Atmospheres*, *93*(D2), 1361–1375, doi:10.1029/JD093iD02p01361.

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.

McKendry, I. G., D. Van der Kamp, K. B. Strawbridge, A. Christen, and B. Crawford (2009), Simultaneous observations of boundary-layer aerosol layers with CL31 ceilometer and 1064/532 nm lidar, *Atmospheric Environment*, *43*(36), 5847–5852, doi:10.1016/j.atmosenv.2009.07.063.

Menut, L., C. Flamant, J. Pelon, and P. H. Flamant (1999), Urban boundary-layer height determination from lidar measurements over the Paris area, *Applied Optics*, *38*, 945–954.

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).

Michalak, A. M., A. Hirsch, L. Bruhwiler, K. R. Gurney, W. Peters, and P. P. Tans (2005), Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, *Journal of Geophysical Research: Atmospheres*, *110*(D24).

Munkel, C., and J. Rasanen (2004), New optical concept for commercial lidar ceilometers scanning the boundary layer, *Proceedings of SPIE, 5571, Remote Sensing of Clouds and the Atmosphere IX*, 364–374, doi:10.1117/12.565540.

Münkel, C., N. Eresmaa, J. Räsänen, and A. Karppinen (2006), Retrieval of mixing height and dust concentration with lidar ceilometer, *Boundary-Layer Meteorology*, *124*(1), 117–128, doi:10.1007/s10546-006-9103-3.

Münkel, C., K. Schäfer, and S. Emeis (2011), Adding confidence levels and error bars to mixing layer heights detected by ceilometer, *Proceedings of SPIE, 8177, Remote Sensing of Clouds and the Atmosphere XVI*, doi:10.1117/12.898122.

Nathan, B., T. Lauvaux, J. Turnbull, and K. Gurney (2018), Investigations into the use of multi-species measurements for source apportionment of the indianapolis fossil fuel $co_2$ signal, *Elem Sci Anth*, *6*(1).

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 (2013a), WRF simulations of the urban circulation in the Salt Lake City area for $CO_2$ modeling, *Journal of Applied Meteorology and Climatology*, *52*, 323–340, doi:10.1175/JAMC-D-12-061.1.

Nehrkorn, T., J. Henderson, M. Leidner, M. Mountain, J. Eluszkiewicz, K. McKain, and S. Wofsy (2013b), 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.

Newman, S., X. Xu, H. P. Affek, E. Stolper, and S. Epstein (2008), Changes in mixing ratio and isotopic composition of $CO_2$ in urban air from the Los Angeles basin, California, between 1972 and 2003, *Journal of Geophysical Research: Atmospheres*, *113*(D23).

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

Oda, T., and S. Maksyutov (2011), A very high-resolution (1 km $\times$ 1 km) global fossil fuel $CO_2$ emission inventory derived using a point source database and satellite observations of nighttime lights, *Atmospheric Chemistry and Physics*, *11*(2), 543–556.

Oda, T., T. Lauvaux, D. Lu, P. Rao, N. L. Miles, S. J. Richardson, and K. R. Gurney (2017), On the impact of granularity of space-based urban $CO_2$ emissions in urban atmospheric inversions: A case study for Indianapolis, IN, *Elem Sci Anth*, *5*.

Pataki, D., D. Bowling, and J. Ehleringer (2003), Seasonal cycle of carbon dioxide and its isotopic composition in an urban atmosphere: Anthropogenic and biogenic effects, *Journal of Geophysical Research: Atmospheres*, *108*(D23).

Patarasuk, R., K. R. Gurney, D. OâĂŹKeeffe, Y. Song, J. Huang, P. Rao, M. Buchert, J. C. Lin, D. Mendoza, and J. R. Ehleringer (2016), Urban high-resolution fossil fuel $CO_2$ emissions quantification and exploration of emission drivers for potential policy applications, *Urban ecosystems*, *19*(3), 1013–1039.

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.

Peters, W., A. R. Jacobson, C. Sweeney, A. E. Andrews, T. J. Conway, K. Masarie, J. B. Miller, L. M. P. Bruhwiler, G. Petron, A. I. Hirsch, D. E. J. Worthy, G. R. var der Werf, J. T. Randerson, P. O. Wennberg, M. C. Krol, and P. P. Tans (2007), An atmospheric perspective on North American carbon dioxide exchange:

CarbonTracker, *Proceedings of the National Academy of Sciences*, *104*, 18,925–18,930, doi:10.1073/pnas.0708986104.

Pillai, D., C. Gerbig, R. Kretschmer, V. Beck, U. Karstens, B. Neininger, and M. Heimann (2012), Comparing Lagrangian and Eulerian models for $CO_2$ transport–a step towards Bayesian inverse modeling using WRF/STILT-VPRM, *Atmospheric Chemistry and Physics*, *12*(19), 8979–8991.

