Emissions of Nitrous Oxide and Methane in North America

Citation	Miller, Scot M. 2015. Emissions of Nitrous Oxide and Methane in North America. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467371
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Emissions of nitrous oxide and methane in North America

a dissertation presented
by
Scot Michael Miller
to
The Department of Earth and Planetary Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Earth and Planetary Sciences

Harvard University
Cambridge, Massachusetts
April 2015

Thesis advisor: Professor Steven C. Wofsy		Scot Michael Miller

Emissions of nitrous oxide and methane in North America

Abstract

Methane (CH_4) and nitrous oxide (N_2O) are the second- and third-most important long-lived greenhouse gas species after carbon dioxide (CO_2) in terms of radiative forcing. This thesis describes the magnitude, spatial distribution, and seasonality of N_2O and CH_4 sources over North America using atmospheric data. We also investigate the environmental drivers and/or anthropogenic source sectors that can explain these emissions patterns. Overall, this thesis provides information on the magnitude, distribution, and likely drivers of greenhouse gas emissions to aid existing or future climate change mitigation policies in the US and Canada.

We estimate anthropogenic N_2O and CH_4 emissions that greatly exceed most existing inventory estimates. Our US budgets for N_2O and CH_4 are approximately 2.8 and 1.5 times higher, respectively, than inventory estimates from the US EPA. Much of the discrepancy in methane appears to stem from oil and natural gas industry and agricultural emissions.

In contrast, we estimate natural CH_4 sources that are smaller than most existing process-based biogeochemical models. These estimated fluxes have a spatial distribution centered around the Hudson Bay Lowlands. Most existing models estimate fluxes that are far more spatially distributed across the Canadian shield. These estimates provide negative information on the spatial distribution of fluxes relative to a spatially-constant model. We find that a simple model using only three environmental variables can describe flux patterns (as seen by the atmospheric observations) as well as any process-based estimate.

Contents

Listing of figures

Acknowledgments

I would like to thank all of my co-authors, without whom this dissertation would not be possible: Eric Kort, Adam Hirsch, Ed Dlugokencky, Arlyn Andrews, Xiaofeng Xu, Hanqin Tian, Thomas Nehrkorn, Janusz Eluszkiewicz, Talya Havice, Jed Kaplan, Patricia Levi, Bowen Zhang, Sebastien Biraud, Marc Fischer, Greet Janssens-Maenhout, Ben Miller, John Miller, Stephen Montzka, Colm Sweeney, Matt Hayek, Inez Fung, Junjie Liu, Josh Benmergui, Joseph Melton, and Doug Worthy. I am also indebted to my PhD advisors, Steven Wofsy and Anna Michalak. Both have been brilliant mentors throughout the past six years. I would also like to thank the entire Wofsy group for their support and for being a amazing group of people to work with. Last but not least, I would like tot thank several organizations for funding my PhD: American Meteorological Society Graduate Fellowship, Department of Energy Computational Science Graduate Fellowship, and the National Science Foundation Graduate Research Fellowship. Because of these fellowships, I had immense freedom in my PhD to pursue my intellectual interests, and for that I am extremely grateful.

0

Introduction

Greenhouse gas emissions play a pivotal role in climate change, but the distribution of these emissions from different sources across the globe is often poorly understood. This thesis focuses on sources of nitrous oxide (N_2O) and methane (CH_4) estimated using atmospheric observations.

N_2O is the third-most important long-lived greenhouse gas after carbon dioxide (CO_2) and CH_4 (Butler, 2014). A significant fraction of N_2O emissions occur in agricultural soils when nitrogen in fertilizer or manure changes chemical form and escapes to the atmosphere (e.g., Rahn & Wahlen, 2000; Rockmann et al., 2003). Large uncertainties remain in knowledge of continental-scale N_2O emissions and the underlying soil processes. For example, existing inventories disagree by a factor of five on total US emissions and on where the largest emissions occur – whether in Iowa or the Gulf Coast (European Comission, Joint Research Center (JRC)/ Netherlands Environmental Assessment Agency (PBL), 2009; Tian et al., 2010).

In contrast to N_2O, CH_4 is emitted to the atmosphere from a diverse array of sources, including wetlands, agriculture, natural gas operations, and landfills (e.g., Ciais et al., 2013; Kirschke et al., 2013). Existing emissions inventories, however, provide little consensus on national-scale

US emissions, and these uncertainties have manifested recently in a contentious public debate over hydraulic fracturing and natural gas extraction. Bottom-up anthropogenic inventories from US EPA and EDGAR give totals ranging from 19.6 to 30 TgC $\mathrm{yr}^{?1}$ (Olivier & Peters, 2005; US EPA, 2013). The most recent EPA and EDGAR inventories report lower US emissions compared to previous versions (decreased by 10% and 35%, respectively) (US EPA, 2013; European Comission, Joint Research Center (JRC)/ Netherlands Environmental Assessment Agency (PBL), 2009); this change primarily reflects lower, revised emissions estimates from natural gas and coal production.

Wetland methane estimates are as uncertain as anthropogenic inventories. Over time, biogeochemical models of these fluxes have become increasingly sophisticated or complex. However, these models still disagree by an order of magnitude over countries like Canada (e.g., Melton et al., 2013a). Furthermore, the inputs into biogeochemical methane models, including maps of soil moisture and soil carbon, are highly uncertain, as are the effects of those inputs on methane fluxes (e.g., Melton et al., 2013a; Wania et al., 2013).

This thesis explores N_2O and CH_4 emissions from a variety of sources across North America in order to determine the relative contribution of different human activities and to facilitate improved projections of future atmospheric greenhouse gas concentrations. Effective reduction strategies for greenhouse gas emissions depend upon reliable estimates of those emissions and a means to track future changes.

In particular, the studies presented here answer these questions using atmospheric greenhouse gas observations, a so-called ‘top-down’ approach. Atmospheric observations ‘see’ the cumulative effect of emissions at the Earth’s surface. Emissions inventories, by contrast, often use an accounting process to tally up individual sources or flux processes. These estimates are based upon detailed emissions factors for individual processes. A challenge of this approach is that uncertainties compound when estimates are summed over multiple source types and extrapolated to national scale. Atmospheric observations provide an objective tool to determine whether the sum of all sources in the inventory equals the total methane measured in the atmosphere.

Traditionally, bottom-up and top-down estimates have been difficult to reconcile because the former relies upon detailed emissions factors and the latter often provides only aggregate emissions totals. In this dissertation, we use atmospheric data to estimate more than just total emissions but also to understand how and why these gases were emitted – a goal that bridges the disconnect between biologists and government regulators who create the bottom-up models and scientists who use top-down atmospheric data.

In addition to analyzing empirical data, the research presented here also explores new ways to utilize limited, available greenhouse gas data. We further develop statistical and meteorological modeling frameworks to extract the most information possible from a diverse set of atmospheric, weather, economic, and land use data. For example, chapter 6 explores how to construct inverse problems that have a known constraint (e.g., the emissions must be non-negative). Another chapter delves into the effects of spatially- and temporally-correlated atmospheric transport errors on estimated greenhouse gas fluxes. In that study, we examined these errors in context of CO_2.

This thesis examines all of the questions posed above through seven individual chapters or studies. Chapter one focuses on emissions of N_2O while chapters two through five explore both anthropogenic and natural methane sources. The last two chapters (six through seven) examine two different facets of how to best utilize limited atmospheric greenhouse gas observations.

1

Regional sources of nitrous oxide over the United States: seasonal variation and spatial distribution

This chapter originally appeared as a research article in the Journal of Geophysical Research – Atmospheres. For the original article, refer to Miller et al. (2012b).

This chapter presents top-down constraints on the magnitude, spatial distribution, and seasonality of nitrous oxide (N_2O) emissions over the central United States. We analyze data from tall towers in 2004 and 2008 using a high resolution Lagrangian particle dispersion model paired with both geostatistical and Bayesian inversions. Our results indicate peak N_2O emissions in June with a strong seasonal cycle. The spatial distribution of sources closely mirrors data on fertilizer application with particularly large N_2O sources over the US Cornbelt. Existing inventories for N_2O predict emissions that differ substantially from the inverse model results in both seasonal cycle and magnitude. We estimate a total annual N_2O budget over the central US of 0.9 – 1.2 TgN/yr and an extrapolated budget for the entire US and Canada of 2.1 – 2.6 TgN/yr. By this estimate, the US and Canada account for 12 – 15% of the total global N_2O source or 32–39% of the global anthropogenic source as reported by the IPCC in 2007.

1.1 Introduction

Nitrous oxide (N_2O) plays a critical role in both stratospheric ozone depletion and climate change. In the stratosphere, it can react with excited oxygen atoms to produce NO_x radicals which, in turn, can catalyze stratospheric ozone destruction (Nevison & Holland, 1997; Ravishankara et al., 2009). Moreover, N_2O is a potent greenhouse gas with a global warming potential of 300 (on a 100 year timescale) (Forster et al., 2007). With the continued decline of atmospheric CFC concentrations, N_2O is now the third most important long-lived anthropogenic greenhouse gas in terms of radiative forcing (0.173 W/m^2 in 2009, an 11% increase since 1998) (Butler, 2014).

Anthropogenic sources of nitrous oxide include agriculture, fossil fuel combustion, and biomass burning (Commisssion, 2006; Denman et al., 2007). Isotopic N_2O measurements suggest that agriculture and associated fertilizer use are the largest anthropogenic N_2O sources globally (Rahn & Wahlen, 2000; Rockmann et al., 2003). The atmospheric lifetime of N_2O is relatively long at about 114 years, and the primary sink of N_2O is loss in the stratosphere (Montzka, 2003). Total annual global emissions (∼17.7 TgN/yr) are approximately 40 – 50% higher than in pre-industrial times and continue to rise (Denman et al., 2007). As a result, global N_2O mixing ratios have been increasing steadily by about 0.2 – 0.3% per year (about 0.73 ± 0.06 ppb/year or 3.5 TgN/yr) (Hirsch et al., 2006).

Because of the long lifetime of N_2O, spatial and temporal changes in N_2O mixing ratios are very small and remain challenging to measure accurately. For example, the seasonal cycle at Mace Head, Ireland, is only 0.1% relative to the mean abundance (Hirsch et al., 2006). The combination of a long lifetime, diffuse sources, and a low signal-to-noise ratio in the measurements means that the N_2O increment due to regional sources is very difficult to estimate and verify (Hirsch et al., 2006). In addition, the percentage of nitrogen released as N_2O after fertilizer application is very uncertain and ranges from 0% – 7 % (Bouwman, 1996). The diversity of fertilizer application processes and conditions results in emissions that show large spatial and temporal fluctuations and are difficult to constrain. As a result, existing estimates of N_2O emissions (both magnitudes and seasonal trends) are highly uncertain (Hirsch et al., 2006; Kort et al., 2008, 2010, 2011). Current emissions estimates over North America, for example, may be too low by as much as a factor of three (Kort et al., 2008, 2010).

The high degree of uncertainty in N_2O emissions complicates efforts to regulate sources under global and regional climate change agreements. At the global scale, the Kyoto Protocol included N_2O as one of the six greenhouse gases targeted for reductions. Within the US, several state and regional regulations target N_2O emissions as well (NEG/ECP, 2001; Western Climate Initiative, 2010). For example, the Midwestern Greenhouse Gas Reduction Accord, an agreement among 6 states in the midwestern US, recommends a 20% reduction in 2005 greenhouse gas emissions by 2020. Advisory group recommendations include N_2O as one of the six targeted categories of species (Advisory group to the Midwestern Greenhouse Gas Reduction Accord (MGGRA), 2009). Given uncertainties in emissions estimates that are ∼20x larger than reduction goals, it is very difficult at present to make reliable baseline inventories and track progress toward designated targets.

Several existing studies provide global to regional-scale constraints on N_2O emissions. Prinn et al. (1990), Hirsch et al. (2006), Huang et al. (2008), and Kort et al. (2011) derived top-down emissions constraints over large global regions. More recent studies provide regional-scale emissions estimates over Europe (Manning et al., 2003; Thompson et al., 2011; Corazza et al., 2011; Manning et al., 2011). Top-down estimates over North America are more limited. Kort et al. (2008, 2010) used tall tower and aircraft data to make continental-scale improvements to existing emissions estimates for North America. Multiple studies consistently find northern hemisphere emissions that are significantly higher than existing inventories (Hirsch et al., 2006; Huang et al., 2008; Kort et al., 2008, 2010).

The present paper provides detailed spatial and temporal information on the magnitude of N_2O emissions over the United States at much finer resolution than previously possible. We combine data from a network of tall tower measurements with high-resolution atmospheric simulations to provide regional-scale emissions estimates in time and space, representing important baseline information for greenhouse gas regulation.

1.2 The Model-Data Framework

1.2.1 Background to the STILT model

This study relies on STILT, the Stochastic Time-Inverted Lagrangian Transport Model, for simulations of atmospheric N_2O concentrations. STILT is a particle-following model based strongly on the HYSPLIT model (Draxler & Hess, 1998). STILT releases an ensemble of imaginary air particles from a receptor point, a single location in space and time. For our study, we set up the model to send 100 particles 10 days backward in time following the winds in an assimilated mesoscale meteorological model. The surface sources that these particles encounter are used to calculate the contribution of continental sources seen at the receptor point. In particular, STILT assumes that all surface sources will be well-mixed up to half the modeled boundary layer height. Particles below this height see influence from surface sources while those above do not. STILT then calculates an influence footprint based on the number and duration of particles in this surface layer. The influence footprint, when multiplied by an emissions inventory and summed over all geographic regions, provides an estimate of the continental source signal seen at the receptor. This continental signal is added to a modeled boundary condition (i.e. – the mixing ratio in 'clean' air before reaching North America) to estimate the total mixing ratio seen at the receptor point. The very detailed rendition of mixing ratio fluctuations provided by STILT can be validated against individual measurements taken at the receptor (usually a tall tower or aircraft mission), providing a powerful framework for assessing upwind surface or volume sources. Previous studies have applied STILT to a wide range of atmospheric trace gases, including CO_2, CO, methane, and N_2O (i.e. – Gerbig et al., 2003; Lin et al., 2003; Lin & Gerbig, 2005; Matross et al., 2006; Kort et al., 2008; Miller et al., 2008; Gourdji et al., 2010; Thompson et al., 2011). Lin et al. (2003) and Gerbig et al. (2003) provide a detailed explanation of STILT model theory and structure.

1.2.2 Underlying meteorological drivers

STILT simulations in this study use Weather Research and Forecasting (WRF) assimilated meteorological fields version 2.2 (Skamarock et al., 2005; Nehrkorn et al., 2010) and the Brazilian Regional Atmospheric Modeling System (BRAMS v. 3.2) (Pielke et al., 1992; Cotton et al., 2003; Sanchez-Ccoyllo et al., 2006). BRAMS simulations (45-km resolution) were only available for the year 2004. The limited STILT-BRAMS runs are complemented by STILT-WRF runs for all time periods.

WRF model simulations use a nested meteorological grid resolution; 10-km resolution wind fields drive particle trajectories over most of the continental US and southern Canada. 40-km resolution wind fields drive trajectories as the particles travel to more distant regions. This nested wind field structure affords higher fidelity source attribution near the receptor while still maintaining reasonable computational costs. Nehrkorn et al. (2010) provide a full description of the WRF simulations used in STILT.

BRAMS wind fields (v. 3.2) were based strongly on the Regional Atmospheric Modeling System (RAMS) (Cotton et al., 2003) with several updates: new parameterizations for convection (shallow and deep) and turbulence along with modified diagnostic outputs to ensure mass conservation to very high accuracy. All simulations apply a mass conservation fix from Medvigy et al. (2005). STILT parameterizes boundary layer turbulence as a Markov chain process (Lin et al., 2003), and both WRF and BRAMS use a Grell - Devenyi scheme for convection (Grell & Devenyi, 2002). BRAMS estimates boundary layer height in accordance with Vogelezang & Holtslag (1996), and WRF uses a Yonsei University Scheme (Fast, 2005). These two approaches give somewhat different estimates for PBL height and exchange rates (see Section 1.4.1).

1.2.3 Emissions inventories

We pair the STILT model with four different N_2O emissions inventories: EDGAR v. 4.0, EDGAR 32FT 2000, GEIA, and the Dynamic Land-Ecosystem Model (DLEM) (see Fig. 1.1). The EDGAR inventories include anthropogenic N_2O sources (including sources from agriculture) on a 0.1° by 0.1° and 1° by 1° longitude-latitude grid, respectively, for v. 4 and 32FT2000 (Olivier & Peters, 2005; European Comission, Joint Research Center (JRC)/ Netherlands Environmental Assessment Agency (PBL), 2009). EDGAR v. 4, incorporated into STILT at a $\frac{1}{4}^\circ$ longitude by $\frac{1}{6}^\circ$ latitude resolution, is the more recent, higher-resolution update to EDGAR 32FT2000; while the newer inventory has a similar spatial distribution, the estimated magnitude of fluxes is lower. GEIA incorporates both anthropogenic and natural sources, also on a 1° by 1° grid (Bouwman et al., 1995). All three inventories are constant in time.

DLEM, a process-based biogeochemical model, estimates N_2O from natural sources and from agriculture (Tian et al., 2010, 2011). The N_2O model simulates nitrification and denitrification processes as a function of ammonium and nitrate concentrations, soil temperature, and soil moisture. According to Tian et al. (2010), DLEM model parameters are then optimized using several Ameriflux sites. Daily emissions estimates are available from 2000 – 2008 at a 32 km x 32 km resolution (incorporated into STILT on a $\frac{1}{4}^\circ$ by $\frac{1}{6}^\circ$ longitude latitude grid) . For all STILT simula-

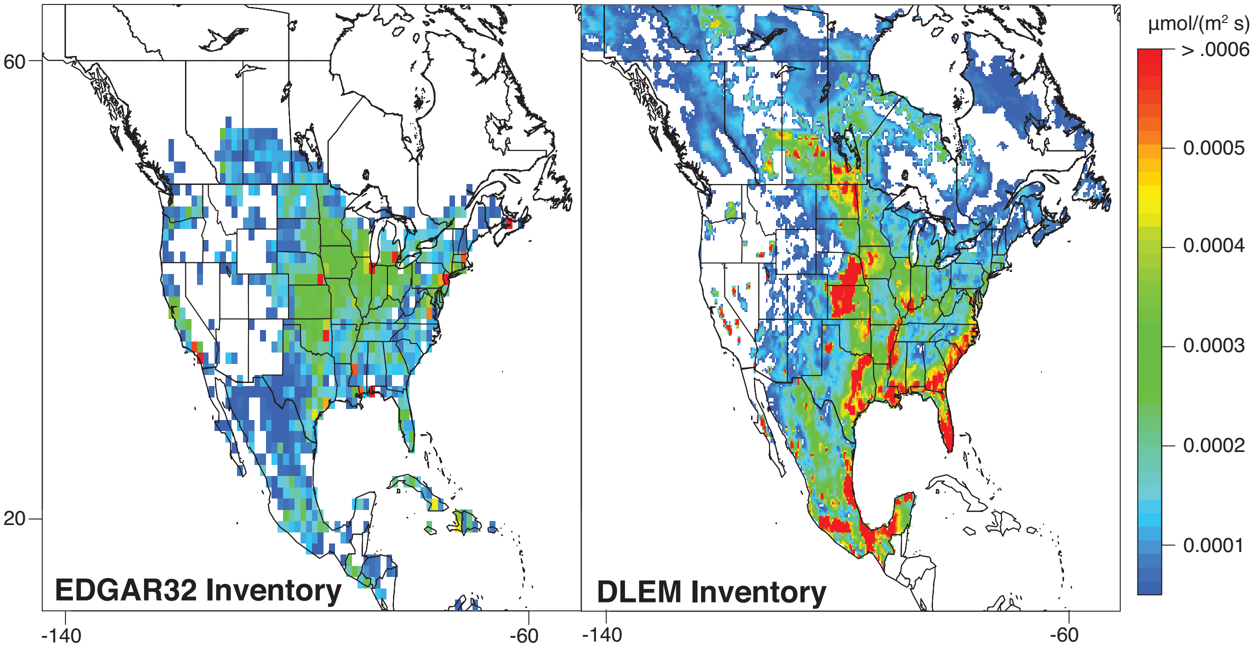


Figure 1.1: Graphical plots of the different N_2O emissions inventories for 2000. Because DLEM is a daily inventory, the plot above shows average fluxes for May – August 2000. EDGAR v.4 and GEIA (not shown) have similar distributions to EDGAR32 but smaller magnitude (particularly EDGAR v.4).

tions with DLEM, we supplement the inventory with anthropogenic source categories from the EDGAR32FT2000 inventory (i.e. – fossil fuel use, waste handling, water treatment, etc.).

1.2.4 Model boundary condition

The model boundary condition represents N_2O mixing ratios in air that is advected from the Pacific or Arctic Oceans before entering North America. We construct an empirical boundary based on measurements from the NOAA cooperative air sampling network. We use monthly mean dry air mole fractions from 6 monitoring stations (Pacific Ocean stations 5°, 10°, 15°, 20°, 25°, and Barrow, AK) to build a time-varying interpolated boundary at 145° W longitude from the equator to the North Pole. Each particle in the ensemble is assigned a boundary condition mixing ratio based on its ending latitude (projected onto the 145° W boundary). Trajectories that end north of 60° latitude and east of $-120°$ longitude are assigned mixing ratios from a northern boundary condition, constructed from measurements at Alert, Canada, and Summit, Greenland, stations. The estimated boundary condition value assigned to each trajectory is averaged over the entire ensemble to produce the final modeled boundary. This latitudinally, zonally, and temporally varying boundary condition allows the modeled background concentration to change when changing synoptic conditions bring air from different regions of the globe. In contrast to an empirical boundary, we could have used the output from a global chemical tracer model (CTM) for the boundary condition. However, known shortfalls in existing inventories and uncertainties in stratosphere exchange over longer time scales results in CTM N_2O estimates that poorly match measurements (Kort et al., 2011).

Data are insufficient to create a boundary condition for the southern edge of the model domain. Weekly N_2O measurements at Barbados in fall 2008 suggest that the boundary concentration estimated for the WKT tower using our Pacific boundary condition could be too low by up to $\sim$0.2 ppb. A low estimated boundary at WKT would imply a higher N_2O increment from continental sources. As a result, the flux estimates presented here could consequently be biased slightly high over the southern US, likely by no more than 25%.

The STILT model does not explicitly account for dilution of N_2O due to the influence of stratospheric air. Stratosphere-to-troposphere exchange reaches a maximum during fall through spring and occurs at a much smaller magnitude during summer (Holton et al., 1995; Sprenger & Wernli, 2003). This exchange could influence concentrations at the receptor sites in two very general ways: through broad-scale exchange that becomes well-mixed in the troposphere and through specific "deep exchange" (stratospheric intrusion) events. Broad-scale exchange that becomes well-mixed should be captured by the measurement-based boundary condition. It would be very difficult to account for deep exchange events, but several indicators might downplay the importance of these events on eventual surface source constraints. First, deep exchange events reach a maximum over the Gulf of Alaska and southern Greenland, areas far from anticipated N_2O sources (Sprenger & Wernli, 2003). Secondly, any depletion due to deep exchange would result in low measurements and modeled concentrations that are too high. This is the opposite of what we see in the actual model (see Section 1.4.4). Finally, the measurements rarely drop below the adjusted modeled boundary condition, suggesting a lack of strong stratospheric influence at the surface sites (e.x. – Figs. 1.4, 1.2).

1.2.5 Data overview

We use two different types of measurement data in the analysis of N_2O emissions: continuous measurements from LEF tower in Wisconsin and daily flask measurements from an ensemble of tall tower sites (Fig. 1.3).

Continuous hourly-averaged measurements of N_2O are available only at NOAA's LEF tower site for 5 months (May – Oct.) in 2004. Despite the limited time frame, the data nonetheless provide a high resolution test to validate the STILT-N_2O model. Measurements at LEF were taken

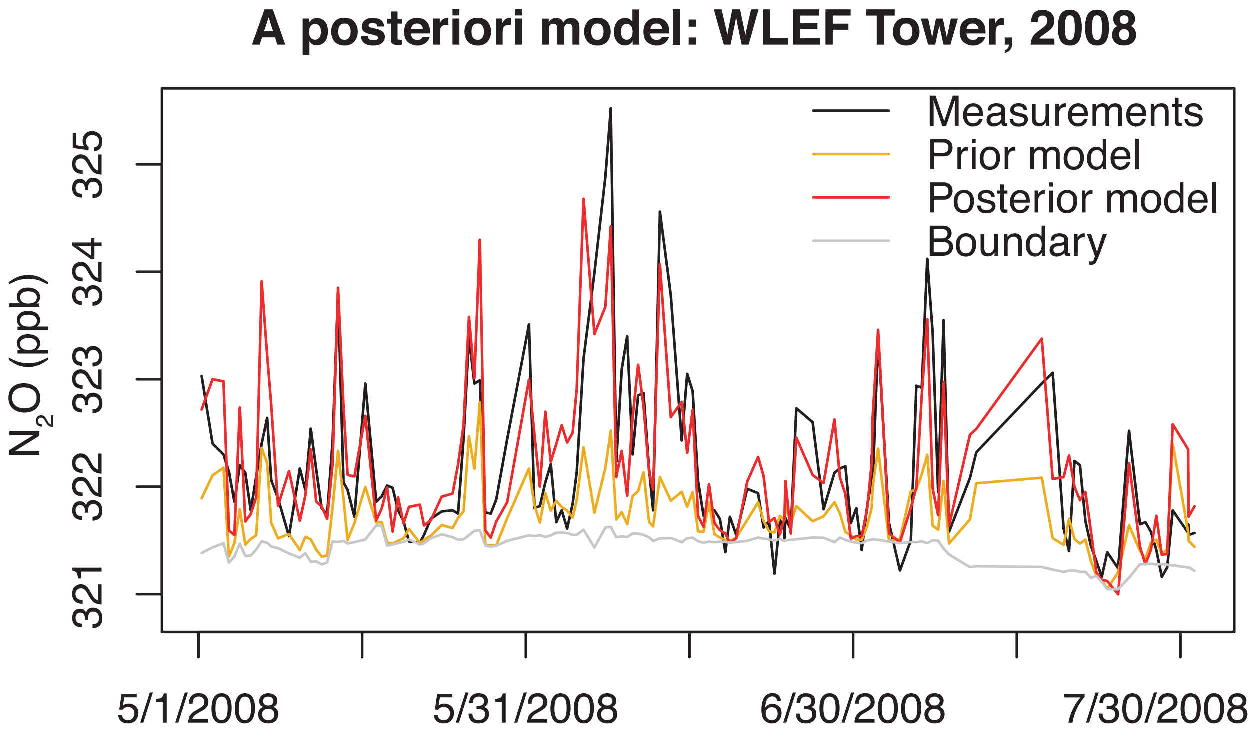


Figure 1.2: An example of the STILT model result at LEF tower with both the a priori and a posteriori EDGAR32FT2000 inventories.

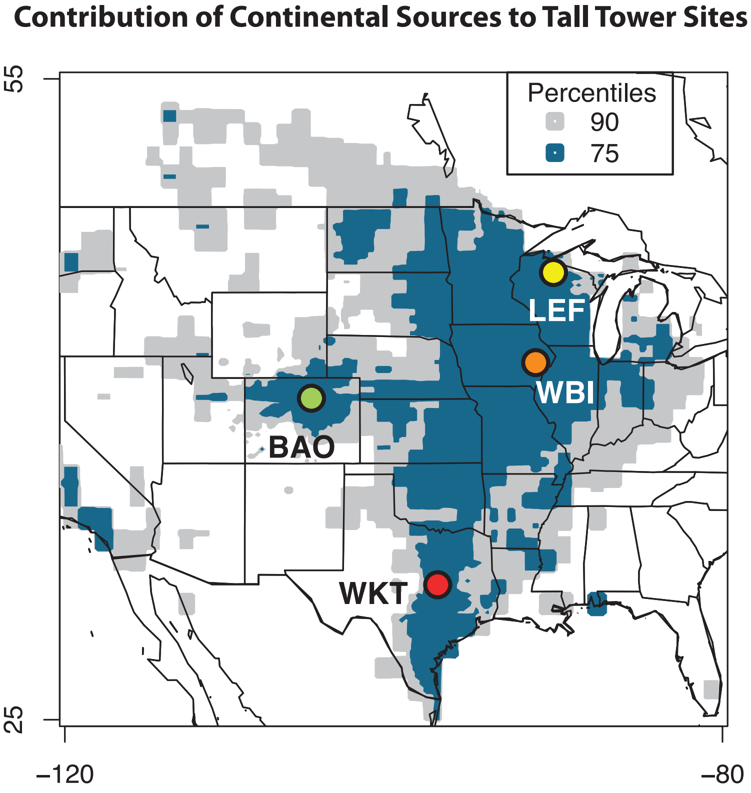


Figure 1.3: The contour lines indicate the regions that contribute 75 and 90% of the mixing ratio signal seen at each of the tall tower sites. The plot reflects the months of May through August, 2008, and was constructed using the EDGAR32FT2000 emissions inventory. The tall tower ensemble sees influence from N_2O sources over much of the central US.

using a gas chromatograph (Hurst et al., 1997, 1998) and had an approximate uncertainty of 0.6 ppb. In order to mitigate significant noise in the hourly-averaged data, we apply a 4th order, 22-point Savitsky-Golay moving average filter (Savitzky & Golay, 1964). The filter effectively preserves the magnitude of peaks and troughs on synoptic time scales in the time series and substantially decreases noise.

In addition to continuous tall tower measurements, we also draw on a network of daily flask measurements from 4 tall tower sites during 2008. These sites include LEF in Park Falls, Wisconsin (244 m above ground level); WBI in West Branch, Iowa (379 m agl); WKT in Moody, Texas (457 m agl); and BAO in Erie, Colorado (300 m agl). The daily flask dataset begins in 2008 with limited data for 2007. We conduct model simulations for the tall tower ensemble over 2008 when both data and meteorological simulations are available. All flask samples were measured using a gas chromatograph and electron capture detector and had a reproducibility of approximately 0.3 ppb. For both hourly and daily datasets, the measurement uncertainty is $30 - 60\%$ of the typical signal from continental surface sources, making it particularly challenging to derive regional source constraints.

1.3 Emissions Estimation Methods

1.3.1 Simple inventory optimization

We first discuss data analysis using a simple optimization of each emissions inventory (i.e. – GEIA, EDGAR, and DLEM). We plot modeled mixing ratios against measurements and fit a linear regression. Variability in the modeled and/or measured signals should result from the transport of varying sources from different regions. Hence, the regression slope provides an approximate corrective scaling factor for each emissions inventory based on the array of model results and measurement sites. In addition, the intercept of the regression can be interpreted as an additive correction to the modeled boundary condition. We use a reduced major axis (RMA) method (Miller & Kahn, 1962) to produce a single scaling factor for each inventory (calculated over all available towers) for each month in 2008 and for each month of the available 2004 LEF dataset (using WRF and BRAMS winds separately). An RMA regression accounts for uncorrelated variance in both the x and y axes. The surface source signal over the central US is dominated by fluxes from agriculture (globally $\sim$4x the fossil fuel source) (Denman et al., 2007); agricultural and urban sources are also co-located in some regions, and therefore we cannot reliably decouple different source types.