Pozzer, A., and R. H. H. Janssen (2015), Description and implementation of a MiXed Layer model (MXL, v1. 0) for the dynamics of the atmospheric boundary layer in the Modular Earth Submodel System (MESSy), *Geoscientific Model Development*, *8*(3), 453–471, doi:10.5194/gmd-8-453-2015.

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

Randall, D., G. Collatz, and P. Sellers (1996), Simulations of terrestrial carbon metabolism and atmospheric $CO_2$ in a general circulation model. Part II: Simulated $CO_2$ concentrations, *Tellus B*, *48*, 543–567.

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*.

Rosenzweig, C., W. Solecki, S. A. Hammer, and S. Mehrotra (2010), Cities lead the way in climate-change action, *Nature*, *467*(7318), 909.

Ryerson, T., A. Andrews, W. Angevine, T. Bates, C. Brock, B. Cairns, R. Cohen, O. Cooper, J. Gouw, F. Fehsenfeld, et al. (2013), The 2010 California research at the Nexus of air quality and climate change (CalNex) field study, *Journal of Geophysical Research: Atmospheres*, *118*(11), 5830–5866.

Schuh, A. E., A. S. Denning, K. D. Corbin, I. T. Baker, M. Uliasz, N. Parazoo, A. E. Andrews, and D. E. J. Worthy (2010), A regional high-resolution carbon flux inversion of North America for 2004, *Biogeosciences*, *7*, 1625–1644, doi:10.5194/bg-7-1625-2010.

Seibert, P., F. Beyrich, S. E. Gryning, S. Joffre, A. Rasmussen, and P. Tercier (2000), Review and intercomparison of operational methods for the determination of the mixing height, *Atmospheric Environment*, *34*, 1001–1027, doi:10.1016/S1352-2310(99)00349-0.

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.

Smith, M. L., E. A. Kort, A. Karion, C. Sweeney, S. C. Herndon, and T. I. Yacovitch (2015), Airborne ethane observations in the Barnett Shale: Quantification of ethane

flux and attribution of methane emissions, *Environmental science & technology*, *49*(13), 8158–8166.

Spinhirne, J. D. (1982), Lidar clear atmosphere multiple scattering dependence on receiver range, *Applied Optics*, *21*(14), 2467–2468, doi:10.1364/ao.21.002467.

Steyn, D. G., M. Baldi, and R. M. Hoff (1999), The detection of mixed layer depth and entrainment zone thickness from lidar backscatter profiles, *Journal of Atmospheric and Oceanic Technology*, *16*(7), 953–959, doi:10.1175/1520-0426(1999)016.

Stohl, A., A. Prata, S. Eckhardt, L. Clarisse, A. Durant, S. Henne, N. I. Kristiansen, A. Minikin, U. Schumann, P. Seibert, et al. (2011), Determination of time-and height-resolved volcanic ash emissions and their use for quantitative ash dispersion modeling: the 2010 Eyjafjallajökull eruption, *Atmospheric Chemistry and Physics*, *11*(9), 4333–4351.

Stull, R. (1988), *An introduction to boundary layer meteorology*, Kluwer Academic Publishers.

Tans, P. P., I. Y. Fung, and T. Takahashi (1990), Observational constraints on the global atmospheric $CO_2$ budget, *Science*, *247*, 1431–1438, doi: 10.1126/science.247.4949.1431.

Tatarov, B., T. Trifonov, B. Kaprielov, and I. Kolev (2000), Dependence of the lidar signal depolarization on the receiver's field of view in the sounding of fog and clouds, *Applied Physics B: Lasers and Optics*, *71*(4), 593–600, doi:10.1007/s003400000265.

Thomson, D., and J. Wilson (2013), History of Lagrangian stochastic models for turbulent dispersion, *Lagrangian Modeling of the Atmosphere*, pp. 19–36.

Tibshirani, R. (1996), Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

Townsend-Small, A., S. C. Tyler, D. E. Pataki, X. Xu, and L. E. Christensen (2012), Isotopic measurements of atmospheric methane in Los Angeles, California, USA: Influence of âĂIJfugitiveâĂİ fossil fuel emissions, *Journal of Geophysical Research: Atmospheres*, *117*(D7).

Turnbull, J., A. Karion, M. Fischer, I. Faloona, T. Guilderson, S. Lehman, B. Miller, J. Miller, S. Montzka, T. Sherwood, et al. (2011), Assessment of fossil fuel carbon dioxide and other anthropogenic trace gas emissions from airborne measurements over Sacramento, California in spring 2009, *Atmospheric Chemistry and Physics*, *11*(2), 705–721.

Turnbull, J. C., C. Sweeney, A. Karion, T. Newberger, S. J. Lehman, P. P. Tans, K. J. Davis, T. Lauvaux, N. L. Miles, S. J. Richardson, M. O. Cambaliza, P. B. Shepson, K. Gurney, R. Patarasuk, and I. Razlivanov (2015), Toward quantification and source sector identification of fossil fuel $CO_2$ emissions from an urban

area: Results from the INFLUX experiment, *Journal of Geophysical Research: Atmospheres*, *120*(1), 292–312, doi:10.1002/2014jd022555.