The RMA regression provides the most straightforward snapshot of seasonality in N_2O emissions. Significant gaps in the data occur at several towers in 2008, specifically in the winter, spring, and fall months. As such, it is difficult to produce a reliable spatially-variant inversion for time periods other than summer of 2008. Therefore, we use the regression method to estimate month-by-month constraints over all of the tower sites.

We apply a boundary correction to each monthly inversion simulation based on the regression intercept. The tall tower data that most strongly constrains the boundary correction are nearly disjoint from data points that constrain the sources, so this procedure has little effect on our main results. If we do not apply a boundary condition correction, the inversions produce obvious bias in the model-data comparisons and unrealistic spatial distributions of the flux fields (e.g. – large emissions over the sea or distant regions).

1.3.2 Geostatistical and Bayesian inversions

We use both a geostatistical inversion (Snodgrass & Kitanidis, 1997; Michalak et al., 2004) and the more conventional Bayesian approach (Rodgers, 2000). The Bayesian inversion incorporates an a priori emissions inventory, along with the measurements, model results, and uncertainty estimates. The resulting emissions inventory optimization (the a posteriori solution) reflects both the a priori estimate and the model/measurements (Rodgers, 2000). In many applications, the a priori emissions estimate provides important, independent information for the inversion, and the inversion improves upon this initial best guess. In other applications, it can be useful to estimate trace gas emissions more directly from the atmospheric data without the help of an a priori estimate (i.e. – making as few initial assumptions as possible). The geostatistical approach is an inversion that can do just that by using a non-informative prior. A geostatistical inversion can be most useful when a priori inventories differ substantially in either magnitude or spatial structure, as is the case for N_2O, indicating significant errors in the spatial distribution of emissions. In this case, the a posteriori solution will vary arbitrarily depending on the highly subjective choice of the prior. It makes sense to learn as much as possible about N_2O fluxes before convolving the results with the influence of a potentially-erroneous prior. When different priors give very disparate flux estimates, the geostatistical inversion can inform the choice of prior for a Bayesian inversion or help to describe the spatial distribution of emissions using information directly from atmospheric data.

In this study, the geostatistical approach helps us examine spatial patterns in fluxes independent of disparate a priori inventories (i.e. – DLEM versus EDGAR32FT2000). We conducted separate geostatistical inversions for each month during the peak period of fluxes (May – July 2008) on a resolution of 1° lat. x 1° lon. to minimize spatial aggregation errors. We focus on the summer time period because the flask dataset was much more sporadic at other times of year. Section 1.3.3 provides a more mathematically rigorous description of the geostatistical approach.

We also conducted comparable monthly Bayesian inversions during the same time period (May – July 2008). We use the geostatistical method to help inform our choice of a priori inventory and then use the Bayesian method to improve on this best initial inventory estimate.

1.3.3 Inversion Mathematics

The STILT trace gas model can be written in linear form as follows (Gerbig et al., 2003):

$$z = \mathbf{H}s + b + \varepsilon \tag{1.1}$$

z is an $n \times 1$ vector of measurements at the receptor location(s) (in this case, from a tall tower), s is an $m \times 1$ vector of fluxes over the entire model domain, and b is an $n \times 1$ vector that represents the model boundary condition. $\mathbf{H}$, the Jacobian matrix (dimensions $n \times m$), relates the surface fluxes to the continental source signal as measured at the tower. Each column of $\mathbf{H}$ describes the statistical influence of fluxes in a certain grid box on the tall tower site (also known as the "influence footprint"). In other words, the influence footprint converts the emissions inventory from a flux to a mixing ratio increment seen at the receptor. This footprint is calculated based on the number of particles that pass over the grid box and the amount of time the particles spend in the box. $\mathbf{H}s$ ($n \times 1$) represents the contribution of continental sources to the mixing ratio at the receptor (in ppb). ε is the model-data mismatch: error caused by model transport, model resolution, uncertainties in the measurements, and boundary condition uncertainties (e.g. – Gerbig et al. (2003), Matross et al. (2006), or Gourdji et al. (2010) for the STILT model).

The solution to the inverse problem minimizes the chi-squared cost function, somewhat similar to a weighted sum of squares approach. Equations 1.2 and 1.3 give the cost function for the Bayesian and geostatistical inversions, respectively.

$$J(s) = \frac{1}{2}(z - \mathbf{H}s)^T \mathbf{R}^{-1}(z - \mathbf{H}s) + \frac{1}{2}(s - s_a)^T \mathbf{Q}^{-1}(s - s_a) \tag{1.2}$$

$$J(s,\beta) = \frac{1}{2}(z - \mathbf{H}s)^T \mathbf{R}^{-1}(z - \mathbf{H}s) + \frac{1}{2}(s - \mathbf{X}\beta)^T \mathbf{Q}^{-1}(s - \mathbf{X}\beta) \tag{1.3}$$

R is the $n \times n$ model-data mismatch covariance matrix, and **Q** is the $m \times m$ a priori covariance matrix.

The two inversion approaches differ in their treatment of the a priori model. The Bayesian method determines the solution with the maximum probability given the prior estimate and the observations and their range of uncertainties, defined by **R** and **Q**, respectively (Rodgers, 2000). The geostatistical method does not include an a priori inventory (s_a) but instead uses a model of the mean ($\mathbf{X}\beta$). Each column of the matrix **X** ($m \times p$) can describe possible spatial trends in the data. β is a $p \times 1$ vector of unknown drift coefficients that scale the columns of **X**. This model of the mean could, for example, include information such as land use patterns or population density (i.e. – "auxiliary variables"; see Gourdji et al. (2008)). In this study, however, we use do not use external information in the model of the mean. In other words **X** is an $m \times 1$ vector of ones, an uninformative prior. In this case, the unknown value of β is the a posteriori mean of the flux field. We want to learn as much as possible about the fluxes from the atmospheric data alone without the aid of any prescribed spatial distribution. Instead, the geostatistical method incorporates broad-scale information within the covariance matrices about the spatial correlation structure of the fluxes. The diagonal elements of **Q** define the spatial variance of the fluxes, and the off-diagonal elements of **Q** describe the anticipated spatial covariance of the fluxes (e.g. – Snodgrass & Kitanidis, 1997; Michalak et al., 2004; Mueller et al., 2008).

For both inversions, we include the 90% influence region (see Section 1.4.1) in the inversion area as well as all land area further west. The data do not fully constrain emissions as far away as the West Coast, but we included the West to avoid biasing the a posteriori flux field due to the cumulative effect of distant emissions. We excluded water bodies from the inversion. Globally, the ocean is a significant source of N_2O, diffusely spread over large areas (Nevison et al., 2004; Denman et al., 2007). However, any fluxes assigned to the ocean by the inversion would more likely reflect boundary condition uncertainties than actual sources and our boundary condition correction should account for the effects of any broad-scale coastal upwelling fluxes.

The analytical solutions to both Bayesian and geostatistical inversions are calculated via a system of linear equations. The solution to the Bayesian approach for the STILT model is described in Gerbig et al. (2003) and Miller et al. (2008), and the solution to the geostatistical inversion is described in detail by Snodgrass & Kitanidis (1997) and Michalak et al. (2004).

1.3.4 Covariance matrices

The covariance matrices (**Q** and **R**) provide crucial information that describes the structure of the a posteriori flux field and/or the relative uncertainties in the inversion inputs. These matrices must be estimated independent of the inversion and are often inferred from the atmospheric data itself (Kitanidis & Lane, 1985; Kitanidis, 1995; Michalak et al., 2005). Mathematically, the individual elements of **R** and **Q** are defined as follows:

$$R_{i,j}(t_{i,j}|\sigma_R, t_R) = \sigma_R^2 exp(-\frac{t_{i,j}}{t_R}) \tag{1.4}$$

$$Q_{i,j}(h_{i,j}|\sigma_Q, l) = \sigma_Q^2 exp(-\frac{h_{i,j}}{l}) \tag{1.5}$$

σ_R^2, the diagonal elements of **R**, represent the variance of the errors caused by model-data mismatch. This includes uncertainties due to the measurements, the forward model, model representation (ex. – aggregation of discreet flux regions), and uncertainties in the boundary condition. In this case, we estimate a single model-data mismatch error over all towers for each month. The off-diagonal elements are the covariances which decay exponentially in time ($t_{i,j}$) according to the temporal decorrelation in model-data mismatch errors (t_R). We constructed a first order autoregressive model on the difference between model and measurements and found a decorrelation time of 4-10 hrs. for hourly data at LEF with the WRF model (depending on the data interpolation method used). This is smaller than the daily measurement frequency in the ensemble data, so we do not include temporal correlations in model-data mismatch (off-diagonal elements of $\mathbf{R}$). We also assume zero error correlation among observations at different tower sites.

The a priori covariance matrix takes on a similar form. For the geostatistical inversion, the diagonal elements (σ_Q^2) give the anticipated variance of the fluxes, and for the Bayesian inversion, they describe the uncertainty in the a priori inventory. The off-diagonal elements of $\mathbf{Q}$ decay exponentially with distance ($h_{i,j}$) according to the decorrelation length parameter (l). In the case of the geostatistical inversion, l describes spatial correlation in the flux field while for the Bayesian inversion, l describes the spatial correlation of uncertainties in the a priori inventory. Spatial correlation becomes negligible at distance $3l$.

We infer most of the covariance matrix parameters from the atmospheric data itself using Maximum Likelihood (ML) and Restricted Maximum Likelihood (RML) methods. In this case, we apply the RML method to infer geostatistical inversion parameters (Kitanidis, 1995; Snodgrass & Kitanidis, 1997) and the ML method to infer Bayesian inversion parameters (Kitanidis & Lane, 1985; Michalak et al., 2005). These closely related techniques provide objective ways to infer any number of inversion parameters. The most likely parameters are those that minimize a given cost function and are normally estimated using an iterative Gauss-Newton algorithm. ML and RML ensure that the reduced chi-squared value of both the model/measurements and fluxes are close to one. This implies that the errors in the a posteriori model are consistent with those estimated by the covariance parameters (e.g. – Michalak et al., 2005). It is important to note, however, that ML and RML can produce unreliable inversion parameters when either the data are sparse or the model/measurement errors are large (Mueller et al., 2008; Gourdji et al., 2010).

This latter concern becomes important in the inversion setup here. In cases where we can-

Table 1.1: Covariance matrix parameters for the geostatistical and Bayesian inversions.

	σ_{R_MAY}	σ_{R_JUN}	σ_{R_JUL}	l (km)	σ_Q
Geostatistical	0.71 ± 0.05	0.97 ± 0.05	0.70 ± 0.04	816 ± 313	$(4.5 \pm 3) \times 10^{-4}$
Bayesian	0.72 ± 0.05	0.99 ± 0.05	0.73 ± 0.04	816 ± 313	$(3.7 \pm 2) \times 10^{-4}$

not infer inversion parameters from the atmospheric data, we infer these values from auxiliary datasets. For example, we assume that the posterior fluxes and errors in the a priori inventory will be correlated over a length scale (l) similar to that of agricultural land use patterns (inferred from cropland maps using RML in kriging form) (Mueller et al., 2008; Ramankutty et al., 2008). Existing emissions inventories of N_2O, corrected in magnitude in accordance with Kort et al. (2008), provide an estimate of σ_Q for the geostatistical inversion (again using RML in kriging form). Additionally, we use the difference between the standard and corrected inventory from Kort et al. (2008) for σ_Q in the Bayesian setup. Table 1.1 lists the full set of parameters used in the geostatistical and Bayesian setups.

1.4 Results and Discussion

1.4.1 Model results – an overview

The STILT model appears very adept at reproducing high resolution N_2O concentration data for a range of different receptor sites. The hourly data from LEF Tower in 2004 demonstrate the model's ability to reproduce high resolution measurements (see Fig. 1.4). In this case, we ran the model using both the WRF and BRAMS meteorological drivers. STILT achieves model-measurement correlation as high as $R = 0.76$ (for EDGAR32FT2000), though correlations for other inventories can be lower ($R = 0.48$ for DLEM) (see Table 1.2). The time series in Fig. 1.4

Table 1.2: Emissions inventory correction factors and model-measurement correlation coefficients for 2004 calculated using an RMA regression based on hourly LEF tower nitrous oxide measurements.

Inventory	Meteorology	Inventory correction factors						R
		May	June	July	August	Sept.	Oct.	
EDGAR32FT2000	BRAMS	2.5 ± 0.2	2.6 ± 0.1	1.5 ± 0.1	1.5 ± 0.1	1.3 ± 0.1	1.4 ± 0.1	0.76
	WRF	3.7 ± 0.2	3.4 ± 0.2	1.6 ± 0.2	1.9 ± 0.2	1.2 ± 0.1	1.4 ± 0.1	0.73
EDGARv4	BRAMS	4.4 ± 0.3	4.8 ± 0.3	2.6 ± 0.2	2.7 ± 0.2	2.3 ± 0.2	2.5 ± 0.2	0.76
	WRF	6.2 ± 0.3	6.1 ± 0.4	2.7 ± 0.2	3.5 ± 0.3	2.2 ± 0.2	2.4 ± 0.2	0.74
GEIA	BRAMS	3.7 ± 0.2	4.1 ± 0.2	2.5 ± 0.2	2.3 ± 0.2	2.0 ± 0.2	1.9 ± 0.1	0.76
	WRF	5.1 ± 0.3	5.5 ± 0.4	2.4 ± 0.2	2.9 ± 0.3	1.8 ± 0.2	2.0 ± 0.2	0.73
DLEM	BRAMS	1.7 ± 0.1	1.1 ± 0.1	0.90 ± 0.1	0.66 ± 0.06	0.50±0.05	1.3 ± 0.1	0.52
	WRF	2.2 ± 0.2	1.3 ± 0.1	0.85 ± 0.06	0.94 ± 0.09	0.63±0.06	1.1 ± 0.1	0.48

displays the model result for both the BRAMS and WRF-driven models after optimization.

Before fully discussing the source optimization results, it is first important to highlight the differences in STILT model runs with BRAMS versus WRF meteorological drivers. In general, the BRAMS runs produce a slightly better model-measurement fit than WRF ($R = 0.76$ and 0.73, respectively, for EDGAR32FT2000), though this difference in fit is only significant in the cases of EDGAR32 and GEIA ($p = 0.01$ in both cases). Secondly, when using the same flux fields, STILT-WRF runs produce modeled mixing ratios that are systematically lower than those produced by STILT-BRAMS, statistically significant for all inventories. This difference appears to stem from the fact that particle trajectories in WRF-driven simulations reach greater altitudes more quickly than in BRAMS simulations and therefore see the influence of surface sources over a shorter interval. For example, in Fig. 1.5, the average influence footprint of WRF and BRAMS simulations is initially the same. However, the surface influence drops to lower values on days 2–4 for WRF, likely because particles move out of the planetary boundary layer (PBL) a bit faster, and thus lower mixing ratios modeled at the receptor. During the first day, and in the far field, the footprints are similar and the overall mean footprints differ by 15–20%.

WRF and BRAMS also display different nighttime mixed layer heights ($\sim$30 m for WRF and

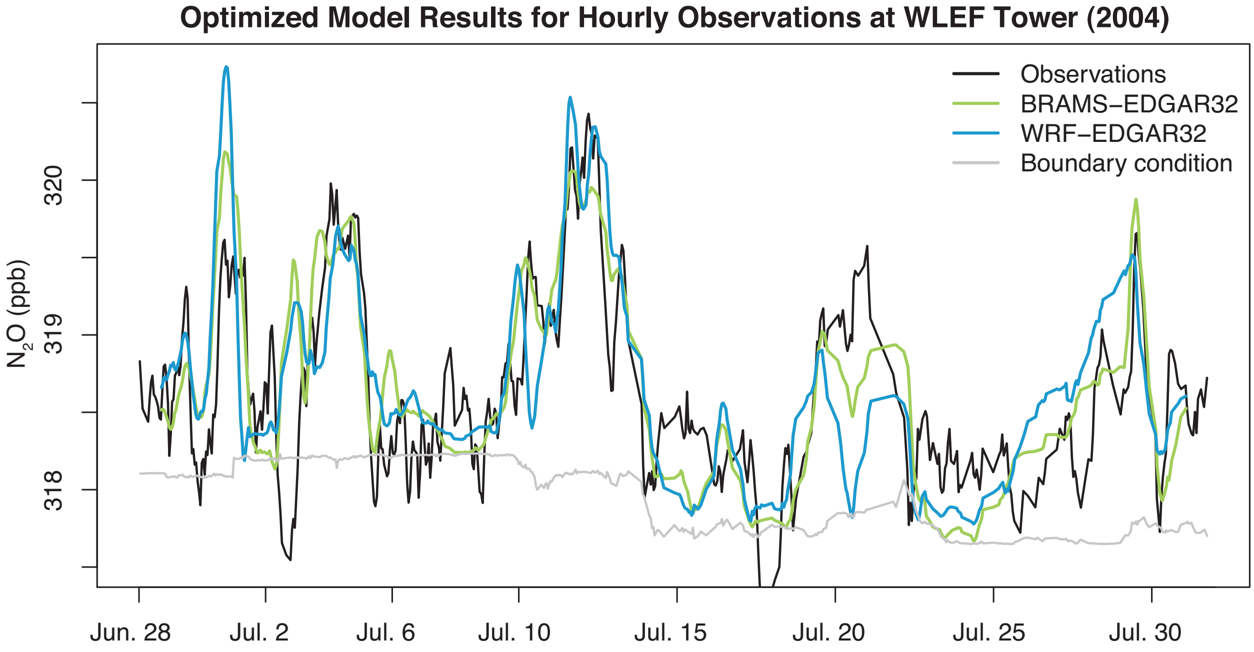


Figure 1.4: STILT model results for hourly observations at LEF Tower in Park Falls, WI, with two different meteorological drivers. Each result is optimized with a reduced major axis regression and smoothed with a Savitsky-Golay filter.

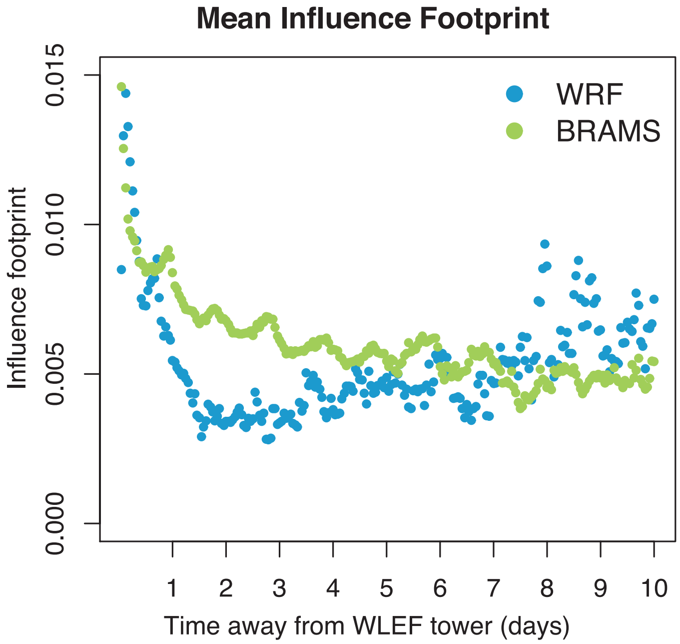


Figure 1.5: The mean footprint of particles (in μmol m^{-2} s^{-1}) by time away from the LEF tower (2004) for both BRAMS and WRF. BRAMS sees a larger footprint than WRF between 0.5 – 6 days away from the tower.

250 m for BRAMS). STILT can set systematically different nighttime mixed layer heights, and differences in estimated concentrations among model versions with different nighttime mixed layers appears to be minimal. Resolution of the discrepancy in particle heights between WRF and BRAMS is beyond the scope of this paper but would be a significant subject to improve the transport model uncertainties.

In addition to differences in meteorology, we also examined possible differences in the model during the afternoon versus nighttime at LEF in 2004. We optimized mid-afternoon (18 – 23 UTC) and nighttime (6 – 14 UTC) data points separately using the RMA regression for each month and each meteorological driver. The inventory correction factors were usually similar, within 5% for afternoon versus nighttime measurements and almost never differed by more than 15%. Evidently, agricultural sources affecting LEF are large area sources and are dominated by synoptic-scale flow.

The emissions constraints presented here apply mostly to the Great Plains corridor and western Great Lakes of the United States. Figure 1.3 shows the mean "regions of influence" for the 2008 tall tower ensemble model runs. At each geographic grid square, we multiply the footprint by the EDGAR 32FT2000 emissions inventory. The result is the mixing ratio increment that each grid cell contributes to the measured signal at the tall tower site. The contour lines provide an estimate of the geographic regions that contribute 75 and 90% of the mean trace gas signal seen at the tall tower sites. Figure 1.3 shows that the 2008 tall tower ensemble is most sensitive to fluxes across the central corridor of the United States. A similar plot for the 2004 LEF dataset (not shown) produces a region of influence that covers the corn belt and northern plains states. Hence, the model inversions that follow in this paper predominantly provide constraints for the central US.

Table 1.3: Emissions inventory correction factors for 2008 calculated using an RMA regression based on an ensemble of tall towers.

Inventory	Inventory correction factors									
	Jan./Feb.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
EDGAR32	1.1 ± 0.1	2.6 ± 0.4	2.8 ± 0.4	5.4 ± 0.6	3.5 ± 0.4	2.2 ± 0.3	1.5 ± 0.3	1.1 ± 0.2	1.1 ± 0.2	1.3 ± 0.2
EDGARv4	1.3 ± 0.2	4.1 ± 0.7	5.1 ± 0.7	10.1 ± 1.2	5.7 ± 0.7	3.8 ± 0.5	1.14 ± 0.2	1.0 ± 0.2	1.1 ± 0.2	1.8 ± 0.3
GEIA	1.9 ± 0.2	3.1 ± 0.5	3.0 ± 0.5	4.5 ± 0.7	3.9 ± 0.6	2.0 ± 0.3	2.1 ± 0.4	1.4 ± 0.2	1.2 ± 0.2	1.6 ± 0.2
DLEM	3.3 ± 0.4	2.9 ± 0.5	2.3 ± 0.4	2.1 ± 0.3	0.9 ± 0.1	0.29 ± 0.04	0.88 ± 0.18	0.83 ± 0.13	1.2 ± 0.2	4.9 ± 0.7

1.4.2 Seasonality of N_2O Emissions

We construct a reduced major axis regression for the 2008 daily tall tower ensemble data in order to make broad monthly corrections to existing inventories (see Table 1.3). Figure 1.6 shows the monthly emissions inventory corrections taken over the entire ensemble of all tower sites. Note that January and February are lumped together because there were only 15 days of model runs for the former month. Also, we did not compute an inventory correction for March because two of the four towers did not report any data.

The results from the entire ensemble provide important information about the seasonal cycle of N_2O over the central US in 2008. Note that since DLEM is a daily inventory, it includes estimated seasonality while GEIA and the EDGAR inventories only provide yearly emissions totals. Our study shows strongly seasonal fluxes that peak in June and fall off at similar rates on either side of the peak. The scaling factors for DLEM represent corrections to the seasonality already included in the inventory. These scalers indicate that DLEM overestimates fluxes during late summer relative to other months and slightly underestimates sources early in the year. The DLEM adjustment for December is anomalous. The inventory in December is 25% lower than in either November of January, and inventory sources during this month shift somewhat eastward, explaining the anomaly at least in part.

Modeled footprints change slightly from one month to another, and some months lack mea-

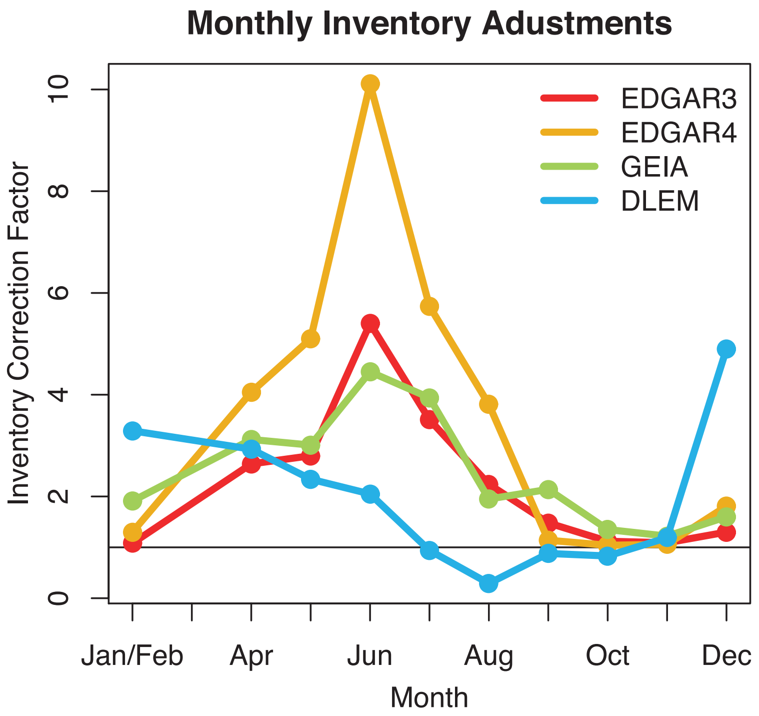


Figure 1.6: Corrective emissions inventory scaling factors produced by the RMA regression for the ensemble of 2008 tall tower data.

surements from one or more towers. Hence, scaling factors for each month could be weighted toward certain regions because of missing towers or seasonal changes in synoptic transport. As a result, the above constraints represent both seasonal changes in surface sources and any artifacts from data availability and seasonal footprint variations.

1.4.3 Location of N_2O Emissions

Different N_2O emissions inventories exhibit very different spatial patterns (see Fig. 1.1). EDGAR inventories place the largest sources over the US Cornbelt (i.e. – Iowa, Illinois, etc.) while DLEM places high emissions in more westerly (ex. – Kansas and Nebraska) and southerly areas (ex. – Texas, Arkansas, Kansas, etc.). The geostatistical inversion provides a very transparent way to learn about the spatial distribution of the fluxes based on atmospheric data independent of existing emissions inventories. We conducted the geostatistical inversion for three different months (May - July, 2008) using STILT-WRF.

Figure 1.7 shows the a posteriori flux fields for May, June, and July, along with associated a posteriori uncertainties as estimated by the geostatistical inversion. We conducted the inversion over a large region that includes most of the continental US and Canada but show results only for the area within the 75% influence region of the tower sites (see Fig. 1.7). The results of the inversion strongly indicate large N_2O sources over the US Corn Belt (i.e. – Iowa, Illinois, etc.). This source region continues to the northwest into North Dakota with a diminished magnitude, and emissions taper off quickly moving westward into Colorado, Wyoming, and the western Plains States.

With limited atmospheric data, the Bayesian inversion cannot fully correct discrepancies in spatial distribution among inventories. In this case, the geostatistical inversion can inform our choice of a priori inventory. Based both on the geostatistical simulations and the 2004 measurement-model correlations, we select EDGAR32FT2000 as the best a priori inventory for the Bayesian

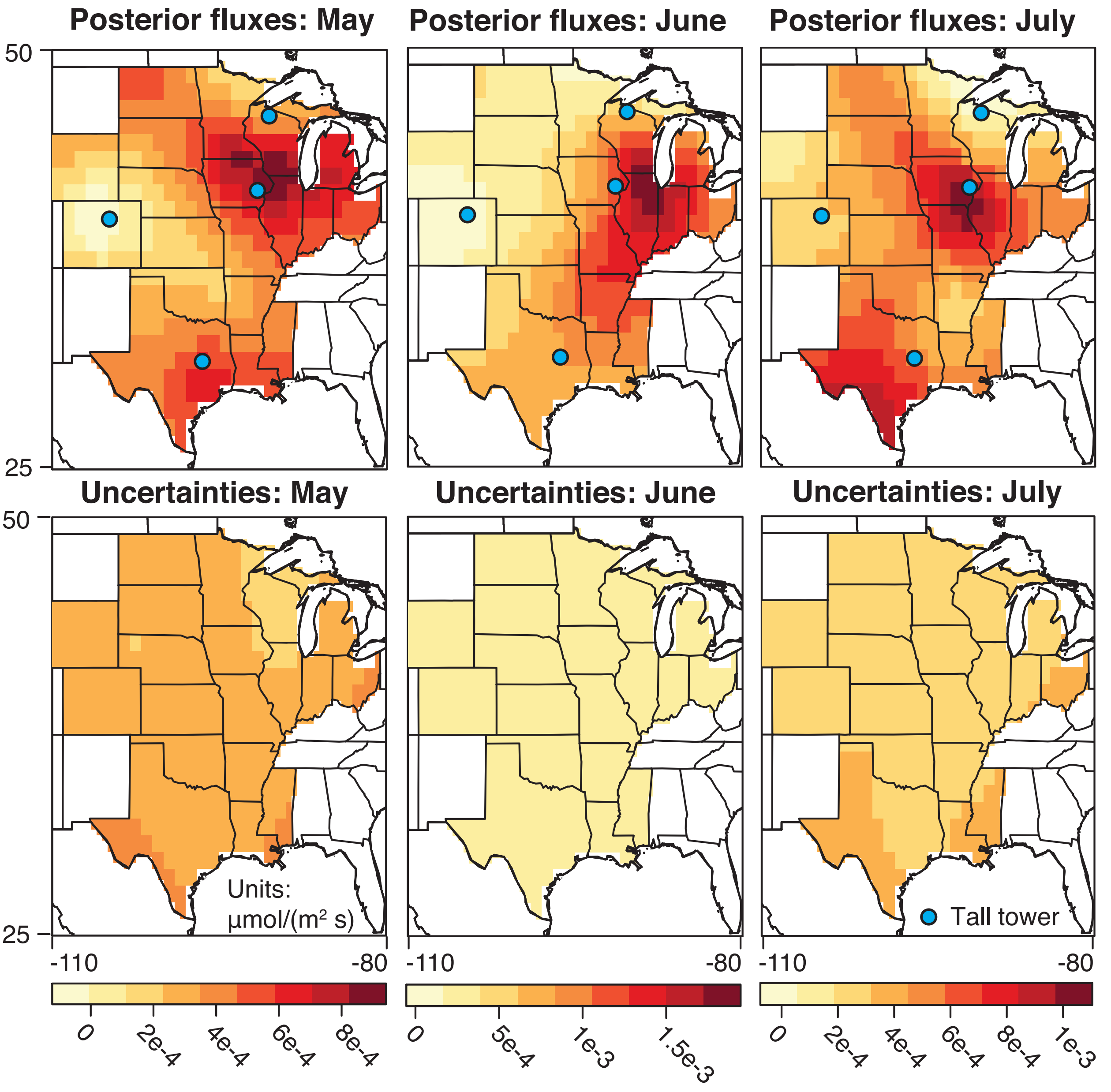


Figure 1.7: Monthly a posteriori fluxes and associated uncertainties (a posteriori standard deviations) estimated for the geostatistical inversion. This figure is plotting using different colors from Fig. 1.1 and on different scales to better highlight spatial trends.

Table 1.4: N_2O budget information for 2004 and 2008 datasets for the inversion region shown in Fig. 1.7 (in TgN/month). Note: The ranges in 2004 and 2008 RMA reflect different inventories and meteorological drivers while 2008 geostatistical- and Bayesian-derived budgets include a posteriori uncertainties calculated by the respective inversions.

	May	June	July	Aug.	Sept.	Oct.
LEF Tower 2004 (RMA)	0.12 – 0.18	0.12 – 0.19	0.073 – .089	0.073 – 0.11	0.056 – 0.068	0.065 – 0.075
Ensemble 2008 (RMA)	0.11 – 0.15	0.16 – 0.29	0.14 – 0.17	0.070 – 0.11	0.032 – 0.074	0.031 – 0.054
Ensemble 2008 (Geostatistical)	0.11 ± 0.08	0.18 ± 0.08	0.13 ± 0.07			
Ensemble 2008 (Bayesian)	0.11 ± 0.07	0.18 ± 0.06	0.13 ± 0.06			

inversion. The EDGAR inventories and GEIA place major sources over the Corn Belt, in agreement with the geostatistical simulations, while DLEM places the largest sources elsewhere. The EDGAR and GEIA inventories also produce the highest model-measurement fit (R^2), and EDGAR32FT2000, in specific, requires less magnitude correction than either GEIA or EDGAR v. 4.

1.4.4 Magnitude of N_2O Emissions

The following section details the results of the Bayesian inversion to improve upon existing N_2O emissions inventories and details estimated emissions budgets from the variety of statistical methods. While the geostatistical setup aimed to discern large scale spatial patterns, the Bayesian setup described here leverages the best existing inventory to produce a more spatially-resolved estimate of fluxes. The inversion was conducted for May, June, and July, the months with peak N_2O fluxes and with nearly complete tall tower data sets. Figure 1.8 plots the EDGAR32FT2000 a posteriori inventory from the Bayesian inversion, and Fig. 1.2 shows an example of the modeled time series from the LEF tower. The results look somewhat similar to the geostatistical setup but are less dispersive and have more spatial detail.

The inversions and regressions suggest sources that are significantly larger than in either EDGAR or GEIA for nearly all geographic regions and times of year. Interestingly, estimated emissions in the newest release of EDGAR (v. 4) are lower than previous releases, requiring even

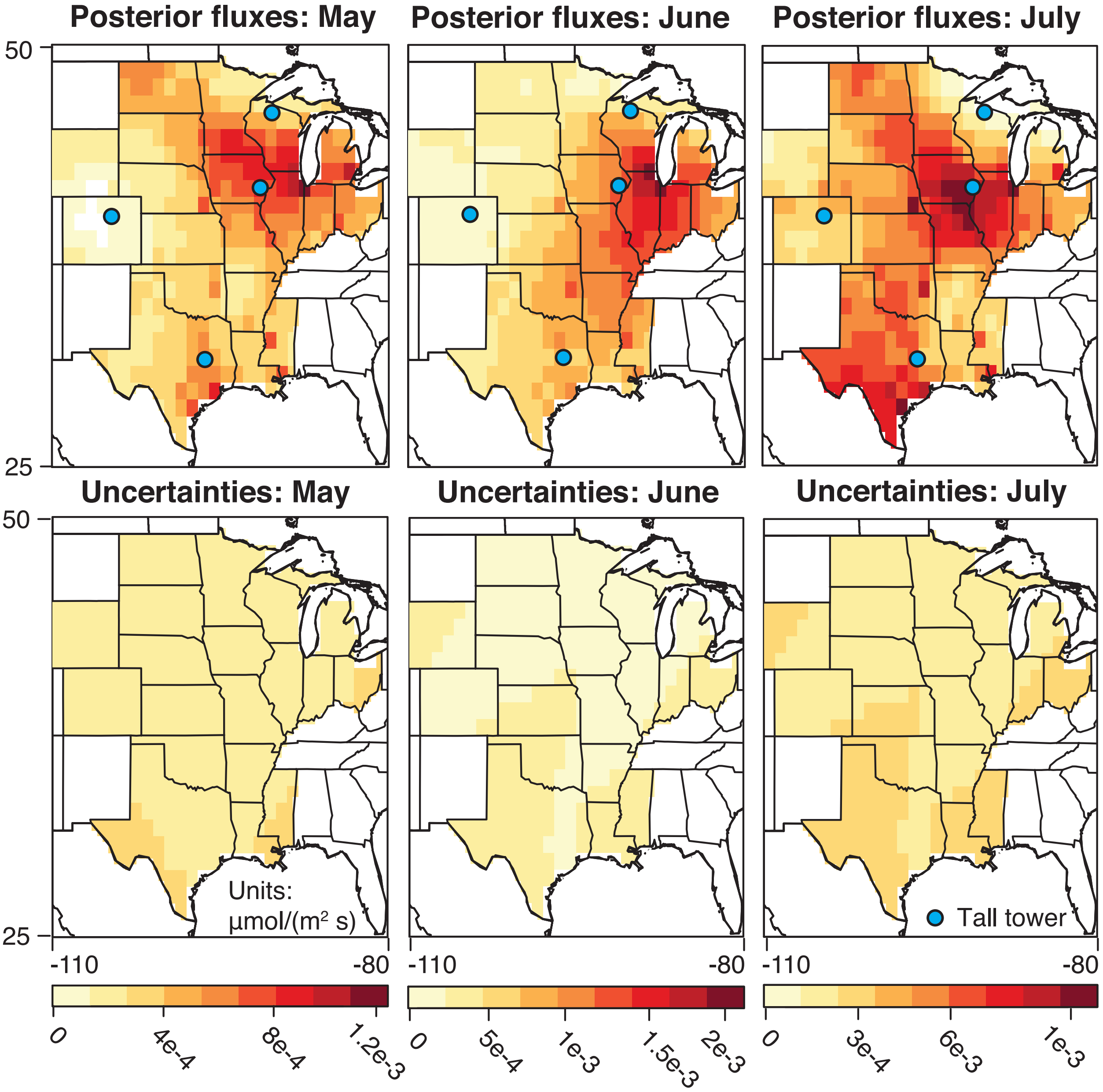


Figure 1.8: The a posteriori source estimates for N_2O using the EDGAR32FT2000 inventory in a Bayesian inversion framework (for May – July 2008). Each month is plotted on a different scale to better highlight spatial patterns.

Table 1.5: The percentage of a priori N_2O fluxes in the US and Canada that lie within the inversion region as displayed in Fig. 1.7.

EDGAR32	EDGAR v.4	GEIA	DLEM
51%	48%	40%	39%

larger modification. Table 1.4 displays monthly N_2O budgets from the inversions and RMA regressions for both 2004 and 2008. Using the RMA regression for the 2008 ensemble of tall tower sites, we estimate an annual N_2O budget of 1.0 – 1.2 TgN/yr for the inversion region. If we divide these results by the percentage of fluxes in the inversion region (see Table 1.5), we can extrapolate to the entire US and Canada. The resulting budget estimate for the United States and Canada is 2.4 – 2.6 TgN/yr. This range of estimates reflects the results of different starting inventories (the EDGAR and GEIA inventories). The different methods produce slightly different budget estimates, in some part due to uncertainties in the boundary condition. The inversions, for example, apply a single boundary condition correction for the entire summer period whereas the regressions calculate a unique intercept for each month.

Differences in the meteorological drivers modestly widen the uncertainty bounds on the budget estimate. Optimal fluxes on average were 12% lower when using BRAMS. The mean scaling factor for EDGAR 2000 over the year is 2.2 for WRF and 1.8 for BRAMS. If we extrapolate to the 2008 dataset, this result suggests that BRAMS would yield a budget of $\sim$1 TgN/yr over the central US and $\sim$2.2 TgN/yr over the US and Canada.

The budget numbers presented here compare well to those estimated in previous top-down studies over North America, notably Kort et al. (2008, 2010). These studies estimated a US and Canada N_2O budget of 0.23 TgN/month for spring and summer months in 2003 and 2004. This compares to a peak June US and Canada budget of $\sim$0.22 – 0.35 TgN/yr presented here across a variety of different source optimization methods. Our annual budget number for the US and Canada corresponds to 13 – 15% of the total global source, or 36 – 39% of the global anthropogenic source (12 –13% and 32 – 34% using BRAMS, respectively) as reported by the IPCC (Denman et al., 2007).

1.5 Conclusions

This study captures nitrous oxide sources over the agricultural belt of the United States at more highly resolved spatial and temporal scales than previous estimates. A diverse set of statistical tools, from a simple regression to Bayesian and geostatistical inversions, provide a closer look at three different aspects of emissions: seasonality, location, and intensity. The largest sources appear over the US Cornbelt (i.e. – Iowa, Illinois, Indiana, southern Minnesota, etc.) with smaller sources likely extending into the Dakotas. This result is consistent with the spatial patterns in the EDGAR and GEIA inventories. N_2O emissions peak in June and taper off quickly both before and after, a seasonality largely missing from existing inventories. The choice of meteorology for the transport model had a significant influence (12 ± 6 %) on the estimated total emissions, with relatively small effect on our derived spatial distribution and little impact on the derived seasonal variation. We estimate a budget over the central US of 1.0 – 1.2 TgN/yr using WRF meteorology (0.9 – 1.1 using BRAMS), a source that is notably larger than the EDGAR and GEIA inventories. Of particular note, the newest release of EDGAR (v. 4), decreases the source strength over previous EDGAR estimates, a revision that is inconsistent with the atmospheric data.

The sources inferred from the tall tower measurements show striking similarity to the anticipated spatial and temporal distributions of fertilizer application and corn production, supporting the view that fertilizer plays a dominant role in generating N_2O emissions over the central US (see Fig. 1.9). Both fertilizer use and N_2O emissions correspond strongly with the distribution of corn production (not shown), likely due to the magnitude of corn production and high fertilizer use relative to other crops (USDA National Agricultural Statistics Service, 2012;

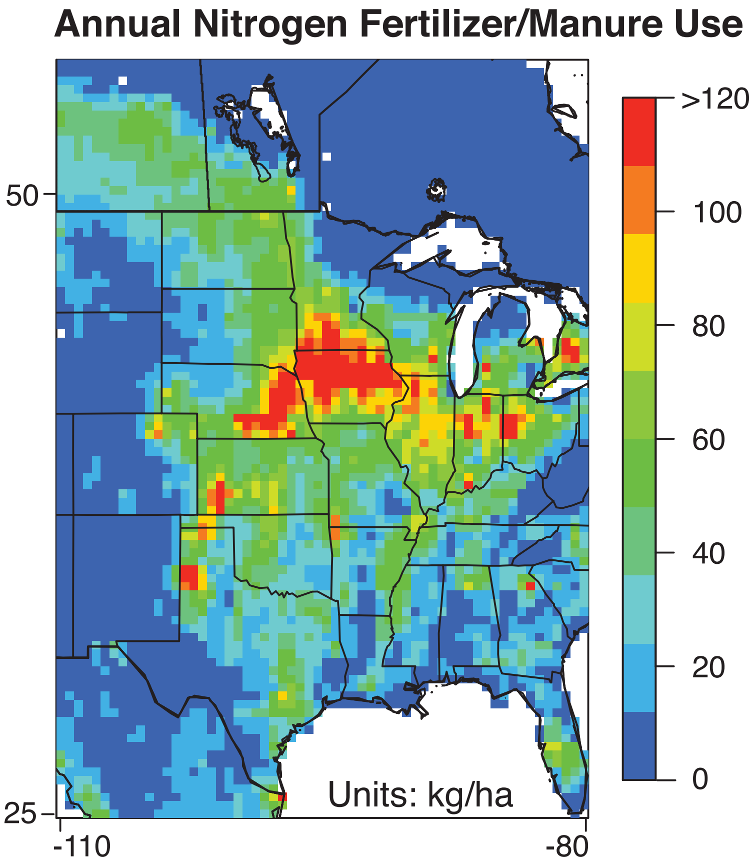


Figure 1.9: An estimate of annual fertilizer use taken from Potter et al. (2010). The a posteriori N_2O fluxes from both the geostatistical and Bayesian inversions are strongly similar to the spatial distribution of nitrogen fertilizer. Potter et al. (2010) estimate manure application for 2007 and synthetic fertilizer for 2000.

Food and Agriculture Organization of the United Nations, Plant Production Protection Division, 2012). Given this correspondence between corn and N_2O, the potential for significant agricultural greenhouse gas emissions are important to consider when weighing the costs and benefits of corn subsidies and ethanol production.

2

Anthropogenic emissions of methane in the United States

This chapter originally appeared as a research article in the Proceedings of the National Academy of Sciences. For the original article, refer to Miller et al. (2013).

This study quantitatively estimates the spatial distribution of anthropogenic methane sources in the United States by combining comprehensive atmospheric methane observations, extensive spatial datasets, and a high-resolution atmospheric transport model. Results show that current inventories from US EPA and EDGAR underestimate methane emissions nationally by a factor of ∼1.5 and ∼1.7, respectively. Our study indicates that emissions due to ruminants and manure are up to twice the magnitude of existing inventories. In addition, the discrepancy in methane source estimates is particularly pronounced in the south-central US, where we find total emissions are ∼2.7 times greater than in most inventories and account for 24% ± 3% of national emissions. The spatial patterns of our emission fluxes and observed methane-propane correlations indicate that fossil fuel extraction and refining are major contributors (45 ± 13%) in the south-central US. This suggests that regional methane emissions due to fossil fuel extraction and processing could be 4.9 ± 2.6 times larger than in EDGAR, the most comprehensive global methane inventory. These results cast doubt on US EPA's recent decision to downscale its estimate of national natural gas emissions by 25–30%. Overall, we conclude that methane emissions associated with both the animal husbandry and fossil fuel industries have larger greenhouse gas impacts than indicated by existing inventories.

2.1 Introduction

Methane (CH_4) is the second most important anthropogenic greenhouse gas, with approximately one third the total radiative forcing of carbon dioxide (Butler, 2014). CH_4 also enhances the formation of surface ozone in populated areas, and thus higher global concentrations of CH_4 may significantly increase ground-level ozone in the Northern Hemisphere (Fiore et al., 2002). Furthermore, methane affects the ability of the atmosphere to oxidize other pollutants and plays a role in water formation within the stratosphere (Jacob, 1999).

Atmospheric concentrations of CH_4 (∼1800 ppb) are currently much higher than pre-industrial levels (∼680–715 ppb) (Butler, 2014; Mitchell et al., 2011). The global atmospheric burden started to rise rapidly in the 18th century and paused in the 1990s. Methane levels began to increase again more recently, potentially from a combination of increased anthropogenic and/or tropical wetland emissions (Dlugokencky et al., 2009; Sussmann et al., 2012; Kirschke et al., 2013). Debate continues, however, over the causes behind these recent trends (Wang et al., 2004; Kirschke et al., 2013).

Anthropogenic emissions account for 50 – 65% of the global CH_4 budget of ∼395–427 TgC yr^{-1} (526–569 Tg CH_4) (Ciais et al., 2013; Kirschke et al., 2013). The US EPA estimates the principal anthropogenic sources in the United States to be (in order of importance) 1. livestock (enteric fermentation and manure management), 2. natural gas production and distribution, 3. landfills,

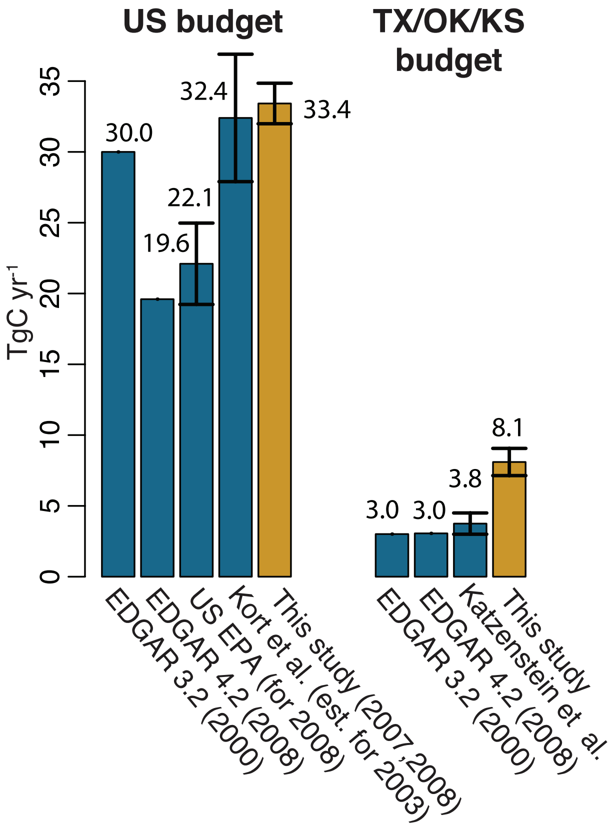


Figure 2.1: US anthropogenic methane budgets from this study, previous top-down estimates, and from existing emissions inventories. The south-central US includes Texas, Oklahoma, and Louisiana. US EPA only estimates national, not regional, emissions budgets. Furthermore, national budget estimates from EDGAR, EPA, and Kort et al. include Alaska and Hawaii while this study does not.

and 4. coal mining (US EPA, 2013). EPA assesses human-associated emissions in the US in 2008 at 22.1 TgC, roughly 5% of global emissions (US EPA, 2013).

The amount of anthropogenic CH_4 emissions in the US, and attributions by sector and region, are controversial (see Fig. 2.1). Bottom-up inventories from US EPA and EDGAR give totals ranging from 19.6 to 30 TgC yr^{-1} (Olivier & Peters, 2005; US EPA, 2013). The most recent EPA and EDGAR inventories report lower US anthropogenic emissions compared to previous versions (decreased by 10% and 35%, respectively) (US EPA, 2013; European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL), 2010); this change primarily reflects lower, revised emissions estimates from natural gas and coal production. However, recent analysis of CH_4 data from aircraft estimates a higher budget of 32.4 ± 4.5 TgC yr^{-1} for 2004 (Kort et al., 2008). Furthermore, atmospheric observations indicate higher emissions in natural gas production areas (Katzenstein et al., 2003; Petron et al., 2012; Karion et al., 2013); a steady 20-year increase in the number of US wells and newly-adopted horizontal drilling techniques may have further increased emissions in these regions (Howarth et al., 2011; US Energy Information Administration, 2013).

These disparities among bottom-up and top-down studies suggest much greater uncertainty in emissions than typically reported. For example, EPA cites an uncertainty of only ±13% for the for US (US EPA, 2013). Independent assessments of bottom-up inventories give error ranges of 50–100% (National Research Council, 2010; Dlugokencky et al., 2011), and values from Kort et al. are 47% ± 20% higher than EPA (Kort et al., 2008). Assessments of CH_4 sources to inform policy (e.g., regulating emissions or managing energy resources) require more accurate, verified estimates for the US.

This study estimates anthropogenic CH_4 emissions over the United States for 2007 and 2008 using comprehensive CH_4 observations from the expanded NOAA/DOE cooperative air sampling network, at the surface, on telecommunications towers, and from aircraft, combined with an atmospheric transport model and a geostatistical inverse modeling (GIM) framework. We utilize auxiliary spatial data (e.g., on population density and economic activity) and leverage concurrent measurements of alkanes to help attribute emissions to specific economic sectors. The work provides spatially-resolved CH_4 emissions estimates and associated uncertainties, as well as information by source sector, both previously unavailable.

2.2 Model and observation framework

We employ the Stochastic Time-Inverted Lagrangian Transport model (STILT) to calculate the transport of CH_4 from emission points at the ground to measurement locations in the atmosphere (Lin et al., 2003). STILT follows an ensemble of particles backward in time, starting from each observation site, using wind fields and turbulence modeled by the Weather Research and Forecasting (WRF) model (Nehrkorn et al., 2010). STILT derives an influence function ("footprint", units: ppb CH_4 per unit emission flux) linking upwind emissions to each measurement. Inputs of CH_4 from surface sources along the ensemble of back-trajectories are averaged to compute the CH_4 concentration for comparison with each observation.

We use observations for 2007 and 2008 from diverse locations and measurement platforms. The principal observations derive from daily flask samples on tall towers in the NOAA measurement network (4984 total observations) and vertical profiles from aircraft (7710 observations, obtained during regular NOAA flights (NOAA ESRL, 2013), regular DOE flights (Biraud et al., 2013), and from the START08 aircraft campaign (Pan et al., 2010)); locations are displayed in Fig. 2.2. We use a GIM framework (Kitanidis & Vomvoris, 1983; Michalak et al., 2004) to analyze the footprints for each of the 12694 observations, and these footprints vary by site and with wind conditions. In aggregate, the footprints provide spatially resolved coverage of most of the continental US, except the south-east coastal region (see suppl. Fig. 8).

The GIM framework, using footprints and concentration measurements, optimizes CH_4 sources separately for each month of 2007 and 2008 on a 1° x 1° latitude-longitude grid for the United States. The contributions of fluxes from natural wetlands are modeled first and subtracted from the observed CH_4 (2.0 TgC yr^{-1} for the continental US); these fluxes are much smaller than anthropogenic sources in the US and thus would be difficult to independently constrain from atmospheric data (see supplement).

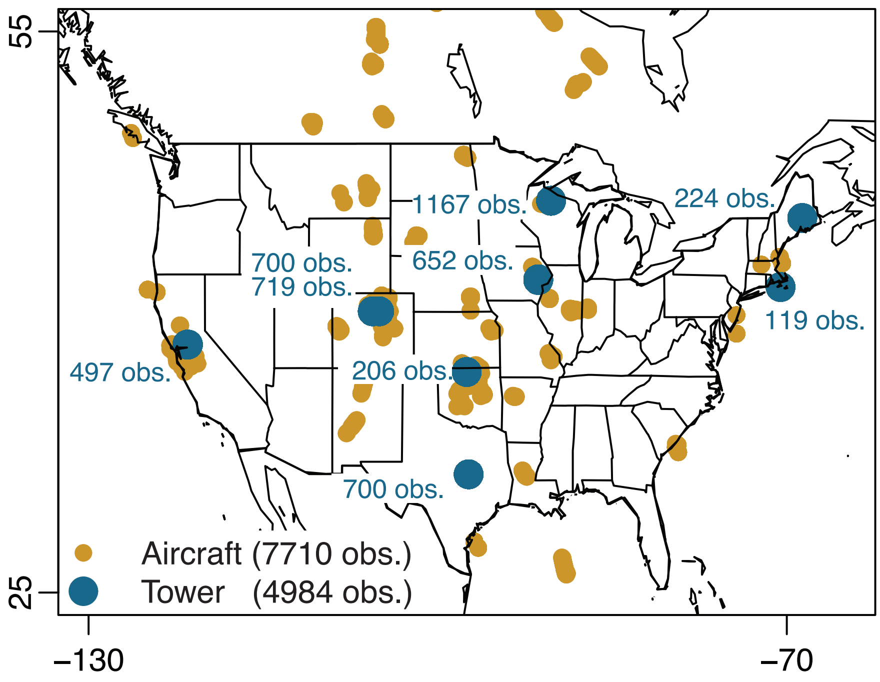


Figure 2.2: CH_4 concentration measurements from 2007 and 2008 and the number of observations associated with each measurement type. Blue text lists the number of observations associated with each stationary tower measurement site.

The GIM framework represents the flux distribution for each month using a deterministic spatial model plus a stochastic spatially-correlated residual, both estimated from the atmospheric observations. The deterministic component is given by a weighted linear combination of spatial activity data from the EDGAR 4.2 inventory; these datasets include any economic or demographic data that may predict the distribution of CH_4 emissions (e.g., gas production, human and ruminant population densities, etc.). Both the selection of the activity datasets to be retained in the model, and the associated weights (emission factors), are optimized to best match observed CH_4 concentrations. Initially, seven activity datasets are included (Table 1) from EDGAR 4.2. We select the minimum number of datasets with the greatest predictive ability using the Bayesian Information Criterion (BIC, see supplement) (Gourdji et al., 2012). BIC numerically scores all combinations of available datasets based on how well they improve goodness of fit and applies a penalty that increases with the number of datasets retained.

The stochastic component represents sources that do not fit the spatial patterns of the activity data. GIM uses a covariance function to describe the spatial and temporal correlation of the stochastic component and optimizes its spatial and temporal distribution simultaneously with the optimization of the activity datasets in the deterministic component (see supplement) (Kitanidis & Vomvoris, 1983; Michalak et al., 2004; Gourdji et al., 2012). Because of the stochastic component, the final emissions estimate can have a different spatial and temporal distribution from any combination of the activity data.

If the observation network is sensitive to a broad array of different source sectors and/or if the spatial activity maps are effective at explaining those sources, many activity datasets will be included in the deterministic model. If the deterministic model explains the observations well, the magnitude of CH_4 emissions in the stochastic component will be small, the assignment to specific sectors will be unambiguous, and uncertainties in the emissions estimates will be small. This is not the case here, as discussed below.

A number of previous studies employed top-down methods to constrain anthropogenic CH_4 sources from global (Chen & Prinn, 2006; Meirink et al., 2008; Bergamaschi et al., 2009; Bousquet et al., 2011; Monteil et al., 2011) to regional (Katzenstein et al., 2003; Kort et al., 2008; Zhao et al., 2009; Kort et al., 2010; Jeong et al., 2012; Peischl et al., 2012; Petron et al., 2012; Wennberg et al., 2012) scales over North America. The regional studies adopted one of three approaches: use a simple box model to estimate an overall CH_4 budget (Katzenstein et al., 2003), estimate a budget using the relative ratios of different gases (Miller et al., 2012a; Peischl et al., 2012; Petron et al., 2012; Wennberg et al., 2012), or estimate scaling factors for inventories by region or source type (Kort et al., 2008; Zhao et al., 2009; Kort et al., 2010; Jeong et al., 2012). The first two methods do not usually give explicit information about geographic distribution. The last approach provides information about the geographic distribution of sources, but results hinge on the spatial accuracy of the underlying regional or sectoral emissions inventories (Law et al., 2002).

Here we are able to provide more insight into the spatial distribution of emissions; like the scaling factor method above, we leverage spatial information about source sectors from an existing inventory, but in addition we estimate the distribution of emissions where the inventory is deficient. We further bolster attribution of regional emissions from the energy industry using the observed correlation of CH_4 and propane, a gas not produced by biogenic processes like livestock and landfills.

2.3 Results

2.3.1 Spatial distribution of CH_4 emissions

Figure 2.3 displays the result of the two-year mean of the monthly CH_4 inversions and differences from the EDGAR 4.2 inventory. We find emissions for the United States that are a factor of 1.7 larger than the EDGAR inventory. The optimized emissions estimated by this study bring

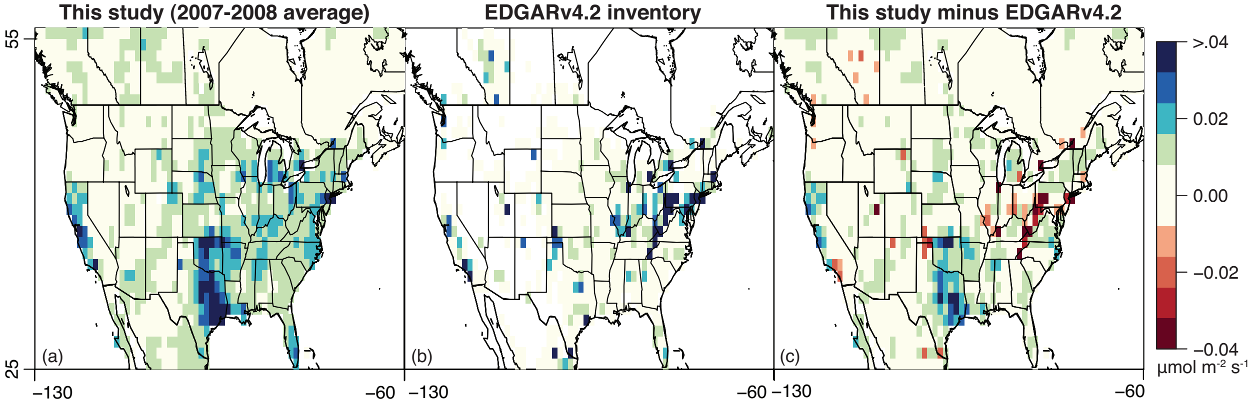


Figure 2.3: The two year averaged CH_4 emissions estimated in this study (a) compared against the commonly used EDGAR 4.2 inventory (b, c). Emissions estimated in this study are greater than in EDGAR 4.2, especially near Texas and California.

the model closer in line with the observations (see Figs. 2.4 and S6). Posterior emissions fit the CH_4 observations ($R^2 = 0.64$, RMSE = 31 ppb) much better than EDGAR v4.2 ($R^2 = 0.23$, RMSE = 49 ppb). Evidently, the spatial distribution of EDGAR sources is inconsistent with emissions patterns implied by the CH_4 measurements and associated footprints.

Several diagnostic measures preclude the possibility of major systematic errors in WRF-STILT. First, excellent agreement between the model and measured vertical profiles from aircraft implies little bias in modeled vertical air mixing (e.g., boundary layer heights, Fig. 2.4). Second, the monthly posterior emissions estimated by the inversion lack statistically significant seasonality (see supplement Fig. 5). This implies that seasonally-varying weather patterns do not produce detectable biases in WRF-STILT. The supplement discusses possible model errors and biases in greater detail.

CH_4 observations are sparse over parts of the southern and central East Coast, and in the Pacific Northwest. Emissions estimates for these regions therefore rely more strongly on the deterministic component of the flux model, with weights constrained primarily by observations elsewhere. Hence, emissions in these areas, including from coal mining, are poorly constrained

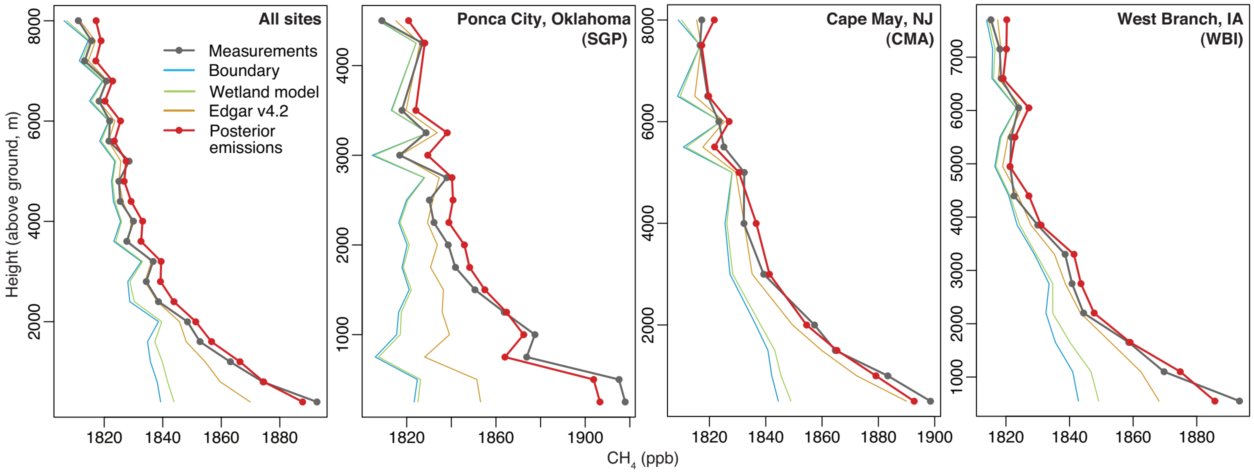


Figure 2.4: A model-measurement comparison at several regular NOAA aircraft monitoring sites (averaged over 2007–2008). Plots include the measurements; the modeled boundary condition; the summed boundary condition and wetland contribution (from the Kaplan model); and the summed boundary, wetland, and anthropogenic contributions (from EDGAR v4.2 and the posterior emissions estimate).