Turner, A., D. J. Jacob, K. Wecht, J. Maasakkers, E. Lundgren, A. Andrews, S. Biraud, H. Boesch, K. Bowman, N. M. Deutscher, et al. (2015), Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data, *Atmospheric Chemistry and Physics*.

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.

Turner, A. J., D. J. Jacob, J. Benmergui, J. Brandman, L. White, and C. A. Randles (2018), Assessing the capability of different satellite observing configurations to resolve the distribution of methane emissions at kilometer scales, *Atmospheric Chemistry and Physics Discussions*, *2018*, 1–23, doi:10.5194/acp-2018-164.

van Stratum, B. J. H., J. V.-G. de Arellano, H. G. Ouwersloot, K. van den Dries, T. W. van Laar, M. Martinez, J. Lelieveld, J.-M. Diesch, F. Drewnick, H. Fischer, Z. Hosaynali Beygi, H. Harder, E. Regelin, Sinha, J. A. Adame, M. Sorgel, R. Sander, H. Bozem, W. Song, J. Williams, and N. Yassaa (2012), Case study of the diurnal variability of chemically active species with respect to boundary layer dynamics during DOMINO, *Atmospheric Chemistry and Physics*, *12*(12), 5329–5341, doi:10.5194/acp-12-5329-2012.

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.

Vimont, I. J., J. C. Turnbull, V. V. Petrenko, P. F. Place, A. Karion, N. L. Miles, S. J. Richardson, K. Gurney, R. Patarasuk, C. Sweeney, et al. (2017), Carbon monoxide isotopic measurements in Indianapolis constrain urban source isotopic signatures and support mobile fossil fuel emissions as the dominant wintertime CO source, *Elementa: Science of the Anthropocene*, *5*.

Vogel, F. R., S. Hammer, A. Steinhof, B. Kromer, and I. Levin (2010), Implication of weekly and diurnal ^{14}C calibration on hourly estimates of CO-based fossil fuel $CO_2$ at a moderately polluted site in southwestern Germany, *Tellus B*, *62*(5), 512–520.

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.

Webster, H., D. Thomson, B. Johnson, I. Heard, K. Turnbull, F. Marenco, N. Kristiansen, J. Dorsey, A. Minikin, B. Weinzierl, et al. (2012), Operational prediction of ash concentrations in the distal volcanic cloud from the 2010 Eyjafjallajökull eruption, *Journal of Geophysical Research: Atmospheres*, *117*(D20).

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.

Wiegner, M., F. Madonna, I. Binietoglou, R. Forkel, J. Gasteiger, A. Geiss, G. Pappalardo, K. Schafer, and W. Thomas (2014), What is the benefit of ceilometers for aerosol remote sensing? An answer from EARLINET, *Atmospheric Measurement Techniques*, *7*, 1979–1997, doi:10.5194/amt-7-1979-2014.

Witt, E. C., III (2015), Geospatial resources for the geology community: the USGS National Map, *The Journal of Geology*, *123*(3), 283–294, doi:10.1086/682008.

Wong, C. K., T. J. Pongetti, T. Oda, P. Rao, K. R. Gurney, S. Newman, R. M. Duren, C. E. Miller, Y. L. Yung, and S. P. Sander (2016), Monthly trends of methane emissions in Los Angeles from 2011 to 2015 inferred by CLARS-FTS observations, *Atmospheric Chemistry and Physics*, *16*(20), 13,121–13,130.

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.

Worden, H. M., Y. Cheng, G. Pfister, G. R. Carmichael, Q. Zhang, D. G. Streets, M. Deeter, D. P. Edwards, J. C. Gille, and J. R. Worden (2012), Satellite-based estimates of reduced CO and $CO_2$ emissions due to traffic restrictions during the 2008 Beijing Olympics, *Geophysical Research Letters*, *39*(14).

Wu, K., T. Lauvaux, K. J. Davis, A. Deng, I. L. Coto, K. R. Gurney, and R. Patarasuk (2018), Joint inverse estimation of fossil fuel and biogenic $CO_2$ fluxes in an urban environment: An observing system simulation experiment to assess the impact of multiple uncertainties, *Elem Sci Anth*, *6*(1).

Wunch, D., P. Wennberg, G. Toon, G. Keppel-Aleks, and Y. Yavin (2009), Emissions of greenhouse gases from a North American megacity, *Geophysical Research Letters*, *36*(15).

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).

Zimnoch, M., J. Godlowska, J. M. Necki, and K. Rozanski (2010), Assessing surface fluxes of $CO_2$ and $CH_4$ in urban environment: a reconnaissance study in Krakow, Southern Poland, *Tellus B*, *62*, 573–580, doi:10.1111/j.1600-0889.2010.00489.x.