(see supplement).

2.3.2 Contribution of different source sectors

Only two spatial activity datasets in Table 1 are selected through the BIC as meaningful predictors of CH_4 observations over the US: population densities of humans and ruminants. Some sectors are eliminated by the BIC because emissions are situated far from observation sites (e.g., coal mining in West Virginia or Pennsylvania), making available CH_4 data insensitive to these predictors. Other sectors may strongly affect observed concentrations but are not selected, indicating that the spatial datasets from EDGAR are poor predictors for the distribution of observed concentrations (e.g., oil and natural gas extraction and oil refining). Sources from these sectors appear in the stochastic component of the GIM (see supplement).

The results imply that existing inventories underestimate emissions from two key sectors: ruminants and fossil fuel extraction and/or processing, discussed in detail below.

We use the optimized ruminant activity dataset to estimate the magnitude of emissions with spatial patterns similar to animal husbandry and manure. Our corresponding US budget of 12.7 $\pm$ 5.0 TgC yr^{-1} is nearly twice that of EDGAR and EPA (6.7 and 7.0, respectively). The total posterior emissions estimate over the northern plains, a region with high ruminant density but little fossil fuel extraction, further supports the ruminant estimate (Nebraska, Iowa, Wisconsin, Minnesota, and South Dakota). Our total budget for this region of 3.4 $\pm$ 0.7 compares to 1.5 TgC yr^{-1} in EDGAR. Ruminants and agriculture may also be partially responsible for high emissions over California (Jeong et al., 2013). EDGAR activity datasets are poor over California (Xiang et al., 2013a), but several recent studies (Zhao et al., 2009; Jeong et al., 2012; Peischl et al., 2012; Wennberg et al., 2012; Jeong et al., 2013) have provided detailed top-down emissions estimates for the state using datasets from state agencies.

Existing inventories also greatly underestimate CH_4 sources from the south-central US (see Fig. 2.3). We find the total CH_4 source from Texas, Oklahoma, and Kansas to be 8.1 $\pm$ 0.96 TgC yr^{-1}, a factor of 2.7 higher than the EDGAR inventory. These three states alone constitute $\sim$24% $\pm$ 3% of the total United States anthropogenic CH_4 budget or 3.7% of net US greenhouse gas emissions (in CO_2 equivalents (US EPA, 2013)).

Texas and Oklahoma were among the top five natural gas producing states in the country in 2007 (US Energy Information Administration, 2013), and aircraft observations of alkanes indicate that the natural gas and/or oil industries play a significant role in regional CH_4 emissions. Concentrations of propane (C_3H_8), a tracer of fossil hydrocarbons (Koppmann, 2008), are strongly correlated with CH_4 at NOAA/DOE aircraft monitoring locations over Texas and Oklahoma ($R^2 = 0.72$, see Fig. 2.5). Correlations are much weaker at other locations in North America ($R^2 = 0.11$ to 0.64).

We can obtain an approximate CH_4 budget for fossil fuel extraction in the region by subtracting the optimized contributions associated with ruminants and population from the total emis-

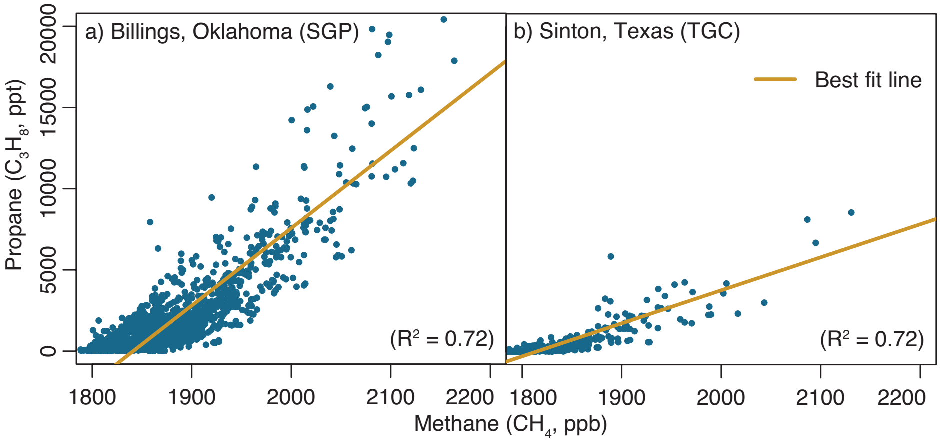


Figure 2.5: Correlations between propane and CH_4 at NOAA/DOE aircraft observation sites in Oklahoma (a) and Texas (b) over 2007 – 2012. Correlations are higher in these locations than any other North American sites, indicating large contributions of fossil fuel extraction and processing to CH_4 emitted in this region.

sions. The residual (see Fig. S2c) represents sources that have spatial patterns not correlated with either human or ruminant density in EDGAR. Our budget sums to 3.7 ± 2.0 TgC yr^{-1}, a factor of 4.9 ± 2.6 larger than oil and gas emissions in EDGAR v4.2 (0.75 TgC yr^{-1}) and a factor of 6.7 ± 3.6 greater than EDGAR sources from solid waste facilities (0.55 TgC yr^{-1}), the two major sources that may not be accounted for in the deterministic component. The population component likely captures a portion of the solid waste sources, so this residual methane budget more likely represents natural gas and oil emissions than landfills. The supplement discusses in detail the uncertainties in this sector-based emissions estimate. We currently do not have the detailed, accurate, and spatially resolved activity data (fossil fuel extraction and processing, ruminants, solid waste) that would provide more accurate sectorial attribution.

Katzenstein et al. (2003) were the first to report large regional emissions of CH_4 from Texas, Oklahoma, and Kansas; they cover an earlier time period (1999–2002) than this study. They used a box model and 261 near-ground CH_4 measurements taken over 6 days to estimate a total Texas-Oklahoma-Kansas CH_4 budget (from all sectors) of 3.8 ± 0.75 TgC yr^{-1}. We revise their estimate upward by a factor of two based on the inverse model and many more measurements from different platforms over two full years of data. The supplement further compares the CH_4 estimate in Katzenstein et al. and in this study.

2.4 Discussion and summary

This study combines comprehensive atmospheric data, diverse datasets from the EDGAR inventory, and an inverse modeling framework to derive spatially-resolved CH_4 emissions and information on key source sectors. We estimate a mean annual US anthropogenic CH_4 budget for 2007 and 2008 of 33.4 ± 1.4 TgC yr^{-1} or $\sim$7–8% of the total global CH_4 source. This estimate is a factor of 1.5 and 1.7 larger than EPA and EDGAR v4.2, respectively. CH_4 emissions from Texas, Oklahoma, and Kansas alone account for 24% of US methane emissions, or 3.7% of the total US greenhouse gas budget.

The results indicate that drilling, processing, and refining activities over the south-central United States have emissions as much as 4.9 ± 2.6 times larger than EDGAR, and livestock operations across the US have emissions approximately twice that of recent inventories. The US EPA recently decreased its CH_4 emission factors for fossil fuel extraction and processing by 25–30% (for 1990–2011) (US EPA, 2013), but we find that CH_4 data from across North America instead indicate the need for a larger adjustment of the opposite sign.

3

Comment: Linking methane emissions inventories with atmospheric observations

The following comment was published in the Proceedings of the National Academy of Sciences (Miller et al., 2014b). It is a reply to a letter that Hristov et al. (2014) wrote in response to Miller et al. (2013).

Hristov et al. (2014) argue that our study "provides a comprehensive, quantitative analysis of anthropogenic methane sources," but that the conclusion "that US EPA [US Environmental Protection Agency] estimates for livestock methane emissions are grossly underestimated appears to be unsubstantiated by ... [a] 'bottom-up' approach" outlined in their letter.

In this reply, we discuss the information provided by atmospheric methane data about methane emissions, and comment on the challenge of connecting "bottom-up" and "top-down" estimates, a conclusion shared by Hristov et al. (2014).

Our study (Miller et al., 2013) used both near-surface and airborne atmospheric measurements of CH_4 concentrations to characterize the total mass of methane added to the atmosphere by surface emissions, discretized in space and time. We conclude that total United States methane emissions in 2007–2008 were 33.4 ± 1.4 TgC/yr (44.5 $TgCH_4$/yr), 45–57% above the most recent US EPA baseline estimate for those years (US EPA, 2013). Furthermore, we estimate "the magnitude of emissions with spatial patterns similar to animal husbandry and manure" (Miller et al., 2013) at 12.7 ± 5.0 TgC/yr (16.9 $TgCH_4$/yr), 11–156% above baseline EPA estimates for those sectors (best estimate 84% above EPA). Our conclusions are generally consistent with previous more limited top-down studies examining total United States (e.g., Kort et al., 2008) and regional livestock/manure methane emissions (e.g., Jeong et al., 2013).

Hristov et al. (2014) argue that "the validity of this 'top-down' approach can be verified by a relatively simple 'bottom-up' method using current livestock inventories and enteric or manure methane emission factors." The authors build this estimate for enteric fermentation by multiplying the US Department of Agricuture (USDA) livestock inventory estimates for 2013 (note that our study covers 2007–2008), by "assumed" feed dry matter intake and "assumed" methane production rates. "With the above assumptions," Hristov et al. (2014) estimate methane emissions from enteric fermentation comparable to the US EPA's inventory for 2011. Similarly, the authors use USDA livestock inventories and Intergovernmetal Panel on Climate Change (IPCC) (Intergovernmental Panel On Climate Change, 2006) manure methane emissions factors to estimate United States manure emissions that are 35% lower than EPA inventory numbers.

The estimates of Hristov et al. (2014) therefore require a series of assumptions, for which errors compound as several factors are multiplied and added. Feed matter intake and emission factors both have substantial uncertainties (Intergovernmental Panel On Climate Change, 2006), as do the IPCC manure methane emission factors (Intergovernmental Panel On Climate Change, 2006),. Given these uncertainties, which are inherent in all bottom-up inventories, we strongly disagree that "the validity of [our] 'top-down' approach can be verified" using the Hristov et al. (2014) estimates.

The method we applied is especially suited to quantifying large-scale total emissions, and uncertainties increase for sector- and region-specific estimates [as outlined above and in our study (Miller et al., 2013)]. Even in light of these uncertainties, the total emissions with spatial patterns consistent with animal husbandry are still likely to be substantially above EPA estimates. Conversely, bottom-up inventories are strongest at detailing individual emission types, but uncertainties compound at larger scales, such as the national scale examined here. This difference is precisely why we argue that careful, detailed assessments are needed to reconcile the emissions clearly visible from atmospheric observations with bottom-up emissions inventories. Hristov et al. (2014) also note a "need for a detailed inventory ... to more accurately estimate ... emissions." On this point we strongly agree.

4

Observational constraints on the distribution, seasonality, and environmental predictors of North American boreal methane emission

This chapter originally appeared as a research article in Global Biogeochemical Cycles. For the original article, refer to Miller et al. (2014c).

Wetlands comprise the single largest global source of atmospheric methane, but current flux estimates disagree in both magnitude and distribution at the continental scale. This study uses atmospheric methane observations over North America from 2007–2008 and a geostatistical inversion to improve understanding of Canadian methane fluxes and associated biogeochemical models. The results bridge an existing gap between traditional top-down, inversion studies, which typically emphasize total emission budgets, and biogeochemical models, which usually emphasize environmental processes. The conclusions of this study are threefold. First, the most complete process-based methane models do not always describe available atmospheric methane

observations better than simple models. In this study, a relatively simple model of wetland distribution, soil moisture, and soil temperature outperformed more complex model formulations. Second, we find that wetland methane fluxes have a broader spatial distribution across western Canada and into the northern US than represented in existing flux models. Finally, we calculate total methane budgets for Canada and for the Hudson Bay Lowlands, a large wetland region (50–60°N, 75–96°W). Over these lowlands, we find total methane fluxes of 1.8 $\pm$ 0.24 TgC yr^{-1}, a number in the mid-range of previous estimates. Our total Canadian methane budget of 16.0 $\pm$ 1.2 TgC yr^{-1} is larger than existing inventories, primarily due to high anthropogenic emissions in Alberta. However, methane observations are sparse in western Canada, and additional measurements over Alberta will constrain anthropogenic sources in that province with greater confidence.

4.1 Introduction

Atmospheric methane (CH_4) is the second-most important long-lived greenhouse gas, and since the pre-industrial era, its radiative forcing has increased to 0.507 W m^{-2}, approximately one third that of CO_2 (Butler, 2014). Therefore, greenhouse gas reduction strategies and future climate predictions will require accurate estimates of methane emissions. Total global emissions are constrained to approximately $\pm$15% using observations of the global CH_4 burden and rate of increase, combined with an estimate of the CH_4 atmospheric lifetime (e.g. Kirschke et al., 2013). However, uncertainties in emissions from individual source types can be greater than a factor of two (O'Connor et al., 2010; Dlugokencky et al., 2011; Melton et al., 2013a). For example, wetlands likely constitute the largest single source of atmospheric methane, but estimates of global fluxes vary from 60 to 213 TgC yr^{-1} (80 to 284 $TgCH_4$ yr^{-1}), meaning they comprise anywhere from 14 to 50% of the total budget (e.g., O'Connor et al., 2010; Melton et al., 2013a; Bridgham et al., 2013;

Kirschke et al., 2013; Ciais et al., 2013, and references therein). Anthropogenic sources (e.g., fossil fuel extraction and processing, ruminants, landfills), by comparision, likely account for 50 – 65% of total global emissions (Ciais et al., 2013; Kirschke et al., 2013). Uncertainties in methane fluxes are larger at the regional scale; estimates of methane from the Hudson Bay Lowlands (HBL), a large boreal wetland region in Canada, range from 0.28 – 8.5 TgC yr^{-1} (Roulet et al., 1992; Worthy et al., 2000; Pickett-Heaps et al., 2011; Melton et al., 2013a).

The present study focuses on improving methane flux estimates from boreal wetlands. These regions are a particular concern because of their large soil carbon stocks. Methane fluxes in wetlands occur primarily in waterlogged, anaerobic soil conditions due to the decomposition of organic material by methanogenic Archaea. Boreal and arctic regions are far less productive than many other ecosystems but nonetheless play a vital role in the global carbon cycle. These northerly regions may contain half of all wetlands and soil carbon in the world ($\sim$1700 PgC), twice the amount of carbon currently held within the atmosphere (Tarnocai et al., 2009).

Evidence suggests that high-latitude wetlands are already changing due to an evolving climate and that ecosystem changes may accelerate (Tarnocai, 2009; Avis et al., 2011; Schuur et al., 2013). For example, most studies predict that climate change will increase methane fluxes from boreal and arctic regions; estimates range from 6% to 35% increase in methane fluxes per °C of global temperature increase (e.g., Gedney et al., 2004; Khvorostyanov et al., 2008; O'Connor et al., 2010; Koven et al., 2011; Zhu et al., 2011).

Three factors may explain the large differences among model estimates of boreal methane fluxes. First, models differ in their underlying environmental variables. For example, existing models of global wetland area range from 2.6 to 9 $\times 10^6$ km^2 (Petrescu et al., 2010) and have differing spatial distributions (especially over boreal North America, Melton et al., 2013a). Second, models further differ in functional form (see section 4.2.4), due in part to uncertainties and/or complexity in biophysical methane processes. For example, many models relate maps of soil temperature to wetland methane fluxes using a coefficient known as Q10. This coefficient describes the factor by which a reaction rate increases per 10°C rise in temperature. Estimates of this coefficient range from 1 to 35, largely due to microbial and soil heterogeneity (van Hulzen et al., 1999; Whalen, 2005; O'Connor et al., 2010; Lupascu et al., 2012). Finally, differences among existing flux models also stem from difficulties extrapolating from plot-level to regional scale. Most flux models calibrate to individual wetland sites and extrapolate to regional or global scales (O'Connor et al., 2010; Zhang et al., 2012). However, small scale study sites exhibit substantial heterogeneity, and fluxes can vary by an order of magnitude over micro-topography on the centimeter scale (Waddington & Roulet, 1996; Comas et al., 2005; Hendriks et al., 2010).

Top-down approaches like inverse modelling provide one means of reducing the wide uncertainty in wetland methane fluxes. Top-down studies use atmospheric methane measurements and meteorological models to improve existing flux estimates at regional (Zhao et al., 2009; Bergamaschi et al., 2010; Villani et al., 2010; Kim et al., 2011) and global (Chen & Prinn, 2006; Bergamaschi et al., 2013; Fraser et al., 2013) scales. Most existing methods emphasize total emissions budgets and provide relatively little information on wetland processes, but two recent publications begin to bridge this gap. Spahni et al. (2011) conduct a global scale inversion that estimates fluxes by wetland type. Pickett-Heaps et al. (2011) use atmospheric methane measurements from northern Ontario to assess the magnitude and seasonal structure of a wetland flux model over the HBL. Results imply a premature seasonal onset of fluxes in this model, referred to as the "Kaplan model." The authors suggest removing fluxes from snow-covered regions as one possible solution. In spite of these recent studies, existing top-down approaches provide limited assessment of the underlying environmental variables or the functional form of existing wetland flux models.

The present study moves closer to integrating top-down flux estimates with process-based, bottom-up modeling methods. First, we explore how atmospheric methane measurements can be used to construct and assess biogeochemical process models at continental scale. Second, we use a broad network of measurement sites in Canada and the US to understand the spatial and seasonal distribution of North American boreal wetland fluxes. To achieve these goals, we combine in situ methane measurements across Canada and the United States from 2007 and 2008, a regional atmospheric transport model, and a geostatistical inverse modeling framework.

4.2 Model and measurements

The methods sections and subsequent discussion are organized as follows. First, we describe the atmospheric model and measurements (sections 4.2.1, 4.2.2, and 4.2.3). Using this model, we compare two existing wetland flux estimates, Kaplan and DLEM, against atmospheric methane observations. Both flux estimates are described in detail below (sections 4.2.4). We subsequently use a geostatistical inverse modeling framework to estimate North American boreal methane fluxes (section 4.3.1). This flux estimate has two components. The first component, termed the deterministic model, is a combination of environmental predictors (e.g., soil moisture, temperature, etc.) that best represents the methane fluxes, as seen through the atmospheric methane observations (section 4.3.2). The second component, termed the stochastic component, estimates the spatial and/or temporal flux patterns that may be lacking in the environmental predictors, and therefore cannot be modeled using the deterministic model. The geostatistical inverse model produces a final best estimate, termed the posterior fluxes, and it is the sum of the deterministic and stochastic components.

4.2.1 The regional atmospheric model

We simulate in situ methane mixing ratios using STILT, the Stochastic, Time-Inverted, Lagrangian Transport model (Lin et al., 2003). STILT is a particle model; an ensemble of air-following particles is released from each methane observation site. In this study, a new 500-particle ensemble is initiated for each of the hourly methane measurements. These particles travel backward in time along the wind fields of a meteorology model, in this case for 10 days. STILT further includes stochastic motions that simulate boundary layer turbulence.

Wind fields from the Weather Research and Forecasting model (WRF version 2.1.2) are used to drive STILT trajectories in this study. Nehrkorn et al. (2010), Hegarty et al. (2013), and the supplement describe this meteorology in greater detail. The WRF fields used here have a nested resolution; 10-km within 24-48 hours of the observation sites and 40-km in more distant regions (see supplement).

STILT subsequently uses the trajectories to calculate a footprint map. The footprints relate the surface fluxes in North America to the concentration increment seen at the measurement location and have units of mixing ratio per unit surface flux. This footprint is based on the number of particles in a region and their altitudes relative to the planetary boundary layer.

The STILT setup here incorporates fluxes from existing inventories on a $\frac{1}{4}^{\circ}$ by $\frac{1}{6}^{\circ}$ longitude–latitude grid (11 to 65°N and 145 to 51°W).

4.2.2 Model boundary condition

STILT only models emissions over the North American continent. The model therefore requires a boundary condition to represent the concentration of methane in incoming air over the Pacific and Arctic oceans before reaching North American sources. This study uses an empirical boundary curtain that interpolates a variety of trace gas measurements from ground-based sites and aircraft in the NOAA ESRL Global Monitoring Division's Cooperative Global Air Sampling Network. The resulting boundary curtain varies latitudinally and vertically and has a daily temporal resolution (see the supplement). The estimated boundary condition value associated with each STILT particle run depends on the ending latitude, altitude, and day of each particle. This boundary value is then added to the modeled methane signal from North American sources. The sum can be directly compared against measured methane mixing ratios at tower sites across Canada and the northern US (e.g., section 4.3.1).

4.2.3 Measurements

This study uses observed methane mixing ratios for 2007 to 2008 from five observation sites sensitive to boreal wetland fluxes: hourly measurements from 4 Canadian observation towers and daily flask measurements from one US tall tower. Sites (from east to west) include Chibougamau, Quebec (CHM, 50°N, 74°W, 30 m above ground level); Fraserdale, Ontario (FSD, 50°N, 83°W, 40 m agl); Park Falls, Wisconsin (LEF, 46°N, 90°W, 244 m agl); East Trout Lake, Saskatchewan (ETL, 54°N, 104°W, 105 m agl); and Candle Lake, Saskatchewan (CDL, 54°N, 105°W, 30 m agl, 2007 only) (Fig. 4.1).

Small scale heterogeneities caused by turbulent eddies and incomplete mixing make it difficult to model hourly-scale variability in the in situ data. STILT also has difficulty estimating the very shallow nighttime boundary layer and therefore rarely captures variations in nighttime concentrations. Hence, this study uses afternoon averages of the methane data and model output (1pm - 7pm local time), a total of 2,485 observations after averaging.

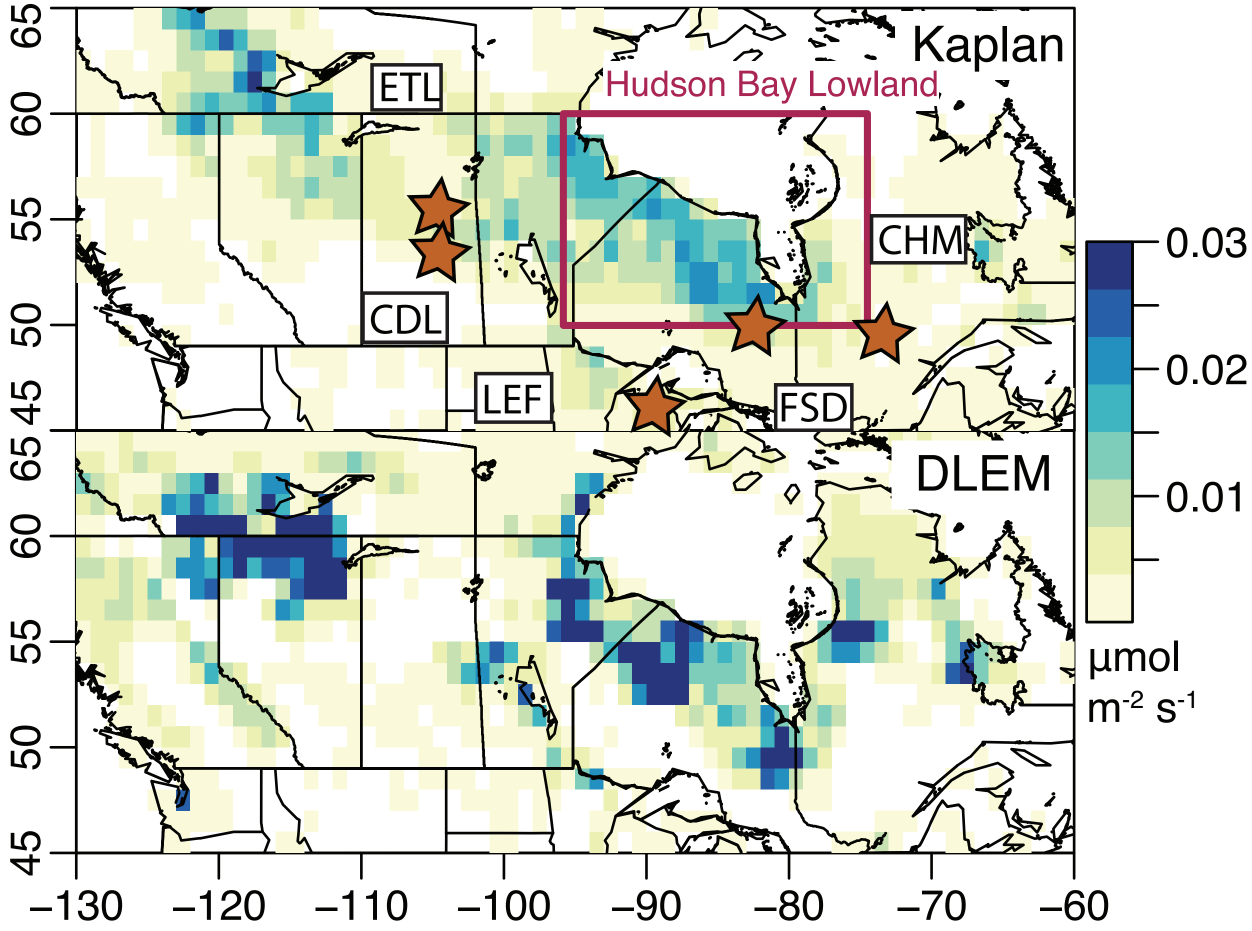


Figure 4.1: Summer mean wetland fluxes from the Kaplan and DLEM wetland methane models (for July, August, and September, averaged over 2007-2008). Both models estimate similar annual totals for the HBL, but DLEM has a more pronounced summer peak.

4.2.4 Existing flux models

The Kaplan model

The first inventory used in this study is the Kaplan model, described in Kaplan (2002) and Pickett-Heaps et al. (2011) (Fig. 4.1). The model has the following functional form:

$$\begin{aligned} E &= \delta b W f(T) \sum_{i=1}^{2} \frac{C_i}{\tau_i} \\ f(T) &= 1.0 \times 10^3 F e^{\frac{-309}{T-227}} \end{aligned} \tag{4.1}$$

where E is the wetland flux expressed here in units of μmol m^{-2} s^{-1}. C_1 and C_2 represent the moles of carbon per unit area in soil and litter, normalized by their respective lifetimes (τ_1 and τ_2). This soil carbon estimate is taken from the Lund-Potsdam-Jena (LPJ) dynamic global vegetation model (Sitch et al., 2003). W is the maximum possible extent of wetlands in a given grid box (a fraction, from LPJ), and δ is a measure of whether wetlands are actually present ($\delta = 0$ if soil moisture (M) $< 10\%$ and $\delta = 1$ otherwise). An emissions factor b represents the fraction of methane per mol carbon respired ($b = 3 \times 10^{-2}$). $f(T)$ represents an Arrhenius equation of temperature (in Kelvin). In this equation, F adjusts the inventory based on soil temperature to better match differences between boreal (B) and tropical (T) wetlands (refer to Pickett-Heaps et al., 2011). We build this inventory with soil moisture (for δ) and soil temperature (T) from WRF (the same meteorology used to drive STILT) at a soil depth of 25cm. This soil depth provided the best match between the Kaplan model and atmospheric methane observations. This setup differs from Pickett-Heaps et al. (2011), who used surface skin temperature instead of soil temperature below the ground surface.

The LPJ model outputs used here for wetland coverage and soil carbon are updated from previous studies that also used the LPJ/Kaplan model (e.g., Bergamaschi et al., 2007; Pickett-Heaps et al., 2011). Among other updates, soil carbon is approximately a factor of four lower than in the previous studies listed above. This adjustment matches the LPJ model against upland soil profiles, but the change appears inconsistent with methane observations over boreal wetlands (e.g., Pickett-Heaps et al., 2011). We re-adjust the LPJ soil carbon estimate upward by a factor of 4.15 to match the LPJ/Kaplan model in Pickett-Heaps et al. (2011). This previous study compares the Kaplan model against measurements from Fraserdale (FSD), Ontario, and likely better represents high latitude soil carbon than the new LPJ estimate.

The DLEM model

DLEM, the Dynamic Land Ecosystem Model, includes more complexity than the Kaplan model described above (Tian et al., 2010, 2012) (Fig. 4.1). It models the production of methane in soil pore water (P, expressed here in μmol m^{-2} s^{-1}), and only a fraction of methane produced is released to the atmosphere (E). This fraction depends on a multitude of factors that are discussed in depth by Tian et al. (2010): plant-mediated transport, diffusive flux, ebullition, oxidation by methanotrophy, and oxidation during plant-mediated transport.

Methane production in soil pore water (P) is simpler to describe:

$$P = P_{\max} \times \frac{[\mathrm{DOC}]}{[\mathrm{DOC}] + k} \times f(T) \times f(\mathrm{pH}) \times f(M) \tag{4.2}$$

where $P_{\max}$ is the maximum possible rate of CH_4 production in soils, a spatially variable parameter (see Tian et al., 2010). [DOC] is dissolved organic content, determined by gross primary productivity, litter fall, and soil organic matter decomposition rates (Tian et al., 2012). k is the half-saturation coefficient, $f(T)$ is the effect of soil temperature, $f(\mathrm{pH})$ is the effect of soil pH, and $f(M)$ is the effect of soil water content. The functions of temperature, pH, and soil moisture have the following forms (Tian et al., 2010):

$$f(T) = \begin{bmatrix} 0 & T < 268.15 \\ 2.5^{(T-303.15)/10} & 303.15 > T \geq 268.15 \\ 1 & T \geq 303.15 \end{bmatrix} \tag{4.3}$$

$$f(pH) = \begin{bmatrix} 0 & \text{pH} < 4.0 \text{ or pH} \geq 10.0 \\ \frac{1.02}{1+1e6\times\exp\left[-2.5pH\right]} & 4.0 < \text{pH} < 7.0 \\ \frac{1.02}{1+1e6\times\exp\left[-2.5(14.0-pH)\right]} & 7.0 < \text{pH} < 10.0 \end{bmatrix} \tag{4.4}$$

$$f(M) = \begin{bmatrix} 0 & M \leq M_{fc} \\ 0.368\left(\frac{M-M_{fc}}{M_s-M_{fc}}\right)^2 \exp\left[\frac{M-M_{fc}}{M_s-M_{fc}}\right] & M_{fc} < M < M_s \\ 1 & M \geq M_s \end{bmatrix} \tag{4.5}$$

where M_{fc} is the field capacity, and M_s is the saturated water content of soil. Tian et al. (2010) provide a graphical depiction of these functional dependences.

The environmental data for DLEM are derived from a number of sources: meteorological data from North American Regional Reanalysis (NARR, Mesinger et al., 2006) and land cover/vegetation data from a combination of sources (see Tian et al., 2010, for more detail).

4.3 Statistical framework

4.3.1 Conceptual overview

We implement a geostatistical inverse model to infer information about methane fluxes and to assess the environmental drivers in existing wetland models. The statistical approach follows that of Kitanidis & Vomvoris (1983), Michalak et al. (2004), and Gourdji et al. (2012). The inversion estimates the spatial and temporal distribution of emissions that is most likely given the atmospheric methane measurements and the transport information provided by the atmospheric model.

The inversion first requires a linear expression for the model-measurement framework:

$$z = \mathbf{H}s + \varepsilon \tag{4.6}$$

where s $(m \times 1)$ are the true, unknown fluxes. Unlike the wetland-specific fluxes estimated by Kaplan and DLEM (E), s encompasses fluxes from all source types. z is the $n \times 1$ vector of observed mixing ratios minus the estimated boundary condition value (see section 4.2.2). $\mathbf{H}$ $(n \times m)$ are the footprints computed by STILT (section 4.2.1). Finally, ε $(n \times 1)$ describes model-data mismatch — all errors unrelated to an imperfect emissions estimate (e.g., transport error, aggregation error, etc.). This vector is assumed to follow a multivariate normal distribution with a mean of zero:

$$\varepsilon = \mathcal{N}(0, \mathbf{R}) \tag{4.7}$$

where $\mathbf{R}$ $(n \times n)$ is the covariance matrix of these errors.

Using the above framework, the inversion then models the unknown fluxes (s, Eq. 4.6) using the following structure:

$$s \;=\; \mathbf{X}\beta + \mathcal{N}(0, Q) \tag{4.8}$$

The first component of the statistical model ($\mathbf{X}\beta$) is a weighted least squares regression and is termed the 'deterministic model' or 'inversion prior' (section 4.3.2). Each column of $\mathbf{X}$ (dimensions $m \times p$) is a predictor in the weighted regression (e.g., Gourdji et al., 2008, 2012). In this study, $\mathbf{X}$ includes datasets termed 'auxiliary data' (e.g., soil temperature, moisture, an anthropogenic emissions inventory, etc.) that help explain the spatial and seasonal distribution of methane fluxes. Additionally, one column of this matrix is constant, equivalent to the intercept of the regression. The regression coefficients (β, dimensions $p \times 1$) are unknown and are estimated in the inversion using the atmospheric methane data.

The second component the geostatistical inverse model, $\mathcal{N}(0, Q)$, is termed the 'stochastic component' or the 'spatially-correlated residual.' The stochastic component adjusts, at grid scale, the fluxes estimated by the deterministic model. This component, for example, can correct the deterministic model if any environmental data in $\mathbf{X}$ have the incorrect distribution. The covariance matrix Q (dimensions $m \times m$) describes the magnitude and the spatiotemporal correlation of the stochastic component. It includes off-diagonal elements that follow an exponential covariance model: any fluxes estimated by the stochastic component will be spatially correlated with a given decorrelation length. This spatial correlation means that the stochastic component can adjust the flux estimate on a fine grid-scale relative to the density of atmospheric observations (e.g., Michalak et al., 2004; Mueller et al., 2008; Villani et al., 2010; Bergamaschi et al., 2013; Miller et al., 2013).

The best estimate of a geostatistical inversion is obtained by minimizing a cost function (L) with respect to the methane fluxes (s) and the coefficients (β) (e.g., Kitanidis & Vomvoris, 1983; Michalak et al., 2004):

$$L_{s,\beta} = \frac{1}{2}(z - \mathbf{H}s)^T \mathbf{R}^{-1}(z - \mathbf{H}s) + \frac{1}{2}(s - \mathbf{X}\beta)^T \mathbf{Q}^{-1}(s - \mathbf{X}\beta) \tag{4.9}$$

The supplement discusses further details of the statistical setup. In particular, we implement the inversion with Lagrange multipliers to prevent negative fluxes (see supplement, Miller et al., 2014a). Furthermore, we estimate the covariance matrices ($\mathbf{R}$ and $\mathbf{Q}$) using restricted maximum likelihood estimation (REML) (Kitanidis, 1995; Michalak et al., 2004).

We use this statistical framework to estimate monthly methane fluxes (s) on a 1° by 1° longitude-latitude grid over the years 2007 and 2008, yielding 41,328 total locations in space and time. The geographic domain of the inversion spans from 35 to 65°N latitude and 145 to 51°W longitude.

4.3.2 The deterministic model of fluxes

The following sections discuss the deterministic model in greater detail.

Auxiliary environmental data

We consider a number of auxiliary datasets or predictors for use in the deterministic model. Ultimately, only a selection of these datasets is used in the inversion depending on how well each explains the atmospheric methane data (see section 4.3.2). These datasets include both environmental drivers of wetland fluxes and inventory data on anthropogenic emissions. The full array of possible datasets for $\mathbf{X}$ are shown in Table 4.1. These include meteorological data from WRF (used in this version of Kaplan model, Nehrkorn et al., 2010) and NARR (used in DLEM, Mesinger et al., 2006). We consider soil carbon estimates from the LPJ model (used in Kaplan, Sitch et al., 2003; Pickett-Heaps et al., 2011) and the Northern Circumpolar Soil Carbon Database (NCSCD) (Tarnocai et al., 2009; Hugelius et al., 2013). Wetland coverage estimates include model output from LPJ and surface water data from the Global Inundation Extent from Multi-Satellites (GIEMS) database (Prigent et al., 2007; Papa et al., 2010). Refer to the supplement for maps of these auxiliary datasets.

In addition to wetland-related datasets, we also consider multiple datasets or predictors for the distribution of anthropogenic emissions. Specifically, we consider including the EDGAR v4.2 anthropogenic inventory in the deterministic model as well as the individual sector-by-sector emissions estimates from EDGAR. A companion study found that EDGAR v4.2 did not match the estimated distribution of anthropogenic emissions in the United States (Miller et al., 2013). Hence, we consider additional proxies other than EDGAR v4.2 for the spatial distribution of

Table 4.1: Auxiliary data or predictors tested for use in the deterministic model. Note: The second column (static/variable) lists whether the auxiliary data in question is seasonally constant or varies temporally. Also, soil moisture and temperature are available at multiple vertical soil levels in WRF and NARR: 5, 25, 70, and 150 cm depth in WRF and 0, 10, 40, and 100cm depth in NARR.

Description	Static/variable	Source model
Liquid soil moisture (e.g., not frozen) (M)	variable	WRF, NARR
Total soil moisture (liquid + frozen) (M_{Tot})	variable	WRF, NARR
Soil temperature	variable	WRF, NARR
Wetland coverage fraction (W)	static	LPJ, GIEMS
Soil carbon content (C)	static	LPJ, NCSCD
Estimated distribution of anthropogenic emissions	static	EDGAR v4.2
Smooth tricubic functions	static	

anthropogenic emissions. For example, we construct smooth tri-cube functions centered over known anthropogenic source regions (e.g., Alberta, Oklahoma, California, the US East Coast; refer to the supplement). The subsequent section discusses how to choose among this array of auxiliary datasets when constructing the deterministic model.

Selection of auxiliary data

It would be ill-advised to use all auxiliary datasets from Table 4.1 in the deterministic model; the resulting model would be an over-fit with problematic colinearity (e.g., Zucchini, 2000). We instead use a statistical selection method to choose an optimal set of auxiliary datasets for the deterministic model. These methods select as many datasets for **X** as can explain variability in the methane fluxes but will prevent an over-fit or unreliable coefficient estimates. We implement one of the most common methods, the Bayesian information criterion (BIC) (as in Gourdji et al., 2012). The BIC numerically scores all possible combinations of auxiliary data based on how well they reduce the model-measurement residuals and applies an increasing penalty for model complexity (refer to the supplement). Specifically, this penalty increases with the number of columns in **X** and with the log of the number of observations. Unlike frequentist statistics, these scores do not support p-values or traditional hypothesis testing. The best model is simply the one with the lowest score. Kass & Raftery (1995) provide a qualitative assessment of model strength based on the difference in BIC scores. A score difference greater than two is "worth mentioning" and greater than 10 is "very strong."

In many cases, one might expect that the product of two or more different environmental variables may be a better predictor than an additive model, so we test multiplicative interactions among the wetland-related auxiliary datasets. Additionally, several of the auxiliary datasets are co-linear (e.g., total soil moisture and unfrozen soil moisture), and we are careful not to include similar or co-linear predictors in the same candidate model for $\mathbf{X}$. For consistency, we do not mix WRF and NARR datasets in the same candidate model.

4.4 Results and discussion

4.4.1 Model-data comparison using existing flux estimates

Methane concentrations modeled with existing flux estimates exhibit a variable fit against the atmospheric data (see Fig. 4.2). For example, both the Kaplan and DLEM models match the general shape of the seasonal cycle at eastern tower sites (LEF, FSD, CHM) but underestimate the magnitude of the measurements. Among these sites, models match observations most closely at Fraserdale, Ontario (FSD), possibly because Pickett-Heaps et al. (2011) validated the Kaplan model at Fraserdale. Existing methane flux estimates, however, perform far worse at the western sites (CDL and ETL). For example, the models underestimate both observed summer and winter maxima at these sites. The observed summer maxima are likely caused by peak summer wetland fluxes while the winter maxima likely reflect a combination of advected anthropogenic emissions and limited vertical mixing within the troposphere. This result implies that existing inventories underestimate both wetland and anthropogenic fluxes in western Canada.

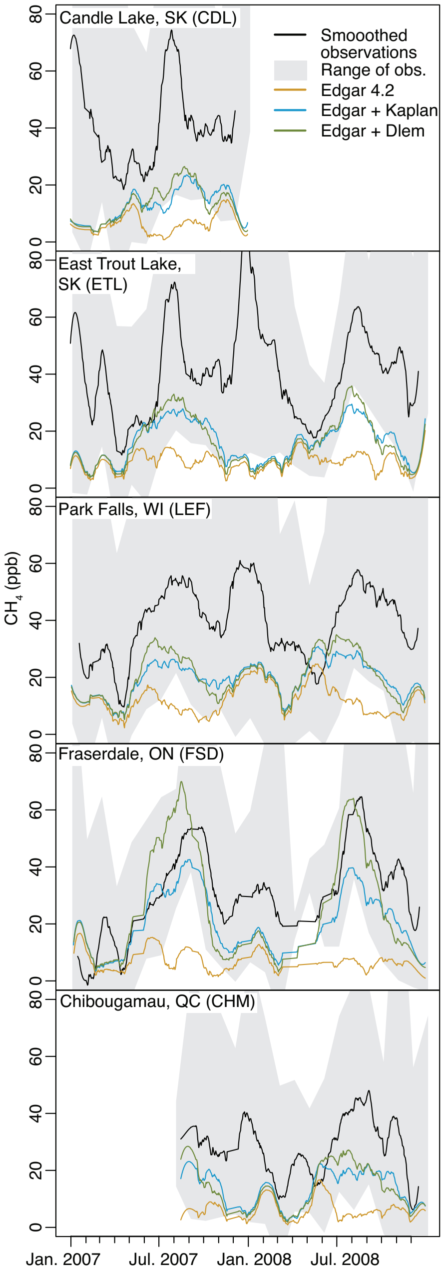


Figure 4.2: A comparison of modeled mixing ratios against measurements at the observation sites. The estimated boundary condition values have been subtracted from the observations; the difference indicates the effect of North American methane sources on the measurement sites. EDGAR v4.2 is an anthropogenic emissions inventory while Kaplan and DLEM model wetlands. The model and observations are smoothed using a 3rd order Savitzky-Golay filter with a 61-point window.

The model-data comparison in Fig. 4.2 also reveals important conclusions about the interdependence of wetland and anthropogenic emissions estimates. Gaps in anthropogenic emissions inventories can affect the perceived amplitude or seasonality of the wetland flux model. Even in remote regions like the HBL, the estimation of wetland fluxes hinges on a reliable anthropogenic emissions estimate. For example, the Kaplan/EDGAR v4.2 modeled concentrations are consistently too low at the Fraserdale site, but the amplitude of the summer maximum is similar to the amplitude of the data. This discrepancy could reflect one of two problems: either the wetland inventory has the incorrect magnitude and seasonal structure or the anthropogenic inventory (EDGAR v4.2) is simply too low. The time series at Park Falls (LEF), Wisconsin, further illustrates the importance of the anthropogenic emissions estimate. It appears that the wetland flux models begin producing methane too early in the spring of 2008 at LEF. A closer examination of Fig. 4.2, however, reveals large ($\sim$25ppb) modeled concentrations from anthropogenic sources during this period. This model-data discrepancy could stem from misspecified anthropogenic emissions, not problems in the seasonal structure of the wetland model. These examples highlight the difficulty of disentangling anthropogenic and wetland methane fluxes.

Subsequent sections discuss the deterministic model and geostatistical inversion results in greater detail.

4.4.2 Environmental predictors of wetland fluxes

This section explores the results of the deterministic model ($\mathbf{X}\beta$, section 4.3.1 and 4.3.2). As discussed in the methods sections, the deterministic model is analogous to a weighted multivariate regression. Model selection methods (like the BIC, section 4.3.2) play a crucial role in constructing this deterministic model; they select auxiliary datasets (Table 4.1) for the deterministic model that can best explain the atmospheric methane data. In this way, model selection provides a means to objectively understand and assess biogeochemical methane models at continental scale.

The BIC selection chooses the following deterministic model for methane fluxes in Canada (Table 4.2):

$\beta_0 + \beta_1[\text{smooth functions}] + \beta_2[W][M]f_{Kaplan}(T)$

This selected model for methane fluxes is relatively simple. The first term (β_0) is a constant component, equivalent to the intercept in a regression. It describes the average magnitude of all sources not explicitly included in other components of the deterministic model. For example, this component might include agriculture, landfills, and wastewater treatment sources (among other possibilities).

The second term (β_1[smooth functions ...]) parameterizes anthropogenic sources (Section 4.3.2). This term places smooth geometric functions over known source regions, including Alberta, California, Oklahoma, and the US east coast. The BIC does not choose the EDGAR v4.2 anthropogenic inventory for the deterministic model because it fits the atmospheric data less well than the smooth geometric functions (Table 4.2). Hence, we do not utilize EDGAR within the atmospheric inversion.

The final component of the deterministic model ($\beta_2[W][M]f_{Kaplan}(T)$) parameterizes wetland fluxes. This term in the deterministic model includes three auxiliary datasets: the distribution of wetlands (W), a map of unfrozen soil moisture (M), and an Arrhenius equation based upon soil temperature ($f_{Kaplan}(T)$). The optimal deterministic model uses the wetland map from the LPJ model and soil variables from NARR (at 10cm soil depth). All other possible combinations and interactions of the auxiliary variables in Table 4.1 produce higher BIC scores (Table 4.2). For example, we test wetland models that include soil carbon, environmental variables at different depths in the soil profile, and different estimates for wetland distribution. Furthermore, we test a deterministic model that uses the functional form of temperature and/or soil moisture from the

Table 4.2: BIC scores for a selection of candidate deterministic models. Note: We test all possible combinations and interactions of the auxiliary variables in Table 4.1 and display only a sample here. The table is intended to show the range of BIC scores for the best-scoring models and a few other notable models. The drift coefficients (β) scale the magnitude of the auxiliary data to match the methane observations. All models above use inputs from NARR (10cm depth) and LPJ, unless otherwise noted. $f_{Kaplan}(\cdots)$ refers to the functional form used in the Kaplan model $f_{DLEM}(\cdots)$ the functional form in DLEM.

Candidate model	BIC
$\beta_0 + \beta_1$[tricubic functions] $+ \beta_2[W][M]f_{Kaplan}(T)$	16725
$\beta_0 + \beta_1$[tricubic functions] $+ \beta_2[W][M_{Tot}]f_{Kaplan}(T)$	16728
$\beta_0 + \beta_1$[tricubic functions] $+ \beta_2[W][M]f_{Kaplan}(T)f_{Kaplan}(C)$	16729
$\beta_0 + \beta_1$[tricubic functions] $+ \beta_2$[full Kaplan model]	16735
$\beta_0 + \beta_1$[tricubic functions] $+ \beta_2[W][M]f_{Kaplan}(T)$ using NARR surface soil layer	16744
$\beta_0 + \beta_1$[tricubic functions] $+ \beta_2[W]f_{DLEM}(M)f_{DLEM}(T)$	16750
$\beta_0 + \beta_1$[EDGAR v4.2] $+ \beta_2[W][M]f_{Kaplan}(T)$	16885

DLEM model.

This selected wetland model is similar to the Kaplan flux model but with soil carbon removed. Section 4.4.4 synthesizes the wetland flux results from this study and highlights what this parameterized wetland model might indicate about biogeochemical methane modeling.

Table 4.3 lists the Canadian methane budget associated with each component of the deterministic model and compares these estimates against existing inventories. It is important to remember that the methane budgets from the deterministic model are estimated using the atmospheric data – via the unknown coefficients, β. The smooth functions represent the largest component of the deterministic model, followed by the constant component and finally the wetland component. When interpreting these budgets, however, it is important to note that the constant component $\hat{\beta}_0$ could represent either anthropogenic emissions or wetland fluxes.

Figures 4.3 and 4.4 visualize the deterministic model, both spatially and in relation to the atmospheric methane data. Figures 4.3 displays the annual average of the deterministic model. The smooth geometric functions to parameterize anthropogenic emissions are evident over the province of Alberta and over the Dakotas. The wetland model is more difficult to distinguish

Table 4.3: Canada methane budgets from the deterministic model (south of 65°N) and several inventory estimates

Flux model	Canada budget (Tg C yr^{-1})
Deterministic model	
$\hat{\beta}_0$	5.4 ± 1.5
$\hat{\beta}_1$[smooth functions]	7.9 ± 0.9
$\hat{\beta}_2[W][M]f_{Kaplan}(T)$	3.2 ± 0.6
Existing wetland models	
Kaplan model	4.4
DLEM model	5.6
Existing anthropogenic inventories	
Environment Canada	3.3
EDGAR v4.2	3.9

in this annual mean plot but is largest south of Hudson Bay in eastern Canada and near Great Slave Lake in Northwest Territories. The deterministic model is non-zero everywhere across Canada, and this reflects the constant term β_0 of the deterministic model. This term has an estimated magnitude of $2 \pm 0.5 \times 10^{-3} \mu\text{mol m}^{-2}\text{ s}^{-1}$ (5.4 ± 1.5 TgC yr^{-1} over Canada, Table 4.3).

Despite the simplicity of the deterministic model, the mixing ratios estimated with this model match favorably against atmospheric measurements (Fig. 4.4). The deterministic model fits the atmospheric methane observations ($R = 0.72$, root mean squared error (RMSE) = 20.9 ppb) better than either the model setup with Kaplan and EDGAR v4.2 ($R = 0.12$, RMSE=37.1 ppb) or DLEM and EDGAR v4.2 ($R = 0.08$, RMSE=37.2 ppb). The formulation of anthropogenic emissions in the deterministic model may account for much of this improved fit against the atmospheric data. Despite the improvement, the deterministic model displays two notable shortfalls. First, the deterministic model does not reproduce the summer maxima observed at western observation sites (CDL and ETL). Second, the deterministic model underestimates the summer maxima at the

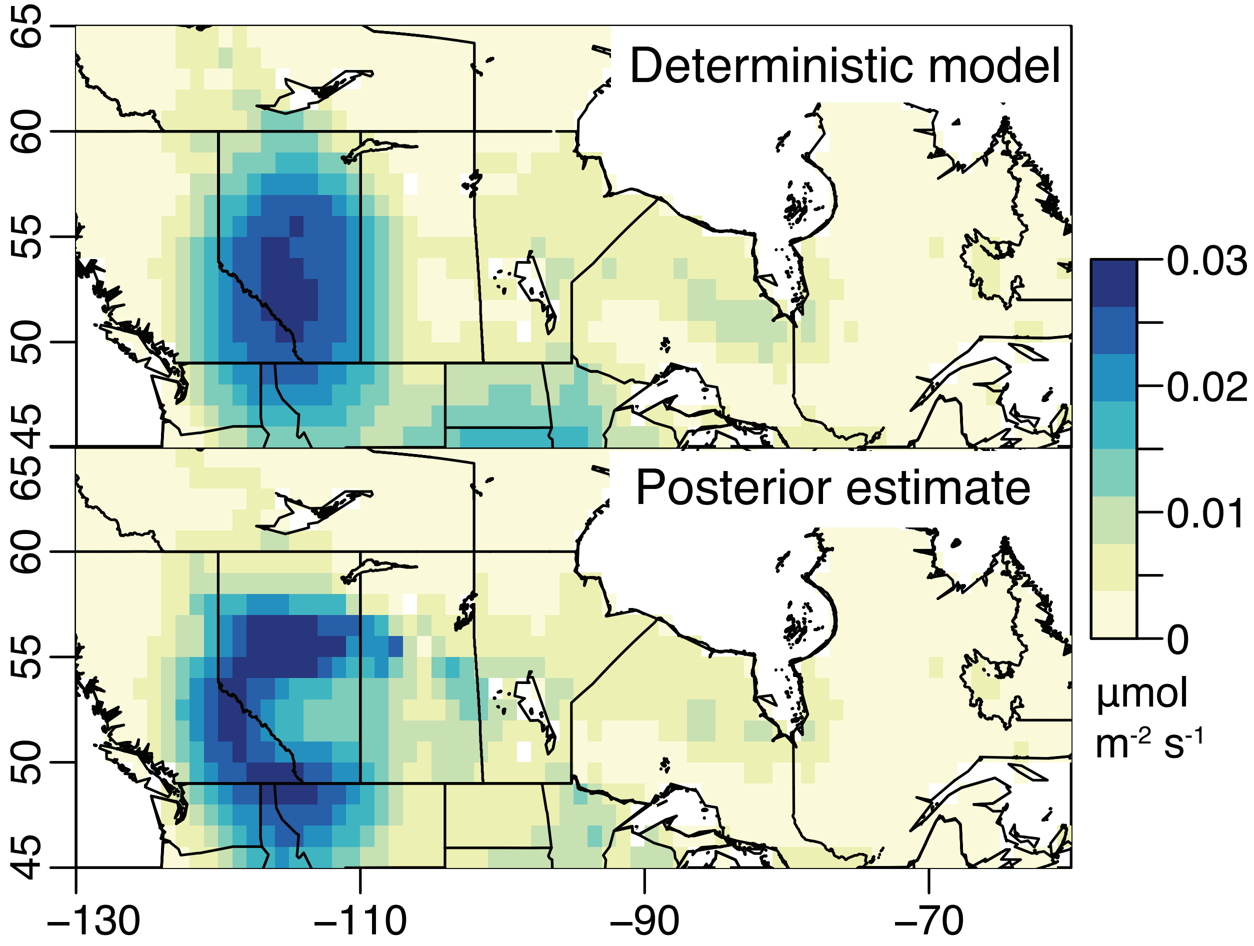


Figure 4.3: The 24-month (2 year) mean estimated methane flux from the deterministic model (top) and the final posterior estimate (bottom).

Wisconsin (LEF) and Quebec (CHM) observation sites. These shortfalls suggest that the spatial distribution of wetland fluxes in the deterministic model may be too restrictive. In other words, wetland fluxes likely extend further west, east, and south than in the deterministic model, which places the largest wetland fluxes in the HBL.

Subsequent sections discuss the final methane flux estimate from the geostatistical inversion ($\hat{s}$). This final, best estimate ($R = 0.89$, RMSE=12.0 ppb) is henceforth referred to as the 'posterior' fluxes.

4.4.3 The spatial and temporal distribution of emissions

The posterior flux estimate identifies two major source regions in Canada (Fig. 4.3): over Alberta in western Canada and over the HBL in eastern Canada. This discussion analyzes each geographic region individually.

In western Canada, the inversion identifies a large, seasonally-constant methane source region over Alberta. In the deterministic model, this source is represented by a smooth function. But in the posterior estimate, this source region becomes a more well-defined crescent shape over Alberta (Fig. 4.3). These emissions likely originate from anthropogenic activity, and a future study will give an in-depth analysis of anthropogenic emissions in Canada. The posterior flux estimate also includes a large summer source in Alberta and Saskatchewan. As discussed previously, these fluxes are not represented by the auxiliary environmental datasets in the deterministic model. This omission in western Canada dominates the discrepancy in summertime Canadian methane between the deterministic model and posterior fluxes (Fig. 4.5). The omission implies that either the LPJ wetland or the NARR soil moisture map is an underestimate in westerly regions of Canada. Unfortunately, the atmospheric data in this region has limited capability to pinpoint the exact location of these western wetland fluxes; atmospheric observations are sparse in western Canada, and wetland fluxes are co-located with large anthropogenic sources.

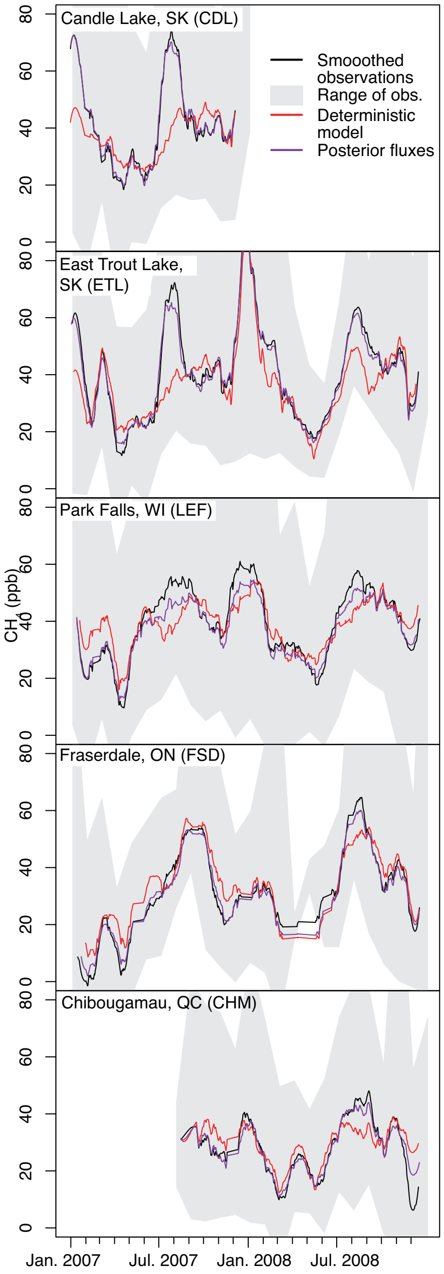


Figure 4.4: A comparison of modeled mixing ratios against measurements at the observation sites. This figure is similar to Fig. 4.2 but compares the deterministic model and posterior emissions estimate instead of existing flux models.

In sum, this study identifies Alberta as a region with poorly-known wetland fluxes and as a possible hot-spot of anthropogenic emissions. We recommend that future methane measurement efforts focus on Alberta because this province is a key uncertainty in current understanding of Canadian methane sources.

Eastern Canada, in contrast, is dominated by seasonal methane fluxes that presumably emanate from wetlands. Figure 4.5 compares the seasonal cycle of DLEM, Kaplan, the deterministic model, and the posterior flux estimate over the HBL. The seasonal cycle of the deterministic model and posterior flux estimate compare similarly to the Kaplan model but have a broader seasonality than DLEM. The posterior flux estimate matches the Kaplan model more closely than the deterministic model over the HBL (though the deterministic model is a better match than Kaplan/EDGAR v4.2 in other regions of Canada and the northern US).

Seasonal structure aside, the flux models also diverge in spatial distribution. Figure 4.6 displays the mean summer (July, August, September) methane flux estimated by the inversion for eastern Canada. It also displays the difference between this estimate and the DLEM and Kaplan models. Our flux estimate is more spatially-dispersive than DLEM across the Hudson Bay region. The differences between the posterior estimate and Kaplan are more subtle. The posterior estimate indicates methane fluxes across a broader region than Kaplan: into Minnesota, Wisconsin, Manitoba, and further west.

Figure 4.7 summarizes the findings of this study as an annual methane budget estimate for the HBL and for all of Canada (south of 65° latitude). Our methane estimate for Canada is a factor of 1.5 to 2.2 times existing estimates. Anthropogenic emissions in western Canada may explain much of this discrepancy. In contrast, our annual HBL budget is consistent with that of DLEM and Pickett-Heaps et al. (2011) who use the Kaplan wetland model, but our estimate diverges from a site-based study by Roulet et al. (1992) and a box model study by Worthy et al. (2000) (see the supplement). Furthermore, the HBL budget estimated here is low compared to the array of

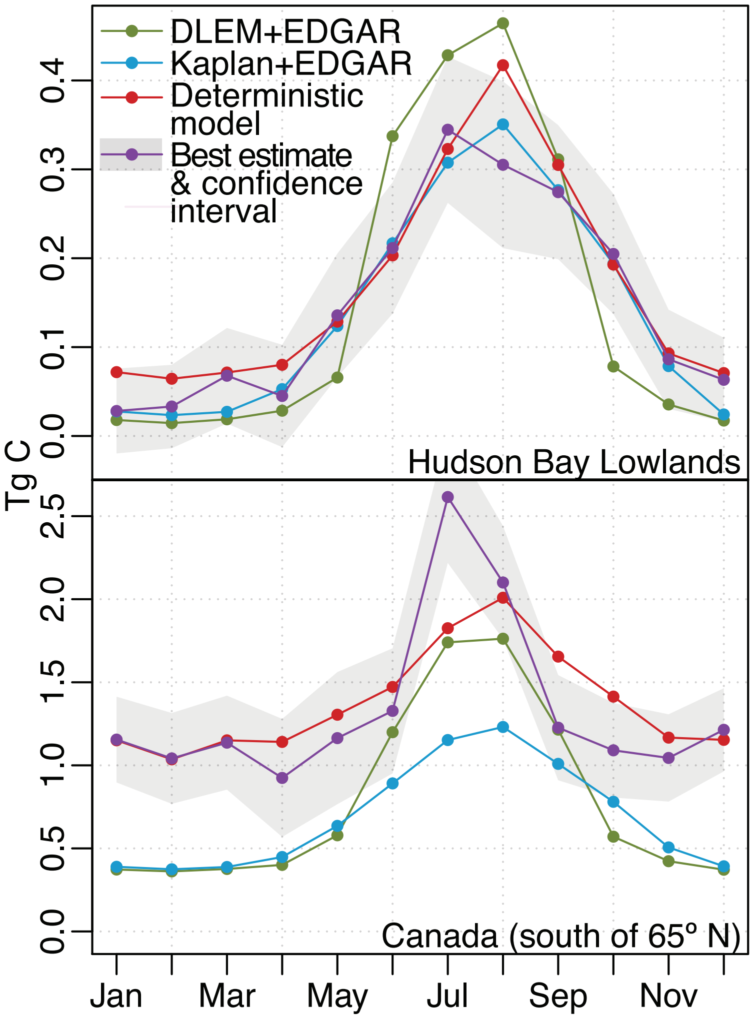


Figure 4.5: The monthly average methane budget estimated for the HBL and Canada in 2007 – 2008. Existing models underestimate wetland fluxes in western Canada. This regional shortfall explains much of the summertime discrepancy between the flux models and the posterior estimate in the lower panel.

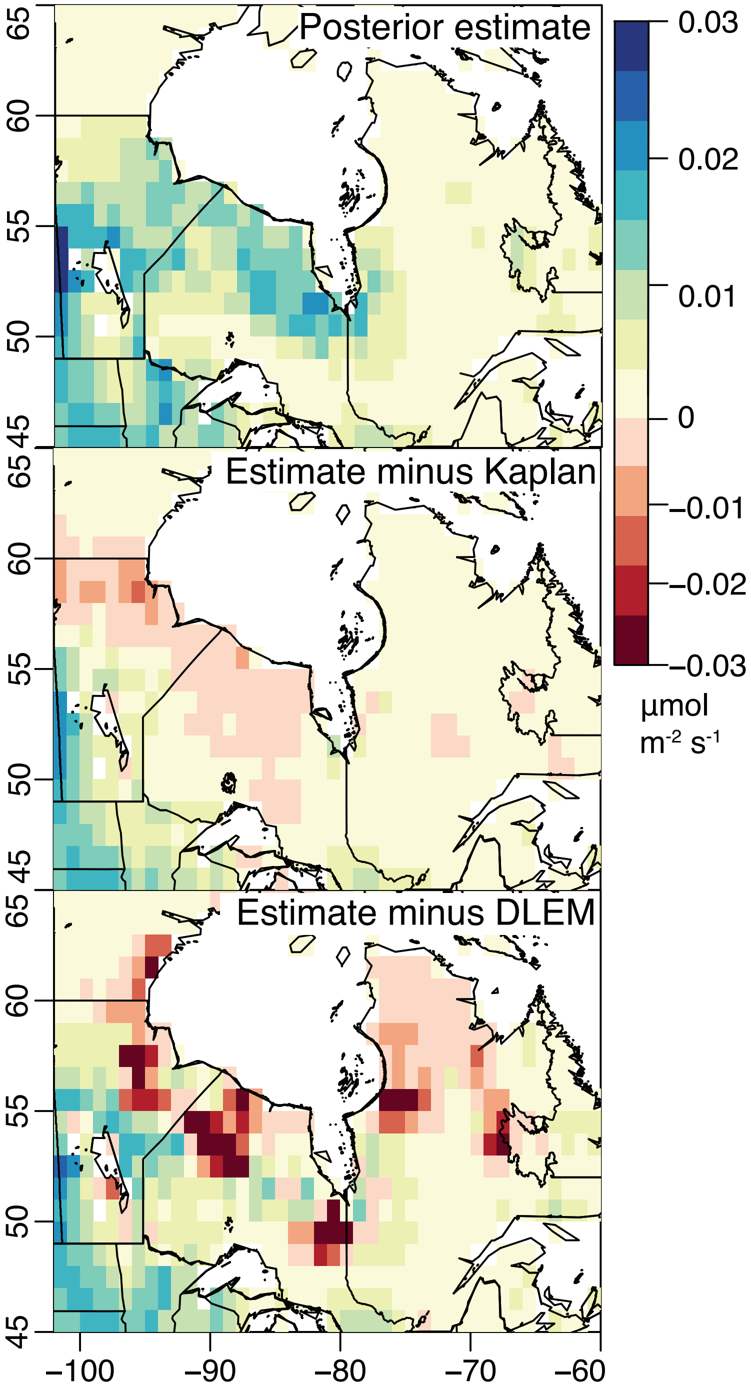


Figure 4.6: The top panel displays the posterior flux estimate averaged over all summer months (July - September for 2007–2008). Bottom panels illustrate the difference between the posterior estimate and the DLEM and Kaplan methane models.

biogeochemical models listed in Melton et al. (2013a). The HBL budgets in those models range from 1.7 – 8.5 TgC yr^{-1}. This range in wetland methane estimates in is likely greater than the inter-annual variability in wetland fluxes. For example, Tian et al. (2010) estimate an 11% standard deviation in annual North American methane fluxes.

4.4.4 A synthesis perspective on biogeochemical methane models

This section explores the study's implications for biogeochemical methane modeling. The inversion results (e.g., section 4.4.2) raise the question of why a simple flux model fits the atmospheric methane data as well as sophisticated process models. The deterministic model developed here excludes a number of factors that can affect methane fluxes: soil carbon, plant-mediated transport, and heterogeneities in microbial communities, among many other processes. This question could be answered in two ways.

First, simple parameterizations may be sufficient when regional-scale flux patterns are the primary goal. For example, a synthesis study of existing chamber measurement sites found that methane fluxes across all sites are influenced most strongly by only a few environmental variables: water table height, soil temperature, and vegetation type (Olefeldt et al., 2013). Furthermore, Bubier et al. (1993) and Waddington & Roulet (1996) argue that most centimeter-scale flux variability ultimately depends on two primary parameters: temperature and water table position. These studies imply that a simple model may adequately parameterize regional-scale flux variability.

A second reason may account for the simplicity of the deterministic model. Complex methane flux processes can be challenging to upscale, meaning that the most complete methane model is not always the most accurate at regional scales. The spatial distribution of many flux-related processes is highly uncertain (e.g. Melton et al., 2013a) due to a paucity of both land surface and methane flux data. This uncertainty means that models with many processes and parame-

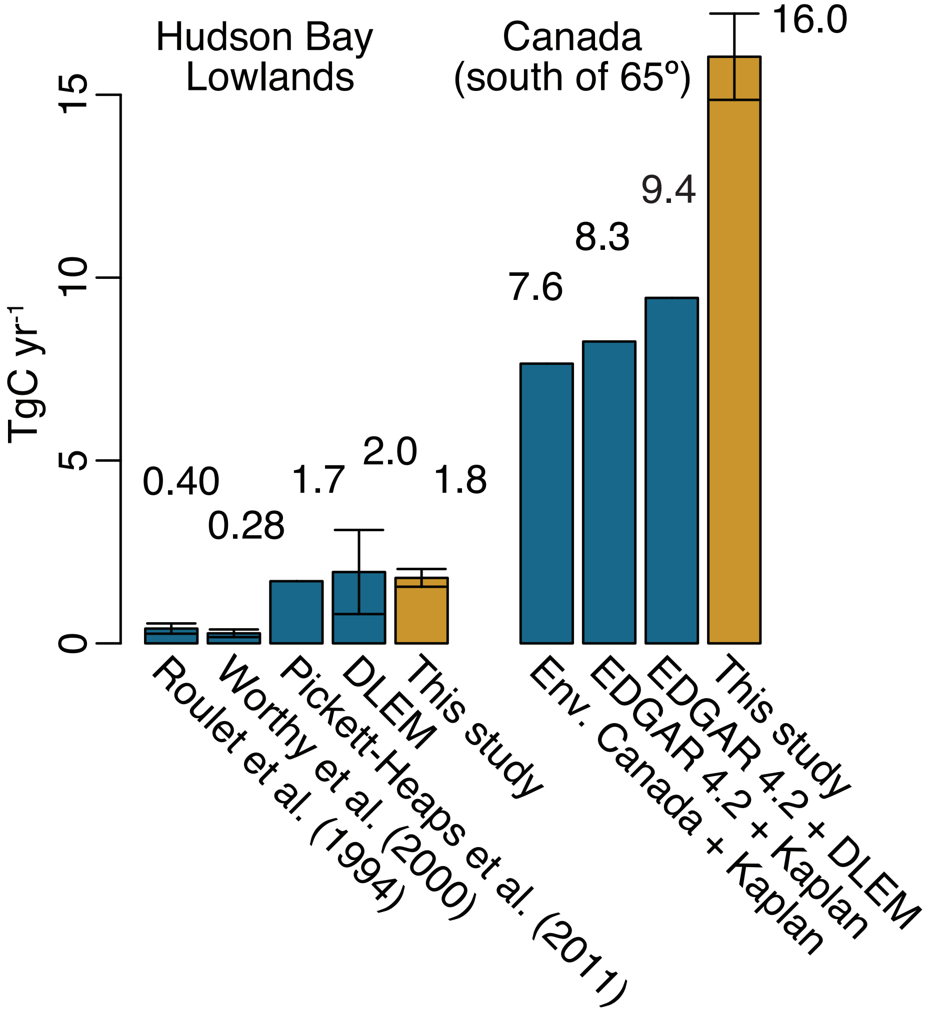


Figure 4.7: Total methane budgets from this study and others for the HBL and for Canada. The HBL estimates listed here are from DLEM and from observational studies. Melton et al. (2013a) list numerous additional model-based HBL methane budgets, which range from 1.7 – 8.5 TgC yr^{-1}. 'Env. Canada' refers to Environment Canada's National Inventory Report [2013]. The confidence intervals for this study do not encompass uncertainties in model selection and therefore may underestimate the total budget uncertainty.

ters could run the risk of over-fitting limited, available data (A paper by Zucchini (2000) illustrates the hazards of over-fitting.). For example, a number of physical processes like ebullition, plant-mediated transport, and microbial community dynamics are all thought to play a role in methane emissions (Bridgham et al., 2013, and references therein), but how these processes or features vary on regional spatial scales is often poorly understood.

To that end, model selection methods, like the BIC used here, provide a means to diagnose weaknesses in flux model upscaling from plot-level to regional or continental scale (e.g. Olefeldt et al., 2013). Model selection methods choose the set of predictors that can best explain variability in any available methane data. If model selection does not choose a given predictor, that outcome implies one of several conclusions. Either the distribution of the predictor does not match against the distribution implied by the methane data, or the available methane data is insufficient to constrain the effect of that predictor. In either case, any conclusions based upon the predictor would likely over-fit the available data at the expense of describing the large-scale flux process of interest.

4.5 Conclusions

This study uses atmospheric methane observations and geostatistical inverse modeling to understand North American boreal methane fluxes and associated biogeochemical models. The conclusions of this study fall under three general themes. First, we find that a simple wetland flux model, when combined with WRF-STILT, provides as good agreement with atmospheric methane observations as more complex flux process models. This result may have several possible causes: either simple models adequately parameterize regional-scale flux patterns, or the spatiotemporal distribution of important but complex flux processes is difficult to accurately model with available data at this geographic scale.

Second, we estimate both the spatial and seasonal distribution of methane fluxes over much of boreal North America. We find wetland fluxes that are more broadly distributed than in existing inventories, even extending into Minnesota, Wisconsin, Manitoba, and western Canada. This result implies that existing maps may under-represent the extent of soil moisture and/or the distribution of wetlands.

Finally, we calculate regional and Canadian methane budgets. Our HBL budget is on the upper range of observational studies but the lower range of biogeochemical model estimates (Fig. 4.7 and Melton et al. (2013a)). In addition, we estimate total Canadian emissions that exceed existing inventories, largely due to sources in or near Alberta. Available atmospheric data is limited near Alberta during the study period, and this work highlights a need for more intensive methane measurements over that region.

5

The ability of atmospheric data to resolve discrepancies in wetland methane estimates over North America

This chapter has been submitted for publication in the journal Biogeosciences Discussions.

Existing estimates of methane fluxes from North American wetlands vary widely in both magnitude and distribution. In light of these disagreements, this study uses atmospheric methane observations from the US and Canada to analyze seven different bottom-up, wetland methane estimates reported in a recent model comparison project. We first use synthetic data to explore how well atmospheric observations can constrain wetland fluxes. We find that observation sites can identify an atmospheric pattern from Canadian wetlands but not reliably from US wetlands. The network can also identify the spatial distribution of fluxes in Canada at multi-province spatial scales. Based upon these results, we then use real data to evaluate the magnitude, temporal distribution, and spatial distribution of each model estimate. Most models overestimate the magnitude of fluxes across Canada. Most predict a seasonality that is too narrow, potentially indicating an over-sensitivity to air or soil temperatures. In addition, the LPJ-Bern and SDGVM mod-

els have a spatial distribution that is most consistent with atmospheric observations, depending upon the season and region. Unlike most models, LPJ-Bern and SDGVM utilizes land cover maps, not just remote sensing inundation data, to estimate wetland coverage. A flux model with a constant spatial distribution outperforms all other existing flux estimates across Canada.

5.1 Introduction

Methane fluxes from wetlands play a critical role in global climate change. Methane is the second-most important long-lived greenhouse gas; the radiative forcing of the current atmospheric burden is approximately 26% of carbon dioxide. Wetlands are possibly the largest single source of this gas to the atmosphere and account for roughly 30% of global emissions (Ciais et al., 2013).

Despite the important role of wetland methane fluxes in climate change, existing estimates of this source disagree markedly on the magnitude, seasonality, and spatial distribution of fluxes, from regional to global scales. In fact, a recent global model comparison project named WETCHIMP (Wetland and Wetland CH_4 Inter-comparison of Models Project) found large discrepancies among existing methane wetland models (Fig. 5.1, Melton et al., 2013a; Wania et al., 2013). For example, existing estimates of maximum global wetland coverage differ by over a factor of 6 – from 4.1×10^6 to 26.9×10^6 km^2. Furthermore, estimates of global natural wetland fluxes range from 92–264 Tg CH_4 yr^{-1}. The relative magnitude of these uncertainties increases at sub-global spatial scales. As a case in point, methane estimates for Canada's Hudson Bay Lowlands (HBL) range from 0.2 to 11.3 Tg CH_4 yr^{-1}. These disagreements in current methane estimates do not bode well for scientists' abilities to accurately predict future changes in wetland fluxes due to climate change (Melton et al., 2013a). A number of studies have used chamber measurements of methane to parameterize or evaluate biogeochemical methane models (e.g., Livingston & Hutchinson, 2009). However, these measurements usually encompass fluxes from a very small

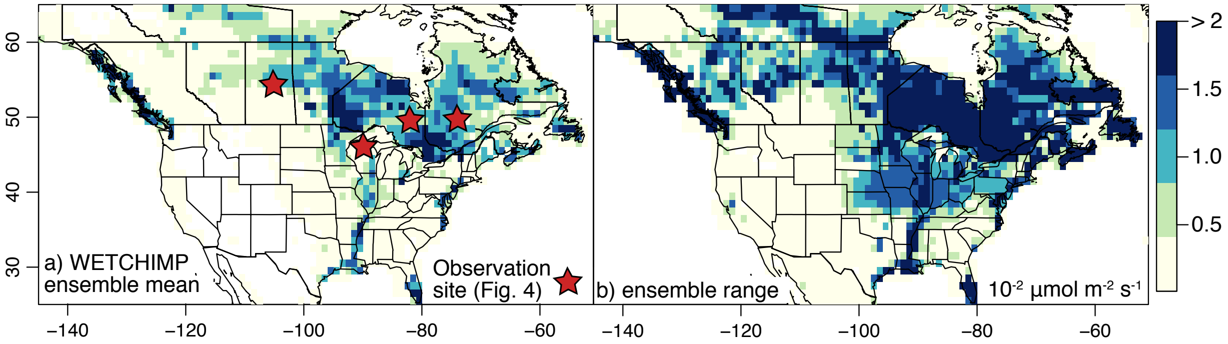


Figure 5.1: Mean of the annual methane fluxes estimated by the WETCHIMP models (a) and the range of fluxes estimated by the ensemble (b). Note that the range in estimates is larger than the mean. The fluxes shown above are averaged over an entire grid cell, not per m^2 of wetlands.

spatial scale, and fluxes can vary by an order of magnitude over one meter or less (Waddington & Roulet, 1996; Hendriks et al., 2010). Methane data collected in the atmosphere, by contrast, sees the cumulative effect of methane fluxes across a much broader region (e.g., Kort et al., 2008; Pickett-Heaps et al., 2011; Miller et al., 2014c). Hence, atmospheric data can provide an important tool for evaluating existing methane flux estimates across different countries or continents.

The present study compares the WETCHIMP methane flux estimates against atmospheric methane data from 2007–2008 through two sets of analyses. First, we construct progressively demanding synthetic data experiments to explore how well available data can constrain wetland fluxes. Can the atmospheric data identify methane patterns from wetlands over distracting patterns in the atmosphere? These patterns include methane from anthropogenic sources or random noise due to model and measurement errors. If yes, can the observation sites detect spatial variability in the wetland fluxes? We seek to understand whether large uncertainties in wetland methane estimates point to a paucity of methane data – data capable of calibrating or evaluating the models. In the alternative, perhaps these disagreements would be much smaller if existing biogeochemical models leveraged all available data. To answer these questions, we utilize a modeling approach based upon the Bayesian Information Criterion (BIC), described in greater detail in Sect. 5.2.2 (Shiga et al., 2014; Fang et al., 2014).

Based on the synthetic experiments, we conduct a second set of analyses using real atmospheric data. We use this data to evaluate the magnitude, seasonal cycle, and spatial distribution of each WETCHIMP methane estimate. Of the seven available models, which have a magnitude, seasonal cycle, or spatial distribution that is most consistent with the available data? We investigate this question over the United States and Canada using methane data collected from towers and regular aircraft flights operated by NOAA and its partners and from towers operated by Environment Canada.

5.2 Methods

This section first describes the atmospheric methane data and the atmospheric model that allows direct comparison between the data and various flux estimates. Subsequent sections describe how we use these tools to construct both the synthetic and real data experiments outlined in the introduction (Sect. 5.1).

5.2.1 Data and atmospheric model

The present study utilizes atmospheric methane observations at Environment Canada and NOAA observation sites (Fig. 5.2). These include regular measurements from tower and aircraft platforms, a total of 14,703 observations from 2007-2008. The observations used here are identical to those in Miller et al. (2013) and Miller et al. (2014c).

We then employ an atmospheric transport model to relate methane fluxes at the Earth's surface to atmospheric concentrations at the observation sites. The modeling approach here combines the Weather Research and Forecasting (WRF) meteorological model and a particle-following model known as STILT, the Stochastic Time-Inverted Lagrangian Transport model (e.g., Lin

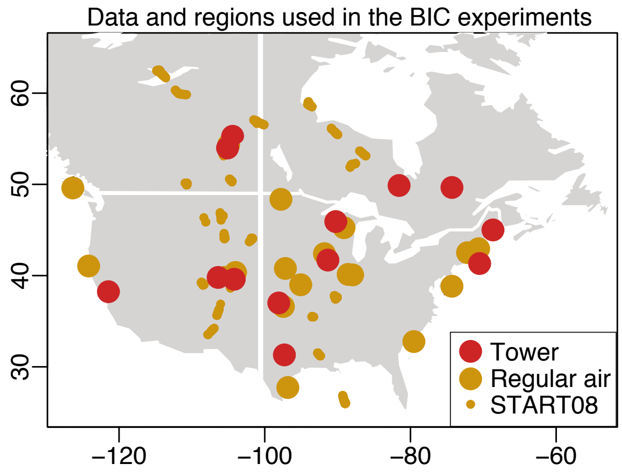


Figure 5.2: The NOAA and Environment Canada atmospheric methane observation network for 2007–2008 (14,703 total observations). Small yellow dots indicate observations from the START08 measurement campaign (Pan et al., 2010). Larger dots indicate tower and aircraft sites with regular observations over the two year period (Andrews et al., 2014). The grey background delineates the four regions used in the synthetic experiments (sect. 5.2.3).

et al., 2003; Nehrkorn et al., 2010; Hegarty et al., 2013). WRF-STILT generates a set of footprints; these footprints quantitatively estimate the sensitivity of each observation to fluxes at each surface location (with units of ppb per unit surface flux). We multiply the footprints by a flux model and add this product to an estimate of the 'background' concentration – the methane concentration of air entering the North American regional domain. The resulting modeled concentrations can be compared directly against atmospheric methane observations. This modeling setup is identical to Miller et al. (2013) and Miller et al. (2014c). Both the observations and the WRF-STILT model are described in greater detail in those papers and in the supplement.

Using this setup, we can compare predicted methane concentrations using the WETCHIMP flux estimates (Fig. 5.1) against observed atmospheric concentrations. Of the WETCHIMP models, seven provide a flux estimate for boreal North America and six provide an estimate for temperate North America. These models include CLM4Me (Riley et al., 2011), DLEM (Tian et al., 2010), LPJ-Bern (Spahni et al., 2011), LPJ-WHyMe (Wania et al., 2010), LPJ-WSL (Hodson et al., 2011), ORCHIDEE (Ringeval et al., 2010), and SDGVM (Singarayer et al., 2011). All model outputs have a temporal resolution of one month. These models are described in Melton et al. (2013a), Wania et al. (2013), and the Supplement.

5.2.2 Model selection framework

This study employs two synthetic data experiments to explore the sensitivity of atmospheric observations to wetland fluxes: can the observations detect an atmospheric pattern from wetlands fluxes over distracting patterns from anthropogenic emitters? If yes, can the observations detect spatial variability in wetland fluxes from different regions? We build a modeling approach based upon the BIC to answer these questions.

The BIC is a model selection technique, and various forms of the BIC are used widely in statistical regression analysis (e.g., Schwarz, 1978; Ramsey & Schafer, 2012). It scores all possible combinations of explanatory variables based on model-data fit, and it penalizes combinations that have a greater number of variables. The best combination or candidate model has the lowest BIC score.

We use a form of the BIC that has been adapted for use within a geostatistical inverse modeling framework (e.g., Gourdji et al., 2008; Miller et al., 2014c). The implementation here parallels that of Fang et al. (2014) and Shiga et al. (2014):

$$\mathrm{BIC} \;=\; \underbrace{\ln|\boldsymbol{\Psi}| + (z - \mathbf{HX}\beta)^T \boldsymbol{\Psi}^{-1} (z - \mathbf{HX}\beta)}_{\text{negative log}-\text{likelihood}} + \underbrace{p\;\ln(n)}_{\text{penalty term}} \tag{5.1}$$

The first term in Eq. 5.1 is the negative log-likelihood, a measure of how well the model fits the data. In that term, z ($n \times 1$) represents the observations minus background concentrations, $\mathbf{H}$ ($n \times m$) the footprints, $\mathbf{X}$ ($m \times p$) a matrix of p explanatory variables, β ($p \times 1$) a set of coefficients assigned to those variables, and $\boldsymbol{\Psi}$ ($n \times n$) a covariance matrix derived from an atmospheric inversion framework. The data (z), footprints ($\mathbf{H}$), and parameters that define the covariance matrix ($\mathbf{\Psi}$) are taken from Miller et al. (2013) and Miller et al. (2014c) (refer to the Supplement). The second term in Eq. 5.1 penalizes the BIC score of a particular model based upon the number of explanatory variables (p).

We employ this model selection framework to understand which explanatory variables from an anthropogenic emissions inventory and from the WETCHIMP ensemble are required to describe either synthetic or real methane data at North American observation sites.

5.2.3 Synthetic data experiments

The experiments described in this section use synthetic data generated at each of the observation sites. We use anthropogenic emissions estimates for the US and Canada from Miller et al. (2013) and Miller et al. (2014c), respectively, and use one of the WETCHIMP models as the wetland flux estimate. We then multiply these fluxes by $\mathbf{H}$ to create the synthetic data at the measurement locations. We further add in randomly-generated error – error that represents uncertainties in the fluxes, the measurements, and the atmospheric transport model, among other error sources. These errors are estimated by Miller et al. (2013) and Miller et al. (2014c) and are described in the Supplement.

The synthetic experiments ask progressively demanding questions that test the limits of available data. In experiment one, we examine whether methane observations in different regions of North America can detect methane patterns in the atmosphere due to wetland fluxes. When given multiple possible explanatory variables (including data from the EDGAR anthropogenic emissions inventory), will the model selection framework choose a wetland estimate? If yes, the observations can identify a pattern in atmospheric methane due to wetland fluxes and that pattern is large enough to be visible over other signals in the atmosphere. If not, then either the contribution of wetlands at that site is small, or the observations cannot differentiate atmospheric patterns due to wetlands over other atmospheric patterns due to anthropogenic sources or model-measurement errors. This setup follows Shiga et al. (2014), who explored the detectability of atmospheric patterns from anthropogenic CO_2 emissions.

For this test, we generate the synthetic data using one of the WETCHIMP models. We then allow the model selection framework to select wetland fluxes and/or the EDGAR data used to generate the synthetic fluxes. We divide the wetland fluxes into four regions (Fig. 5.2) and four seasons (winter, spring, summer, fall). The model selection can choose none, some, or all of these sixteen wetland variables. We run this experiment 1000 times, generating new synthetic data each time, and calculate the percentage of all trials in which the model selection chooses a wetland model. In this experiment, the coefficients (β) are fixed to one. Note that several of the WETCHIMP models overestimate the magnitude of fluxes (Sect. 5.4.2), so we only use models with a smaller magnitude to generate the synthetic data in this experiment (CLM4Me, DLEM, SDGVM, and LPJ-WSL).

In experiment two, we investigate whether the observation network is sensitive to spatial variability in the wetland fluxes, independent of magnitude or seasonality. In this setup, we do not fix the coefficients (β) but rather estimate coefficients that minimize the log-likelihood in Eq. 5.1. We also include a spatial constant or intercept term in $\mathbf{X}$ that can change by month. As a result of this setup, the magnitude and seasonality of the intercept can be adjusted to match the data, but any spatial variability in the fluxes can only come from the wetland model. As in experiment one, the model selection framework can choose among 16 wetland variables – variables that represent different regions and seasons. If model selection chooses a wetland variable, then the spatial distribution in that variable is necessary to reproduce the synthetic data. If not, then the observations are not sensitive to spatial variability in wetland fluxes for that region/season. This approach follows that of Fang et al. (2014), who employed a model selection framework to evaluate the spatial distribution of biospheric CO_2 flux models.

5.2.4 Real data experiments

If experiment two is successful on synthetic data, we then apply the experiment to real data. We use the model selection framework to determine which, if any, bottom-up models have a spatial distribution that can describe the methane observations more effectively than a spatial constant.

We also include a number of model-data time series to evaluate both the magnitude and seasonality of the fluxes. We model methane concentrations at a number of US and Canadian observation sites using WRF-STILT, WETCHIMP, and EDGAR v4.2FT2010 (Olivier & Janssens-Maenhout, 2012; European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL), 2013). We average the observations and model output at the monthly scale and then compare the magnitude of these model estimates for each month against the averaged observations.

Several studies indicate that EDGAR may underestimate emissions in certain regions of the US and Canada (e.g., Kort et al., 2008; Miller et al., 2013, 2014c; Wecht et al., 2014). Therefore, we scale the magnitude of EDGAR v4.2FT2010 to match wintertime observations (November–April) at each site using a standard major axis (SMA) regression. During those months, fluxes from wetlands are small and any model biases are likely due to anthropogenic emissions. We then apply this scaling factor, estimated for each site from winter data, to anthropogenic emissions in all seasons. Miller et al. (2013) found that anthropogenic emissions in the US lack significant seasonality, so the wintertime scaling factors should be applicable to other seasons.

We further compare the seasonality of existing bottom-up models against the seasonality of a recent inverse modeling estimate by Miller et al. (2014c). We plot the monthly methane budget as a fraction of the annual total for both the bottom-up models and the inversion estimate. We only conduct this analysis for wetland flux regions that are visible to the observation network (synthetic experiments one and two).

Note that inter-annual variability in existing methane flux models is small relative to the differences among these models; as a result, conclusions from the 2-year study period (2007–2008) likely hold for other years. For example, the inter-annual variability in the total US/Canadian budget is $\pm 7.3 - 9.7\%$ (standard deviation), depending upon the model in question (Note that LPJ-Bern has even larger inter-annual variation due to an issue with model spin-up (Wania et al., 2013).).

5.3 Results and discussion: synthetic experiments

The synthetic experiments presented here explore the limits of existing atmospheric data for constraining wetland fluxes. We first leverage synthetic data to examine whether the atmospheric observation sites can distinguish an atmospheric pattern from wetland fluxes above other patterns due anthropogenic emissions or simulated model, measurement, and emissions uncertainties. If atmospheric observations are to constrain wetland methane fluxes, those observations must, at minimum, identify an atmospheric pattern from wetland fluxes from other distracting patterns in the model and/or data.

The results of this experiment are summarized in Fig. 5.3a. The four columns in Fig. 5.3a display the results from an individual season in each of four geographic regions. In this experiment, the observation network can detect a summertime methane pattern from wetlands in both Eastern and Western Canada in greater than 75% of all trials. In the eastern US, the model selection framework chooses a wetland model in 50–75% of all trials in multiple different seasons. By contrast, the observations are least sensitive to wetland fluxes in the western US, and the model selection framework chooses wetland fluxes from that region in less than 25% of all trials irrespective of the season. This result may be due, in part, to the relatively dry climate and scant wetlands in much of the west. The methane signal from resource extraction and/or agriculture may

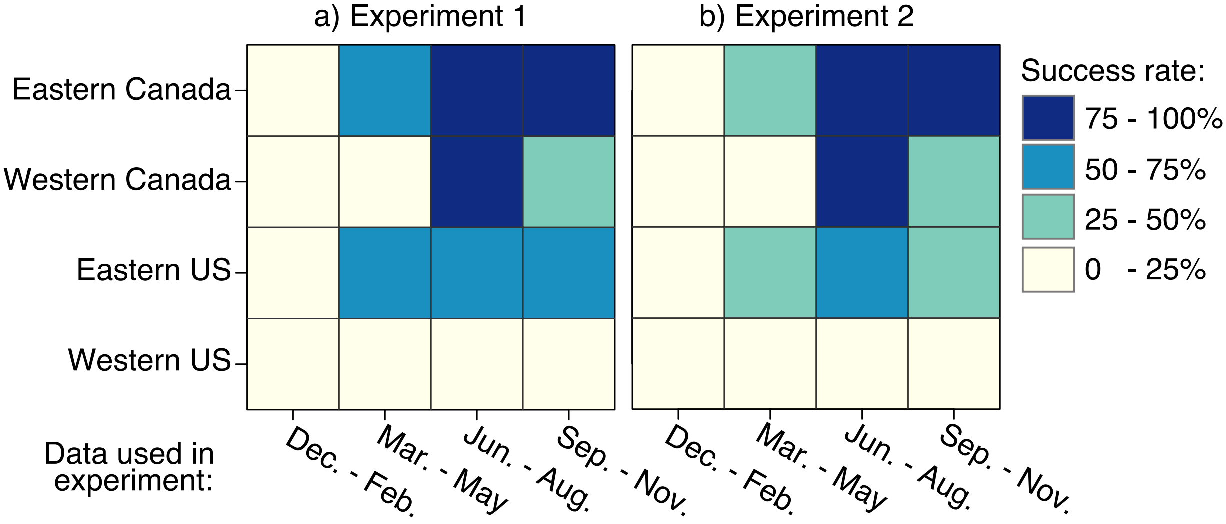


Figure 5.3: This figure displays the results of the synthetic data experiments. These experiments examine whether the observation network can (a) identify a methane pattern from wetland fluxes and (b) identify spatial variability in the wetland fluxes. The figure shows the percentage of trials that are successful. Darker shades indicate that the network is more sensitive to fluxes in the given region and season.

also overshadow any patterns from wetlands.

The results also contain a number of seasonal trends. Of any region, the observation network is best able to constrain the seasonal cycle in eastern Canada. The largest wetland fluxes estimated for the US and Canada are in Ontario and Quebec. It is therefore unsurprising that the network is so sensitive to fluxes from this region, even though there are relatively few observation sites within the region. In other regions, the observation network is less sensitive to wetlands during the winter, fall, and spring seasons. For example, the model selection framework chooses a wetland model in less than 25% of all trials during the winter in all regions.

The density of the observation network may also play a role in these results. Wetlands in the Eastern US are sparse relative to Canada, but the higher density of observations in the Eastern US may contribute to a relatively high success rate for that region. A recent observation network expansion could play a key role in future efforts to constrain wetland fluxes across these regions.

Environment Canada has recently been expanding their observation network across western and Arctic Canada (i.e., Saskatchewan, Alberta, Northwest Territories, and Nunavut). In addition, Earth Networks has begun to install new observation sites across the eastern US in a privately-funded initiative.

Compared to experiment one, the second experiment asks a more demanding question of the observation network: is the observation network sensitive to spatial variability in the wetland fluxes from each region? Alternately, can a spatially-constant model reproduce the synthetic atmospheric observations as well? Existing bottom-up estimates disagree markedly on the spatial distribution of wetland fluxes, but perhaps atmospheric data can provide guidance.

Figure 5.3b displays the results of this experiment for each region and season. The available data is only sensitive to spatial variability in certain cases. The model selection framework chooses a wetland model in $> 75\%$ of all trials in eastern Canada during summer and fall and in western Canada during summer. Eastern Canada is the only region that can distinguish different spatial patterns in the shoulder seasons. In remote regions of northern Ontario and Quebec, large wetland fluxes dominate variability in atmospheric methane. Hence, it is understandable that observations are most sensitive to the spatial distribution of fluxes in this region. By contrast, the observation network is largely insensitive to spatial variability in wetland fluxes across the US; in most instances, the model selection framework favors a spatially-constant model over a wetland model for the two US regions.

These results indicate that the observation network has limited capacity to evaluate wetland fluxes over the United States. Across Canada, the results are far more promising, despite the relative sparsity of the observation network there. Existing bottom-up methane estimates are highly uncertain across Canada, and the synthetic experiments indicate that atmospheric observations can reduce these uncertainties.

Table 5.1: Spatial patterns chosen by the model selection framework

Region	season	Models chosen over a constant	model name(s)
E. Canada	summer	1 of 7	LPJ-Bern
E. Canada	fall	0 of 7	
W. Canada	summer	2 of 7	LPJ-Bern, SDGVM

5.4 Results and discussion: comparisons with atmospheric data

5.4.1 Spatial flux patterns

We first compare the spatial distribution of the existing wetland flux estimates against methane data from the atmospheric observation network. We apply experiment two to real data and report the results for regions and seasons that had a high success rate in the synthetic experiment. That experiment examined whether the spatial variability in a wetland model is more useful at describing the atmospheric data than a spatial constant. We now apply this question to real data: do the WETCHIMP models have spatial variability that describe the real data better than a spatial constant? If so, which models? This approach indicates whether each model contributes positive information on the location of wetland fluxes or if a spatial constant is a more effective descriptor.

The results of this real data experiment are displayed in Table 5.1. This table only lists the regions and seasons that had a success rate >75% in synthetic data experiment two. If a wetland model describes the distribution of fluxes better than a spatial constant in those regions/seasons, then the model selection framework should select that model.

Only a small number of WETCHIMP models are able to describe the distribution of wetland fluxes (as seen via the atmospheric observations) better than a spatial constant – between 0 – 28% of the available models depending upon the region and season. The model selection framework chooses LPJ-Bern in eastern Canada and LPJ-Bern and SDGVM in western Canada. The spatial patterns in the remaining WETCHIMP models do not perform better than a spatial constant when compared to atmospheric data.

The LPJ-Bern and SDGVM models have several unique spatial characteristics that could explain these results. Over eastern Canada, the LPJ-Bern model concentrates the largest fluxes in the HBL. Other models, by contrast, often distribute the fluxes more broadly across Ontario and Quebec or put the largest fluxes in Ontario outside of the HBL. In western Canada, the LPJ-Bern and SDGVM models distribute fluxes broadly across both northern Saskatchewan and Alberta. A number of other estimates like DLEM or CLM4Me assign relatively small fluxes in these provinces relative to other regions.

The LPJ-Bern and SDGVM models share another common characteristic: both model wetland area independently instead of relying solely on remote sensing inundation datasets. LPJ-WSL, ORCHIDEE, DLEM, and CLM4Me use remote sensing inundation datasets like GEIMS (Global Inundation Extent from Multi-Satellites, Prigent et al., 2007) to construct a wetland map. Other models, like LPJ-Bern, LPJ-WHyMe, and SDGVM also use land cover maps and/or land surveys to estimate wetland (or at least methane-producing) area (Melton et al., 2013a; Wania et al., 2013). Wetland maps generated using the two approaches show substantial differences. Remote sensing datasets estimate relatively high levels of inundation in regions of Canada that are non-forested or have many small lakes (see further discussion in Melton et al., 2013a; Bohn et al., 2015). Independently-generated wetland maps, by contrast, assign more wetlands over regions with high water tables but little surface water. As a result of these differences, models like LPJ-Bern assign more wetlands and methane fluxes in the Hudson Bay Lowlands relative to other regions of Eastern Canada.

5.4.2 Flux magnitude

We next compare the magnitude of predicted concentrations using the WETCHIMP models against atmospheric observations. Unlike previous sections that utilized model selection, this section employs several simple model-data timeseries, displayed in Fig. 5.4. We model methane concentrations at a number of US and Canadian observation sites using WRF-STILT, the WETCHIMP flux estimates, and anthropogenic emissions from the EDGAR v4.2FT2010 inventory. This model estimate consists of several components: the background (in green) is the estimated concentration of methane in clean air before entering the model domain as in Miller et al. (2013) and Miller et al. (2014c). The estimated contribution of anthropogenic emissions from EDGAR v4.2FT2010 is added to this background (in red). Note that the estimated scaling factors for EDGAR (Sect. 5.2.4) are 1.7 ± 0.3 at Chibougamau, 5.6 ± 0.5 at East Trout Lake, 2.4 ± 0.3 at Fraserdale, and 2.5 ± 0.3 at Park Falls. The contribution of wetland fluxes from the WETCHIMP models is then added to the previous inputs, and the sum of all components (blue lines) can be compared directly against measured concentrations.

The various WETCHIMP flux estimates produce very different modeled concentrations at the observation sites (Fig. 5.4). Overall, modeled concentrations with the WETCHIMP fluxes usually exceed the methane measurements during summer. At Chibougamau, Fraserdale, and Park Falls in early summer, all six WETCHIMP models predict methane concentrations that equal or exceed the observations. The ORCHIDEE, LPJ-WHyMe, and LPJ-Bern models always exceed the measurements during summer while DLEM and SDGVM better match the observations at these sites. In contrast to these results, a recent study by Bohn et al. (2015) found that the ensemble average is not biased over the Western Siberian Lowlands relative to inverse modeling estimates. The models also show a large spread in that region.

Methane models that overestimate fluxes in North America do not always compensate with

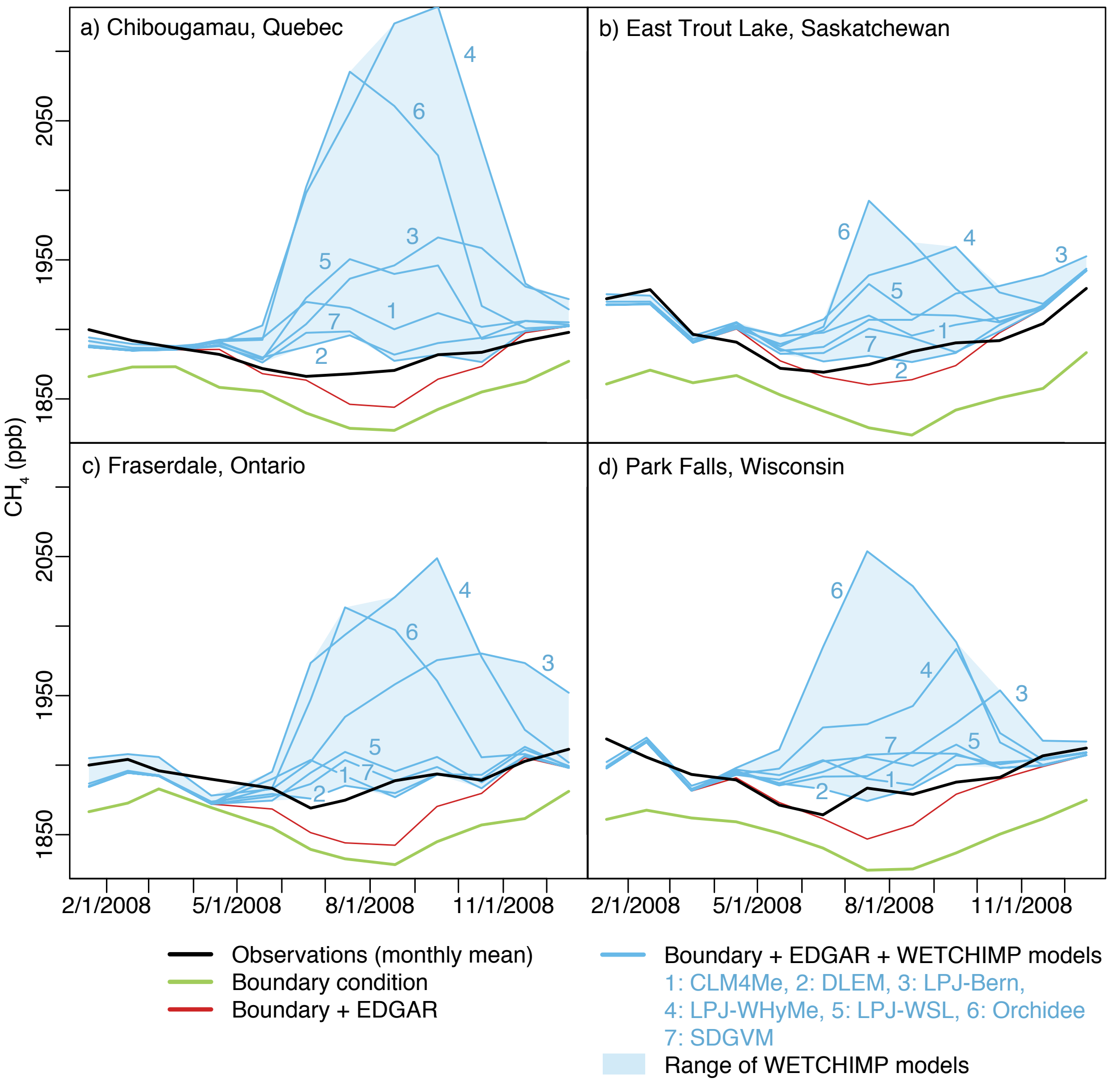


Figure 5.4: These time series compare atmospheric methane measurements at several observation sites against model estimates using the WETCHIMP ensemble and the EDGAR v4.2FT2010 anthropogenic emissions inventory. The range of estimates from the various WETCHIMP models is large.

smaller fluxes elsewhere. For example, the ORCHIDEE model not only estimates large fluxes over North America but also estimates higher fluxes over the tropics than any other model (Melton et al., 2013a). Hence, the disagreement in magnitude over North America not only reflects uncertainty in the global distribution of wetland fluxes but also reflects uncertainty in the global wetland budget.

5.4.3 Seasonal cycle

Bottom-up methane flux estimates show variable performance when compared against atmospheric observations, and the temporal distribution of these estimates is no exception. Figure 5.5 compares the seasonal cycle of the existing estimates over Canada's HBL. Eastern Canada is one of the largest wetland regions in North America (Fig. 5.1), and unlike other regions, the observation network there can detect a clear wetland signal through most of the seasonal cycle (Fig. 5.3).

In this region, the bottom-up estimates diverge on the seasonal cycle of fluxes. Most estimates predict peak fluxes in July or August, though two variations of the LPJ model predict seasonal peaks in September and October (LPJ-WHyMe and LPJ-Bern, respectively). Discrepancies among models are also notable during the fall and spring seasons. For example, fluxes in June account for anywhere between 6% and 21% of the annual methane budget, depending upon the model. Fluxes in October account for between 1% and 23% of the annual budget.

The figure also displays the seasonality of an inverse modeling estimate from Miller et al. (2014c) for comparison. That estimate incorporates observations from Chibougamau and Fraserdale, atmospheric measurement sites that are strongly influenced by fluxes from the HBL. The discrepancies among the WETCHIMP models often exceed the 95% confidence interval of the inversion estimate (Miller et al., 2014c). On whole, the WETCHIMP estimates have a narrower seasonal cycle than the inversion estimate, which assigns comparatively larger fluxes to the fall

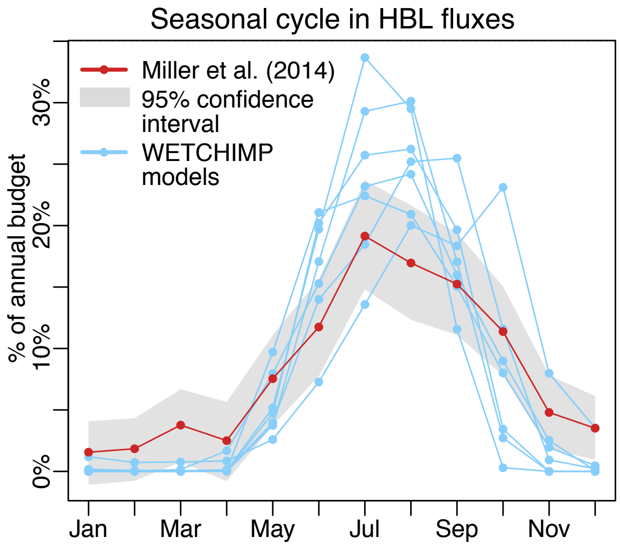


Figure 5.5: The seasonal cycle in methane fluxes estimated for the Hudson Bay Lowlands (HBL; 50–60°N, 75–96°W). We include both the WETCHIMP estimates and an inverse modeling estimate from Miller et al. (2014c). Each month is displayed as a percentage of the annual budget estimated by a given model. This approach highlights differences in the seasonality of the models and controls for differences in magnitude. In general, the WETCHIMP models estimate a narrower seasonal cycle relative to Miller et al. (2014c).

and spring shoulder seasons. A recent inverse modeling study of the Western Siberian Lowlands found parallel results for that region – existing models also under-predict the shoulder seasons relative to summer months (Winderlich, 2012; Bohn et al., 2015).

Numerous possible explanations could underly this discrepancy. For example, the bottom-up models could be too sensitive to soil/air temperature, and may therefore shut off methane emissions too early. Compared to the inversion estimate, the bottom-up models predict small or minimal fluxes during fall/spring months when air temperatures are near freezing but soils are still unfrozen (Fig. C.3). According to estimates from the North American Regional Reanalysis (NARR) (Mesinger et al., 2006), surface soils in the HBL (0 and 10cm depth) begin to thaw in April and are largely unfrozen in May (Fig. C.3). In the fall, surface soils (0 cm depth) begin to freeze in November, but deeper soils (10cm and 40cm) remain largely unfrozen until December. Compared to the bottom-up models, the inversion estimate predicts a wider seasonal window, a result that is consistent with soil freeze/thaw.

5.5 Conclusions

A recent model comparison study revealed substantial differences in existing estimates of wetland methane fluxes – differences at global to regional scales. In the first component of this study, we use two increasing stringent synthetic data experiments to understand how sensitive the atmospheric observation network is to regional-scale wetland fluxes. We find that the network can reliably identify an atmospheric pattern from Canadian wetlands. The network can identify a methane pattern from the eastern US in a majority (though not all) trials and rarely from the western US. The network can also detect spatial variability in the Canadian wetland source but rarely in the US wetland source. This analysis also accounts for distracting signals or patterns in the atmosphere from anthropogenic sources or simulated modeling errors. These results indicate that uncertainties in current methane models can be reduced, if those models begin to leverage available methane data. Furthermore, these discrepancies indicate a disconnect between scientists who build process-based and/or biogeochemical models and scientists who collect or use atmospheric methane data. Improved collaboration between these two groups could help reduce present uncertainties in natural methane fluxes, at least over Canada.

In a second component of the study, we evaluate each existing bottom-up methane model at regional scale using real atmospheric data. We find that only 0–28% of all models have a spatial pattern that describes the atmospheric data more effectively than a constant. The LPJ-Bern and SDGVM models have spatial distributions that are most consistent with atmospheric observations, depending upon the region and season of interest. In addition, almost all existing models overestimate the magnitude of wetland methane fluxes when compared against atmospheric data at individual observation sites. The ensemble of models also appears to estimate a seasonal cycle that is too narrow across the HBL, a large region of methane fluxes in North America. Overall, this study indicates numerous areas for improvement in existing bottom-up wetland methane estimates.

6

Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions

This chapter originally appeared as a research article in the journal Geoscientific Model Development. For the original article, refer to Miller et al. (2014a).

Many inverse problems in the atmospheric sciences involve parameters with known physical constraints. Examples include nonnegativity (e.g., emissions of some urban air pollutants) or upward limits implied by reaction or solubility constants. However, probabilistic inverse modeling approaches based on Gaussian assumptions cannot incorporate such bounds and thus often produce unrealistic results. The atmospheric literature lacks consensus on the best means to overcome this problem, and existing atmospheric studies rely on a limited number of the possible methods with little examination of the relative merits of each.

This paper investigates the applicability of several approaches to bounded inverse problems. A common method of data transformations is found to unrealistically skew estimates for the examined example application. The method of Lagrange multipliers and two Markov chain

Monte Carlo (MCMC) methods yield more realistic and accurate results. In general, the examined MCMC approaches produce the most realistic result but can require substantial computational time. Lagrange multipliers offer an appealing option for large, computationally intensive problems when exact uncertainty bounds are less central to the analysis. A synthetic data inversion of US anthropogenic methane emissions illustrates the strengths and weaknesses of each approach.

6.1 Introduction

Inverse modeling and data assimilation have become ubiquitous in the atmospheric sciences, and one of the most common applications is the estimation of trace gas surface fluxes. These top-down approaches optimize emissions or flux estimates such that modeled atmospheric concentrations reproduce observed concentrations. Most methods are based on Bayesian statistical principles and assumptions of Gaussian probability density functions (pdfs), implemented in a variety of ways (e.g., Gurney et al., 2002; Michalak et al., 2004; Henze et al., 2007; Peters et al., 2007; Gourdji et al., 2008; Stohl et al., 2012).

Many applications require estimating emissions or fluxes that have known physical limits, often referred to simply as inequality constraints. For example, there are few anthropogenic sinks of carbon dioxide or methane, and the release history of air toxins from an industrial hazard site is never negative. In many cases, predicted sources that violate inequality constraints are not only meaningless but distort prediction in surrounding regions or times. For example, if an inversion estimates an unrealistic negative emissions region, emissions in adjacent regions may become larger than expected due to mass conservation (e.g., Michalak, 2008). Hence, it would not be sufficient to simply reset negative emissions to zero. Doing so would not correct for distorted sources elsewhere and would erroneously increase the overall estimated emissions budget (i.e., would violate the mass balance or budget as constrained by the atmospheric observations).

Additionally, enforcing inequality constraints is often necessary for obtaining realistic uncertainty estimates. Even if the posterior emissions themselves do not violate the inequality constraints, their confidence intervals could very well extend beyond known limits under Gaussian assumptions. In such cases, an unconstrained inversion will produce both upper and lower confidence intervals that are unrealistically large (e.g., Snodgrass & Kitanidis, 1997; Michalak & Kitanidis, 2003). The problem occurs because unrealistically low emissions within the lower confidence interval must be balanced by larger emissions elsewhere in the upper confidence interval, or vice versa.

In response to the problems associated with unconstrained inversions, existing trace gas flux estimation studies typically use one of four methods to apply inequality constraints. One method employed in previous studies is a data transformation (refer to Sect. 6.3.1, e.g., Muller & Stavrakou, 2005; Bergamaschi et al., 2009). A second method decreases the uncertainty assigned to many of the prior fluxes until the posterior fluxes obey the known bounds (e.g., Eckhardt et al., 2008; Stohl et al., 2012). This adjustment may run counter to the modeler's physical understanding of the prior estimate or associated uncertainties and therefore is not discussed in great detail here. A third method is that of Lagrange multipliers (refer to Sect. 6.3.2, e.g., Henze et al., 2007; Kopacz et al., 2009; Göckede et al., 2010). Finally, two recent studies use a class of Markov chain Monte Carlo (MCMC) methods known as Metropolis–Hastings (Rigby et al., 2011; Burrows et al., 2013), implemented in a manner that enforces nonnegativity. Existing atmospheric studies provide little guidance on the merits of one approach over another.

The objective of this study is thus to investigate the merits of the above approaches and two additional MCMC implementations. MCMC algorithms are ubiquitous in Bayesian statistics but are not commonly applied to atmospheric studies. The remainder of this paper is organized as follows: Sect. 6.2 examines the statistical assumptions of common inversion methods that are incompatible with inequality constraints. Section 6.3 discusses several possible alternatives to mitigate these statistical assumptions, including data transformations, Lagrange multipliers, and two specific MCMC implementations – a multiple-try Metropolis–Hastings algorithm and a Gibbs sampler. Finally, Sects. 6.4 and 6.5 discuss the costs and benefits of each approach in the context of a synthetic case study estimating North American anthropogenic methane emissions.

6.2 Common Bayesian approaches to inverse modeling

This section describes common approaches to inverse modeling and indicates which statistical assumptions are incompatible with known bounds.

In a typical inverse problem, the unknown quantity to be estimated (s, dimensions $m \times 1$) is different from the quantity actually observed (z, dimensions $n \times 1$), and the two are related to one another by a function $h(s)$. In the case of trace gas inversions, s are the true, unknown emissions or fluxes; z are observations of atmospheric concentration; and the function h is often defined by an atmospheric transport and/or chemistry model:

$$z \sim h(s) + \mathcal{N}(0, \mathbf{R}), \tag{6.1}$$

where $\mathcal{N}(0, \mathbf{R})$, in this case, represents the combined model, measurement, representation, and spatial/temporal aggregation errors, collectively termed model–data mismatch. These errors are most commonly assumed to be random and normally distributed with a mean of zero and an $n \times n$ covariance matrix $\mathbf{R}$.

Any a priori information on the spatial or temporal distribution of s can be incorporated into a model of the mean, $E[s]$:

$$s \sim E[s] + \mathcal{N}(0, \mathbf{Q}). \tag{6.2}$$

This model, $E[s]$, rarely matches the unknown s, and the $m \times m$ covariance matrix $\mathbf{Q}$ describes the magnitude and structure of the residuals between s and $E[s]$. As with the model–data mismatch, these residuals are also typically assumed to be normally distributed with a mean of zero.

The model of the mean can be formulated in a number of ways, but one common method, used in this paper, is as follows:

$$E[s] \; = \; \mathbf{X}\beta, \tag{6.3}$$

where the $m \times p$ matrix $\mathbf{X}$ includes p different covariates, and the unknown $p \times 1$ drift coefficients (β) adjust the magnitude of these covariates to best match the observations. The model of the mean could be uninformative (e.g., $\mathbf{X}$ is an $m \times 1$ vector of ones as in Mueller et al., 2008) or could include any number of covariates, including climatological information or an existing emissions inventory (e.g., Gourdji et al., 2012; Miller et al., 2013, 2014c). Some inversion approaches assume that the drift coefficients are known, in which case $E[s]$ becomes an $m \times 1$ vector (e.g., Rodgers, 2000; Enting, 2002; Tarantola, 2005). An inversion with unknown coefficients has typically been used within the context of a "geostatistical" representation of the inverse problem (used in this study), while the coefficients are usually assumed in the "synthesis Bayesian" approach, though both approaches are Bayesian in nature.

Equation (6.1) can be expanded using the formulation of s described in Eq. (6.3):

$$z \; \sim \; \mathbf{H}\,\mathcal{N}(\mathbf{X}\beta, \mathbf{Q}) + \mathcal{N}(0, \mathbf{R}), \tag{6.4}$$

where $\mathcal{N}(\mathbf{X}\beta, \mathbf{Q})$ represents the distribution of s relative to the prior or model of the mean ($\mathbf{X}\beta$). The $n \times m$ sensitivity matrix, $\mathbf{H}$, is a linearized form of h. This setup assumes, as in most existing studies, that the measurement residuals ($z - \mathbf{H}s$) and flux residuals ($s - \mathbf{X}\beta$) follow a multivariate normal distribution, as will the posterior probabilities of s and β.

The optimal estimate of unknown s can be obtained by minimizing the sum of squared residuals subject to the covariances:

$$\begin{aligned} L_{s,\beta} &= \frac{1}{2}\left(z-\mathbf{H}s\right)^T \mathbf{R}^{-1}\left(z-\mathbf{H}s\right) \\ &+ \frac{1}{2}\left(s-\mathbf{X}\beta\right)^T \mathbf{Q}^{-1}\left(s-\mathbf{X}\beta\right). \end{aligned} \tag{6.5}$$

If **H** does not depend on the unknown value of s, then s can typically be estimated by solving a system of linear equations (refer to Michalak et al. (2004) and Tarantola (2005) for more in-depth discussion on estimating s and the associated posterior uncertainties). Otherwise, the algorithm is usually iterative.

If s has known bounds, then Eq. (6.4) must be reformulated in a way that honors the inequality constraints. Some deterministic methods permanently remove elements of s from the optimization if they violate the bounds (e.g., Lagrange multipliers). In a purely stochastic approach, the first term in Eq. (6.4) will instead follow some multivariate distribution (f_l^u) that is restricted to within the lower and upper constraints (l and u, respectively):

$$z \sim \mathbf{H}\, f_l^u(s|\mathbf{X},\mathbf{Q}) + \mathcal{N}(0,\mathbf{R}). \tag{6.6}$$

This alteration modifies the prior probability of the fluxes; the fluxes (s) will deviate from the prior or expected value ($E[s]$) only to the extent that the fluxes remain within the bounds l and u. In contrast, the distribution of the model-measurement residuals, $\mathcal{N}(0,\mathbf{R})$, remains the same as in the case without inequality constraints. In Eq. (6.6), f_l^u could be formulated as a multivariate truncated normal, exponential, or gamma distribution, among many other choices. Most formulations of f_l^u result in a cost function that does not have a straightforward analytical minimum, unlike the multivariate normal case in Eq. (6.5). Instead, the unknown quantity (s) must be estimated using an algorithm that samples the posterior probability density. Even then, it can be difficult to sample this density efficiently. The next section describes example deterministic and stochastic approaches in greater detail.

6.3 Strategies for enforcing inequality constraints

6.3.1 Data transformations

Data transformations can enforce inequality constraints with relatively easy implementation, but transformations typically render a linear inverse problem nonlinear and therefore require an iterative solution. A number of different data transformations exist, but the power transformation is a common approach because it is defined at zero (unlike log transformations; e.g., Snodgrass & Kitanidis, 1997; Wilks, 2011, Chap. 3):

$$\tilde{s} = \alpha \left(s^{1/\alpha} - 1 \right), \tag{6.7}$$

where s are the fluxes in normal space and α can be any scalar value such that $\tilde{s} > -\alpha$, though larger values of α cause more extreme transformations. This formulation approaches the natural logarithm for large values of α.

For the power transform, the Jacobian or sensitivity matrix ($\mathbf{H}$) is not linear in the transformed space. The algorithm, as a result, becomes iterative and requires updating $\mathbf{H}$ at every iteration until both $\mathbf{H}$ and the best estimate in the transform space ($\tilde{s}$) converge (described in detail by Snodgrass & Kitanidis (1997) and Fienen et al. (2004), among others). Most transformations assume a skewed pdf and therefore lead to skewed posterior uncertainty estimates, and such asymmetry can have a number of implications as discussed in Sect. 6.5. Furthermore, most common transformations can only enforce a single upper or lower bound that is the same for all elements of s.

6.3.2 Lagrange multipliers and the trust region algorithm

The method of Lagrange multipliers is commonly used in deterministic optimization problems to enforce equality or inequality constraints. The approach has also been adapted to a number of stochastic inverse problems in hydrology (Barnes & You, 1992; Walvoort & de Gruijter, 2001; Michalak & Kitanidis, 2004) and more recently in an atmospheric context (Henze et al., 2007; Kopacz et al., 2009; Göckede et al., 2010). The method of Lagrange multipliers can be applied to an inversion by modifying the original cost function $L_{s,\beta}$:

$$L_{s,\beta,\lambda} = L_{s,\beta} - \lambda^T(s - l), \tag{6.8}$$

where l in this case is a lower bound on s, where the bound can be spatially and temporally variable, and λ are the unknown Lagrange multipliers.

A number of implementations exist, but all methods share many similarities. Any element of s that would otherwise violate the inequality constraints becomes fixed on one of the bounds. Most algorithms are iterative and add or remove these elements from the "active" set at each iteration. The optimization proceeds only on the active set and ignores all other elements that have been fixed (e.g., Gill et al., 1981). A large difference among algorithms is the way in which elements are removed or added to the active set.

One result of this setup is that elements in the fixed set are not modeled as continuous random variables. Estimated emissions in these regions have no associated posterior uncertainty. In other words, Lagrange multipliers compromise the stochastic nature of the inverse problem in order to enforce the desired constraints.

Several numerical methods are available for solving constrained optimization problems via the method of Lagrange multipliers, but many are restricted to small or medium-sized problems (e.g., s has fewer than 1000 elements). These include the method of Theil and Van de Panne

(Theil & Panne, 1960; Snyman, 2005, Chap. 3.4) and the active set method (e.g., Gill et al., 1981, or Antoniou & Lu, 2007, Chap. 13.3).

A number of algorithms can efficiently solve large, bounded optimization problems, including the trust region method and the bounded, limited-memory Broyden–Fletcher–Goldfarb–Shanno approach (L-BFGS-B; e.g., Byrd et al., 1995). This paper implements a trust region algorithm. The method can efficiently handle large, bounded optimization problems because it adds and/or subtracts multiple elements from the fixed set at each iteration (e.g., More, 1988; Lin & More, 1999). A trust region algorithm approximates the objective function at each iteration and estimates the range over which this approximation can be trusted (referred to as the trust region). The algorithm optimizes s within the trust region and compares the approximated improvement to the actual reduction in the cost function (in this case Eq. 6.5). If the cost function approximation performs well, the algorithm is allowed to make more aggressive moves at each iteration. In other words, the algorithm may increase the size of the trust region if the approximation does well and vice versa. Though it was originally developed for unconstrained problems, Gay (1984) extended the trust region method to constrained optimization. For additional discussion of general trust region algorithms, see Sorensen (1982), Lin & More (1999), Conn et al. (2000, Chap. 1, Chap. 6), or Yuan (2000).

This paper adopts a general algorithm outlined in Lin & More (1999). The reader is referred to a review article (Yuan, 2000) for a broad discussion of possible trust region implementations. Most require the gradient (∇L) and Hessian ($\nabla^2 L$) of the cost function. For reference, these equations are listed below for the geostatistical approach:

$$\begin{aligned} \nabla L_{s,\beta} &= -\frac{1}{2}\mathbf{H}^T\mathbf{R}^{-1}(z - \mathbf{H}s) \; + \; \frac{1}{2}\mathbf{G}s \\ \nabla^2 L_{s,\beta} &= \frac{1}{2}\mathbf{G} \; + \; \frac{1}{2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \\ \mathbf{G} &= \mathbf{Q}^{-1} \; - \; \mathbf{Q}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Q}^{-1}. \end{aligned} \tag{6.9}$$

To construct these equations, we take the derivative of the original cost function (Eq. 6.5) with respect to β and set this derivative equal to zero (Kitanidis, 1986). We then rearrange the cost function to omit the unknown drift coefficients β. The resulting Hessian and gradient above are written only in terms of the unknown vector s. Rodgers (2000) presents analogous equations for a prior model setup that has predetermined coefficients (β).

6.3.3 MCMC algorithms applied to bounded inversions

The following sections discuss two possible MCMC implementations for inequality-constrained problems. In general, MCMC algorithms make it possible to generate realizations of the unknown quantity from high-dimensional probability density functions. These algorithms make problems with non-Gaussian distributions and/or complex joint pdfs tractable (e.g., Andrieu et al., 2003).

MCMC algorithms simulate a Markov chain with an equilibrium distribution that matches the distributions of the quantities being estimated. The methods rely on the generation of conditional realizations; each realization is a guess of the unknown (e.g., s) that should represent a random draw from the posterior probability distribution. The algorithms create a new realization based only upon the previous one, and the means of doing so differentiate the various MCMC methods. Many conditional realizations are typically generated to adequately sample or represent the equilibrium distribution (Geyer, 2011). The point-wise properties of the equilibrium probability density (e.g., mean, median, percentiles, standard deviation) can be used to represent the statistics of the unknown state, including its uncertainties. A thorough introduction to MCMC approaches is given by Geyer (2011).

MCMC methods can also be used for the solution of bounded problems. Each individual realization of the unknown quantity is restricted by the inequality constraints (Gelfand et al., 1992), ensuring that both the posterior best estimate and associated uncertainties will honor known physical limits.

Metropolis–Hastings

Metropolis–Hastings algorithms have become widespread in Bayesian statistics (see Chib & Greenberg, 1995; Bolstad, 2012, for in-depth discussion). The modeler uses an existing, accepted realization of the unknown quantity (in this case s) to generate a new proposed realization with a Markov chain whose properties are defined by the modeler. One possible approach might generate many realizations of s by using slightly modified inputs for Eq. (6.5). Instead of using z in Eq. (6.5), sample randomly from $\mathcal{N}(z, \mathbf{R})$. Instead of using $E[s]$, use random, sequentially correlated samples from $\mathcal{N}(E[s], \mathbf{Q})$ (see the Supplement for further discussion).

Each subsequent proposed realization is accepted or discarded based on its likelihood relative to the previous, accepted realization. Realizations with a relative likelihood greater than one are always accepted, while those with a relatively likelihood less than one are only sometimes accepted. A large number of realizations is sequentially generated in this way to sample across the probability space of the unknown (in this case the posterior probability distribution of s).

The modeler must carefully balance two considerations when proposing new realizations. If each proposed realization is too close to the previous accepted realization, the algorithm will sample the probability space very slowly. However, if the proposed realization is too far from the previous accepted realization, it will likely have a low relative likelihood and be rejected (e.g., Chib & Greenberg, 1995).

A number of studies have implemented the general algorithm with an adaptation for inequality constraints, both in hydrology (e.g., Michalak & Kitanidis, 2004; Wang & Zabaras, 2006; Zanini & Kitanidis, 2009) and more recently in the atmospheric sciences (e.g., Rigby et al., 2011; Burrows et al., 2013).

The cited hydrology studies use an implementation suitable for large problems (refer to the Supplement). For the specific implementation in these studies, each proposed realization is first constrained to be nonnegative with Lagrange multipliers before being tested for acceptance. This implementation is a compromise between a purely stochastic approach that would represent all elements as continuous random variables and the method of Lagrange multipliers that completely removes some elements from the optimization.

Though ideal for small problems, Metropolis–Hastings algorithms can often become stuck in local regions of high probability when there are many quantities being estimated (i.e., when m is large). The acceptance rate can become so small as to make implementation impractical (Liu et al., 2000). This study implements a multiple-try Metropolis–Hastings algorithm (Liu et al., 2000) suitable for larger-scale inverse problems, described fully in the Supplement.

Gibbs sampler

Unlike the Metropolis–Hastings algorithm, the Gibbs sampler calculates a new realization for each element of the unknown state sequentially (in this case, each of m elements in s). This method involves calculating a probability distribution for an individual element of s conditional on the current realization for all other elements. The algorithm takes a random sample from the element-wise conditional probability density, and this sample becomes the new guess for the given element. Using this method, the Gibbs sampler sequentially calculates a conditional distribution and random sample for each of m elements in s until an entire new, full conditional realization has been formed (see the Supplement). Like the Metropolis–Hastings algorithm, the Gibbs sampler requires generating a large number of conditional realizations, and the statistics of these realizations can be used to define a best estimate and associated uncertainties. For an in-depth review of the Gibbs sampler, refer to Casella & George (1992) or Bolstad (2012, Chap. 10).

Several studies in hydrology apply Gibbs sampler methods to constrained inverse problems (e.g., Michalak & Kitanidis, 2003; Wang & Zabaras, 2005; Fienen et al., 2006; Michalak, 2008). Michalak (2008) describes a flexible implementation in the context of groundwater problems that can incorporate any kind of spatial or temporal correlation in the a priori covariance matrix $\mathbf{Q}$, and this implementation is adapted for the case study here. The implementation uses a multi-dimensional truncated normal as the prior pdf (Michalak, 2008). However, because the Gibbs sampler uses element-wise pdfs, it avoids the mathematical challenge of explicitly maximizing a high-dimensional truncated pdf. The approach thereby enforces the inequality constraints in a computationally tractable manner.

The implementation in this study differs from Michalak (2008) in one important way. Some regions of the United States and Canada have zero anthropogenic methane emissions, and we alter the shape of the pdfs to allow a high probability at zero. The implementation here draws a random sample from a Gaussian conditional distribution for sequential elements in s. If the sample is positive, it becomes the new realization for that element of s. If the sample is negative, we use zero as the realization for that element. The approach is equivalent to modeling the prior, and subsequently the posterior, distributions as truncated Gaussian with an added Dirac delta (a function that is zero at every point except zero). This modification relative to Michalak (2008) results in a peak in the posterior densities at zero.

6.4 Methane case study setup

A synthetic case study of estimating US anthropogenic methane emissions illustrates the comparative drawbacks and benefits of the approaches described above: the power transformation, Lagrange multipliers, and two MCMC implementations, with an unconstrained inversion for comparison. The synthetic study setup uses an existing methane emissions inventory and a model of atmospheric transport to create an estimation problem with known true emissions. The prescribed methane emissions are always nonnegative, so the constraints on this inversion are simple; the estimated emissions must also be nonnegative ($l = 0$). The remainder of this section describes the case study setup in detail.

6.4.1 Model and synthetic data setup

This study employs a regional-scale, particle-following model known as STILT, the Stochastic Time-Inverted Lagrangian Transport model (Lin et al., 2003), to quantify the sensitivity of atmospheric observations to surface sources, and thereby to estimate the sensitivity matrix $\mathbf{H}$. STILT simulations are driven by Weather Research and Forecasting (WRF) wind fields, version 2.2 (Skamarock et al., 2005). In other words, the combination of WRF and STILT serves as the forward model for the methane case study. Nehrkorn et al. (2010) provide a detailed description of the WRF fields used here.

We use WRF-STILT to generate synthetic methane mixing ratios in the same locations as aircraft and tall tower observation sites in the United States (4600 total observations). The tower sites are those in the NOAA Earth Systems Research Laboratory and DOE monitoring network and are displayed in Fig. 6.1. Aircraft data include methane measurements from the NOAA Earth Systems Research Laboratory aircraft program at a variety of locations over North America, DOE flights over the US southern Great Plains (Biraud et al., 2013), and observations from

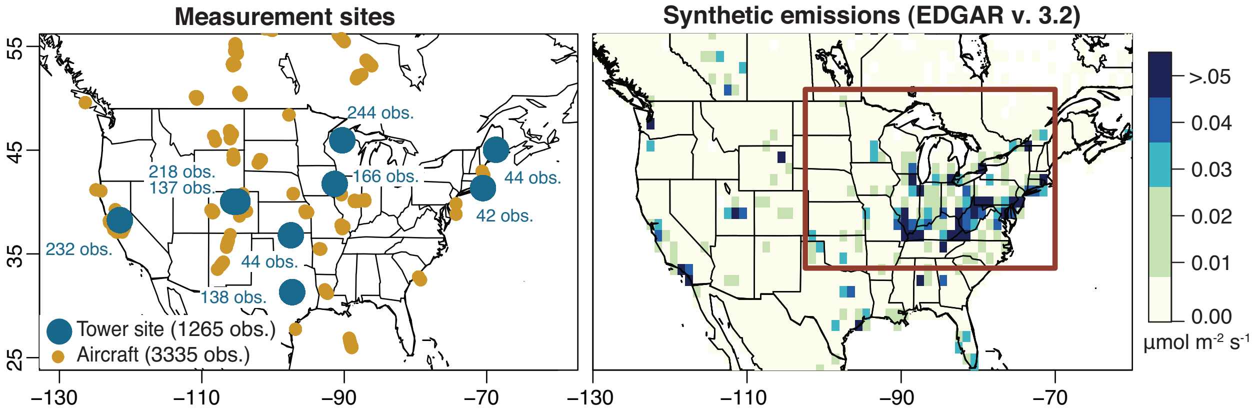


Figure 6.1: The synthetic measurements and synthetic emissions used in this study. Blue numbers (left) indicate the observation count at each tower site. The red box (right) displays the region represented in Table 6.2.

the START08 measurement campaign (Pan et al., 2010). This study includes only aircraft measurements within 2500 m of the ground – measurements that are consistently sensitive to surface fluxes (see Fig. 6.1, Miller et al., 2013).

The study generates synthetic methane measurements using the EDGAR v3.2 FT2000 anthropogenic inventory (Olivier & Peters, 2005). Newer EDGAR inventories are available (e.g., EDGAR v4.2), but top-down studies suggest that version 3.2 best captures the magnitude of anthropogenic sources over the United States (Kort et al., 2008; Miller et al., 2013, see Fig. 6.1).

We add noise to each synthetic measurement, randomly sampled from the model–data mismatch covariance matrix (**R** with diagonal elements σ_R^2). The companion study Miller et al. (2013) statistically infers this information from in situ methane measurements using restricted maximum likelihood estimation (REML) (e.g., Kitanidis & Lane, 1985; Michalak et al., 2004). Table 6.1 summarizes the model–data mismatch values inferred for the towers and aircraft.

Table 6.1: Covariance matrix parameters. (Note: Each tower has a different estimated σ_R. Refer to Miller et al. (2013). Furthermore, in the case of the power transform, we set $\alpha = 6$. This value brings the posterior emissions estimate closest to being normally distributed in the transformed space.)

Parameter	Value
$\sigma_{R,\text{tower}}$ (ppb)	13.1–68.9
$\sigma_{R,\text{aircraft}}$	19.8
Untransformed space	
σ_Q (μmol m^{-2}s^{-1})	0.017
l (km)	101
Transformed space ($\alpha = 6$)	
σ_Q	0.81
l	261

Table 6.2: Eastern US / Canada anthropogenic budgets and 95 % confidence intervals (Tg C month^{-1}) for the true synthetic emissions and inversion estimates.

Type	Budget	% of true emiss. encapsulated in the given confidence interval	
		68.2 %	95 %
True emissions	1.61		
Unconstrained inversion	1.60 ± 0.13	90	97
Transform	1.59 ± 0.20	64	87
Lagrange multipliers	1.60		
Metropolis–Hastings	1.60 ± 0.08	86	97
Gibbs sampler	1.58 ± 0.08	86	96

6.4.2 The inversion setup

The inversion covers much of North America (25–55° N latitude, 145–5° W longitude) on a 1° × 1° spatial resolution over the months May–September 2008. Anthropogenic methane sources do not change markedly from one season to another (Miller et al., 2013). Therefore, the synthetic data study here estimates a single set of emissions over the entire five-month period.

All inversions presented here utilize an uninformative prior (e.g., Michalak et al., 2004; Mueller et al., 2008). In other words, the inversion prior is a single unknown constant across the entire geographic inversion domain. This method makes as few a priori assumptions as possible and relies on the atmospheric data to the fullest extent to infer information about the emissions. This framework is particularly well suited to a synthetic data study; any a priori inventory would be arbitrary since the true emissions are already known.

Despite the lack of information in the prior itself, the inversion incorporates important structural information about the fluxes in the a priori covariance matrix ($\mathbf{Q}$). Specifically, the diagonal elements of $\mathbf{Q}$ describe the total variability of the fluxes (σ_Q^2 – the variance over long spatial scales), and the off-diagonal elements describe the degree of spatial correlation in the posterior flux field, assuming an exponential covariance function. The spatial characteristics of the known emissions field are listed in Table 6.1 and are used to construct $\mathbf{Q}$ (σ_Q^2 and l, the decorrelation length parameter). The parameters for the untransformed space are used in the unconstrained, Metropolis–Hastings, and Gibbs sampler inversions.

6.5 Results and discussion

The inversion implementations discussed in this study produce variable results. All methods place large methane emissions in Kentucky, West Virginia, and along the eastern seaboard, similar to the true synthetic fluxes (see Fig. 6.2), but the methods differ in many other regards. The remainder of this section highlights these differences to illustrate the relative merits of each approach.

6.5.1 Unconstrained inversion

The unconstrained case causes several undesirable side effects, including but not limited to negative emissions estimates (Fig. 6.2). It may appear counterintuitive that the emissions estimate ($\hat{s}$) could have negative components when the observations (z) and model transport ($\mathbf{H}$) contain only positive elements. However, these negative emissions are not necessarily caused by any violation of the statistical assumptions in the inversion. Rather, the estimate ($\hat{s}$) may contain negative elements due to the effect of model–data mismatch errors. When these errors are present, the gradients in the observations can be consistent with adjacent positive and negative sources. In the methane case study, we synthetically generate model–data mismatch errors, so these errors are guaranteed to obey all assumptions of the statistical model. Negative emissions, in this case, could not be caused by biases in wind fields, chemistry, or by a biased prior. In this study, the true winds are known, chemistry is absent, and the geostatistical inverse model does not assume any prior value for the emissions.

Aside from unrealistic negative emissions, the uncertainties are also too large. Conditional realizations and confidence intervals based on multi-Gaussian probabilities extend well beyond the known bounds on the problem (i.e., are not strictly nonnegative), even in regions where the best estimate itself falls within these bounds (Fig. 6.3). Figure 6.4 visualizes this problem in terms of the marginal probability distributions – the probability of an individual element in the emissions field integrating over all possible values of the remaining elements. Even emissions estimated over source regions (Fig. 6.4b) include negative values in the 95% confidence interval.

Additionally, the unconstrained confidence intervals and conditional realizations extend too high for reasons noted in Sect. 6.1. Figure 6.5 shows sample conditional realizations from each method. Emissions in the unconstrained realization extend both lower and higher than the realizations estimated by either MCMC algorithm.

6.5.2 Data transformations

Transformations can be straightforward to implement, but this class of methods skews the probability distributions in the inversion: three of the most important implications are discussed here. First, the posterior covariances (e.g., prior and posterior uncertainties) cannot be directly transformed back to normal space; instead, upper and lower estimation bounds (i.e., percentiles) must be back-transformed to produce posterior confidence intervals. In other words, the covariances become central-value-dependent and are otherwise difficult to physically interpret in back-transformed units. Second, because the covariances are central-value-dependent, it can be difficult to estimate the a priori covariance matrix ($\mathbf{Q}$) in the transformed space, particularly for two of the most common estimation methods. One could use existing knowledge of the emissions to estimate the covariances, but this approach becomes difficult when the covariance matrix has little physical meaning in the transformed space. The covariance matrices can also be inferred from the data and model itself using statistical approaches such as REML. The transformation necessitates iterating between covariance parameter estimation and flux estimation until both converge (Snodgrass & Kitanidis, 1997). The nonlinearities created by the transformation often hinder convergence.

Third, the skewness implied by the power transformation is, in many cases, not representative of actual uncertainties in the emissions best estimate. The uncertainties can become too large in regions of high emissions and too small in regions of low emissions (Fig. 6.2; e.g., Snodgrass & Kitanidis, 1997; Fienen et al., 2004; Muller & Stavrakou, 2005). For example, conditional realizations follow the lower bounds in the methane case study but produce estimates of the sources that are too large in some high-emissions areas. As a result, the back-transformed conditional realizations have an average eastern US budget of 2.1 ± 0.2 Tg C month^{-1}. The mean of these realizations is much higher than the emissions best estimate for the transform inversion (Table 6.2). In this case, the best estimate is computed by minimizing the inversion cost function directly. In contrast, the mean of the conditional realizations is identical to the emissions best estimate for all other methods discussed in this paper. In addition, the average uncertainties in Table 6.2 are larger than any other method, yet these uncertainties capture a lower percentage of the synthetic fluxes at grid scale than other methods. These pitfalls illustrate difficulties in interpreting uncertainties or realizations in the power transform case. Snodgrass & Kitanidis (1997), Fienen et al. (2004) and Muller & Stavrakou (2005) provide further discussion on several of the above challenges associated with data transformations.

6.5.3 Lagrange multipliers

The emissions estimated via Lagrange multipliers reproduce the magnitude and distribution of the true sources. This method is not truly stochastic, however, and removes many elements from the optimization entirely (e.g., emissions over most of Manitoba, Ontario, and Quebec, Canada, in Fig. 6.2). As such, there is no way to calculate either uncertainty bounds or conditional realizations using this approach. The uncertainties assigned to the posterior emissions are typically borrowed from the unconstrained case, though the uncertainties could be borrowed from any other method. Hence, estimation via Lagrange multipliers resolves the problem of unrealistic emissions, but it does not address the challenge of estimating bounded posterior uncertainties or confidence intervals.

6.5.4 MCMC implementations

The MCMC implementations discussed here provide an appealing option when robust uncertainty bounds are a priority in the analysis. Both of the explored implementations ensure that the best estimate (Fig. 6.2), conditional realizations (Fig. 6.5), and confidence intervals respect the known bounds.

Both MCMC implementations produce much narrower uncertainty bounds relative to the other methods (Fig. 6.2, Table 6.2). As discussed in the introduction, the reason for this is twofold. First, the confidence intervals must be smaller because they cannot include values outside the inequality constraints. Second, if the lower range of the confidence intervals is limited, then the maximum emissions values in the interval will also be less extreme (and vice versa; see Sect. 6.1). The uncertainties are smaller, and yet 96–97 % of the synthetic fluxes still fall within the inversion's 95 % confidence interval (Table 6.2). For these reasons, the smaller confidence intervals estimated by the MCMC implementations are most realistic.

The estimated emissions and marginal distributions look very similar between the two MCMC implementations, but the methods show several subtle differences. Unlike the Gibbs sampler, the implementation of the Metropolis–Hastings algorithm here uses Lagrange multipliers and therefore does not explicitly model every element of s in every realization as a continuous random variable. As a result, this Metropolis–Hastings method will always produce a high probability at the inequality constraints (e.g., Fig. 6.4). For smaller problems, in contrast, it may be feasible to use any shape of prior pdf that is defined only within the inequality constraints. That setup would model every element of s as a continuous random variable, but the approach would likely become computationally intractable for larger problems (see the supplement for further discussion). Instead, the Metropolis–Hastings implementation used here is an extension of Lagrange multipliers to circumstances that require bounded confidence intervals. The Gibbs sampler implementation produces a similar peak in the pdf at zero due to the Dirac delta. In general, however, the Gibbs sampler allows more flexibility in setting the shape of the marginal distributions near the bounds.

Appropriate distributional assumptions are important for any type of inversion, and the inversion with inequality constraints is no different. The Gibbs sampler in this case study models the marginal distributions as a truncated Gaussian with a Dirac delta function (see Sect. 6.3.3). If the fluxes or emissions are unlikely to be zero, an implementation without the Dirac delta would be more suitable.

Furthermore, the choice of a truncated normal distribution may not always be appropriate. If the total budget is poorly constrained by the data, this distributional choice could increase estimated emissions in remote regions far from measurement sites. A Gaussian pdf that has been truncated at zero will have a higher mean than the equivalent, full Gaussian distribution, and this effect can shift the posterior mean in poorly constrained problems. One solution could be to fix the drift coefficients (β) at predetermined values, but these coefficients are rarely known in practice. In contrast, if measurement sites are sensitive to emissions across the entire geographic domain (as indicated by $\mathbf{H}$), then either distributional assumption will produce the same trace gas budget.

The MCMC implementations produce the most realistic best estimate, conditional realizations, and uncertainty bounds, but one drawback can be computational cost. The generation of j conditional realizations using the Gibbs sampler requires a for loop with jm iterations, and j is usually 1000 or greater to adequately sample the posterior probability space. The computational time of the multiple-try Metropolis–Hastings depends on the convergence rate of the Lagrange multiplier algorithm and on the number of trial realizations (denoted k; see the Supplement) computed in each step of the multiple-try implementation. The often large ratio of trial to accepted realizations means that the multiple-try Metropolis–Hastings may be less efficient than the Gibbs implementation. This comparison may seem counterintuitive because Metropolis–Hastings does not require the posterior pdf to be sampled one element at a time, as is the case for the Gibbs sampler.

In general, the computational cost of MCMC algorithms increases both with the number of unknown fluxes (m) and the number of realizations (j) required to sample the posterior probability space. Furthermore, the recommended number of realizations changes depending on the size of the problem and the degree of correlation among successive iterations; Gelman (2004) provides a number of guidelines for choosing this number. One approach requires initiating multiple independent Markov chains. The MCMC algorithm reaches convergence when each individual chain has a similar distribution to the combination of all chains. Refer to Gelman (2004, Chap. 11) for more detail. Parallelization can also alleviate some time expense for MCMC algorithms.

In summary, the Gibbs and Metropolis–Hastings implementations produce similar results, but the Gibbs sampler can afford two advantages: greater flexibility in determining the shape of the marginal distributions at the bounds and reduced computational time for the case examined here.

6.6 Conclusions

For inverse problems with parameters that have known physical limits, an unconstrained inversion presents difficulties that go beyond just an unrealistic estimate, and the common remedy of using data transformations can have many undesirable side effects. This study uses anthropogenic methane emissions as a lens to evaluate this approach, as well as several less common alternative approaches.

Inverse problems can be constructed to honor known bounds without compromising the integrity of the estimate. Lagrange multipliers are a viable approach for large problems in which computational time is paramount. However, this method does not provide an explicit means for calculating uncertainty bounds. Uncertainties are usually borrowed from the unconstrained case instead, and these are generally unrealistically large.

Markov chain Monte Carlo (MCMC) methods show the most promise but are rarely applied in the existing atmospheric literature. Both MCMC implementations here produce similar results for the methane case study, but the Gibbs sampler offers better computational efficiency and more flexibility in determining the shape of posterior probability at the bounds. In general, MCMC algorithms can be applied to inverse problems with known bounds to produce the most realistic best estimates, confidence intervals, and conditional realizations of any of the aforementioned approaches.

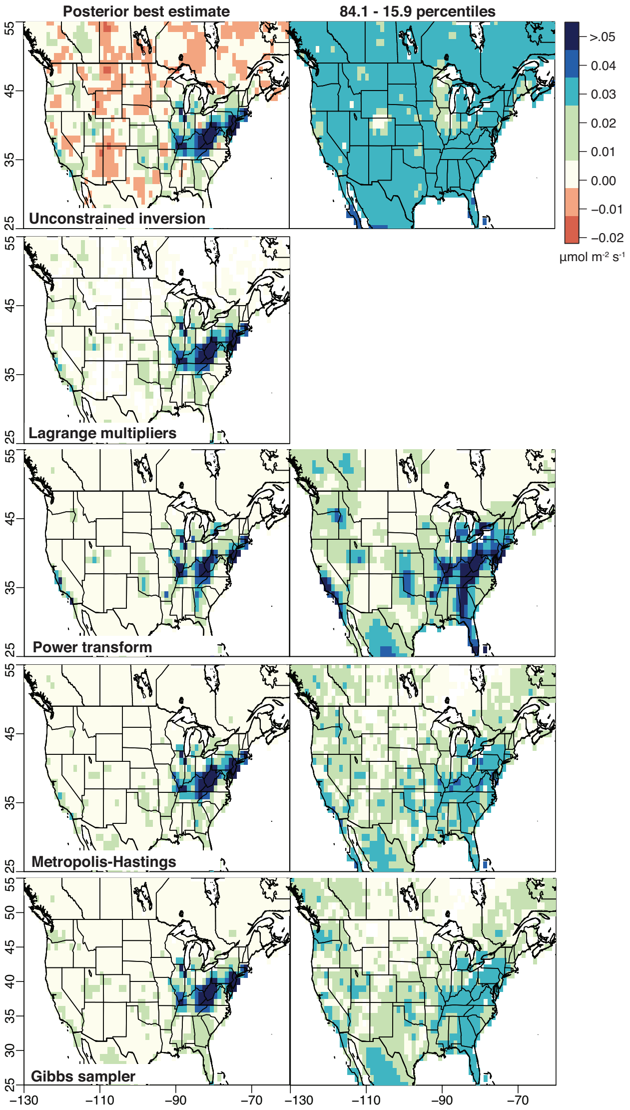


Figure 6.2: The posterior best estimate of the emissions and uncertainties associated with each methodological approach. The method of Lagrange multipliers does not support a direct means of estimating uncertainties.

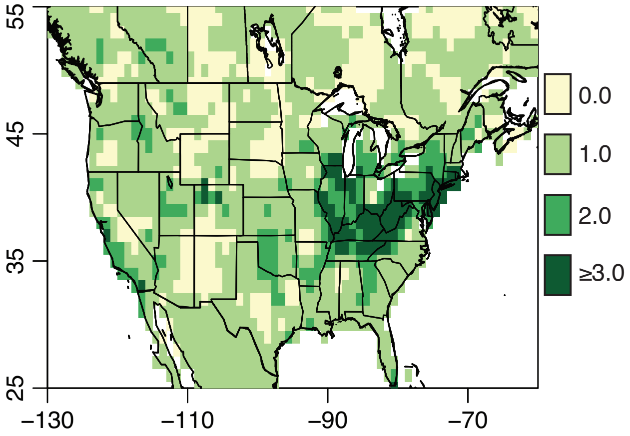


Figure 6.3: The number of posterior uncertainty standard deviations before the methane emissions become negative in the unconstrained estimate.

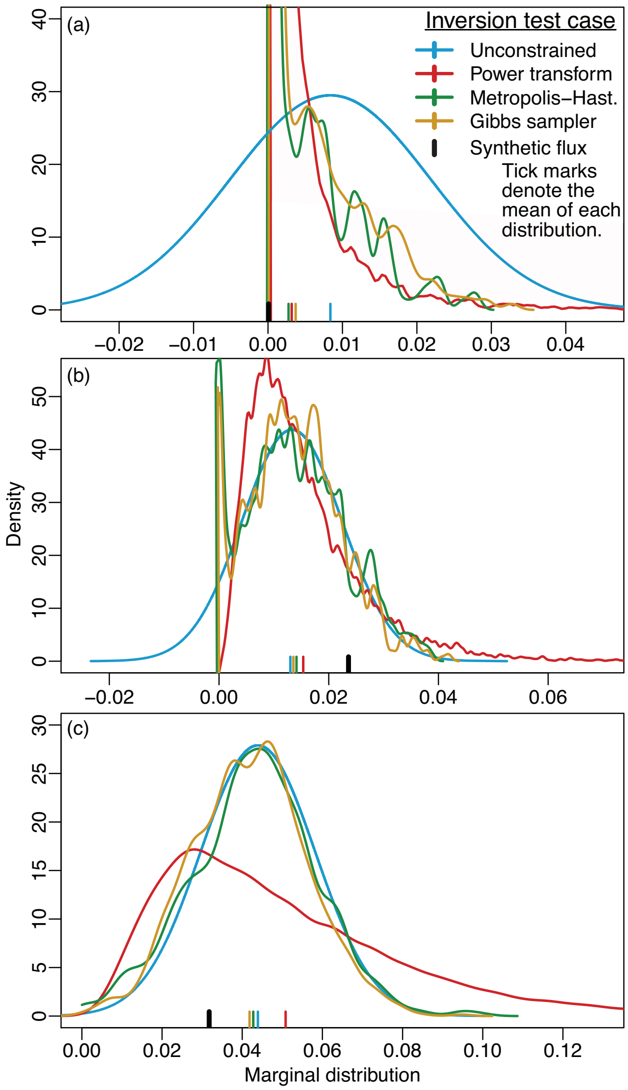


Figure 6.4: The marginal posterior density for the estimate of methane emissions at three individual locations. Case (a) is an estimate of emissions north of Thunder Bay, Ontario, Canada; case (b) is over Indianapolis, Indiana; and case (c) is over eastern Kentucky. The unconstrained case is plotted as a normal distribution, and the other plotted probability densities are produced by applying a kernel smoother to the histogram of realizations. Note that this figure does not include the Lagrange multipliers case because this deterministic approach produces only a best estimate with no associated marginal densities.

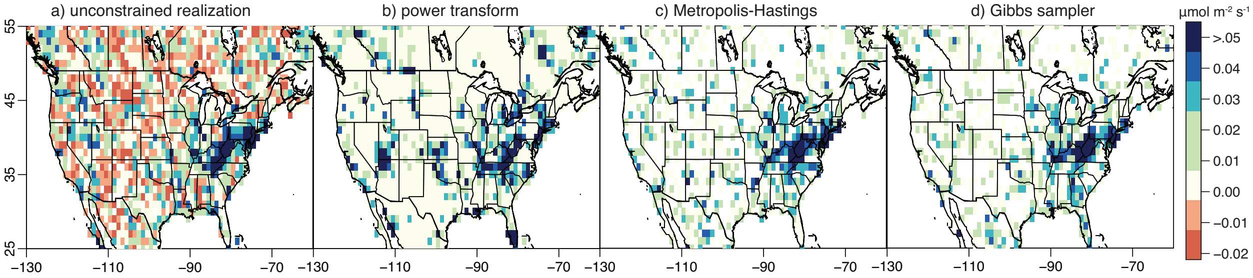


Figure 6.5: Example conditional realizations from each different optimization approach.

7

Biases in atmospheric CO_2 estimates from correlated meteorology modeling errors

This chapter originally appeared as a research article in Atmospheric Chemistry and Physics. For the original article, refer to Miller et al. (2015).

Estimates of CO_2 fluxes that are based on atmospheric measurements rely upon a meteorology model to simulate atmospheric transport. These models provide a quantitative link between the surface fluxes and CO_2 measurements taken downwind. Errors in the meteorology can therefore cause errors in the estimated CO_2 fluxes. Meteorology errors that correlate or covary across time and/or space are particularly worrisome; they can cause biases in modeled atmospheric CO_2 that are easily confused with the CO_2 signal from surface fluxes, and they are difficult to characterize. In this paper, we leverage an ensemble of global meteorology model outputs combined with a data assimilation system to estimate these biases in modeled atmospheric CO_2. In one case study, we estimate the magnitude of month-long CO_2 biases relative to CO_2 boundary layer enhancements and quantify how that answer changes if we either include or remove error correlations or covariances. In a second case study, we investigate which meteorological conditions are associated with these CO_2 biases.

In the first case study, we estimate uncertainties of 0.5–7 ppm in monthly-averaged CO_2 concentrations, depending upon location (95 % confidence interval). These uncertainties correspond to 13–150 % of the mean afternoon CO_2 boundary layer enhancement at individual observation sites. When we remove error covariances, however, this range drops to 2–22 %. Top-down studies that ignore these covariances could therefore underestimate the uncertainties and / or propagate transport errors into the flux estimate.

In the second case study, we find that these month-long errors in atmospheric transport are anti-correlated with temperature and planetary boundary layer (PBL) height over terrestrial regions. In marine environments, by contrast, these errors are more strongly associated with weak zonal winds. Many errors, however, are not correlated with a single meteorological parameter, suggesting that a single meteorological proxy is not sufficient to characterize uncertainties in atmospheric CO_2. Together, these two case studies provide information to improve the setup of future top-down inverse modeling studies, preventing unforeseen biases in estimated CO_2 fluxes.

7.1 Introduction

Scientists increasingly use atmospheric CO_2 observations to estimate CO_2 fluxes at Earth's surface (e.g., Gurney et al., 2002; Michalak et al., 2004; Peters et al., 2007; Gourdji et al., 2012). This "top-down" approach contrasts with "bottom-up" studies that rely primarily on expert knowledge of biological processes (e.g., Huntzinger et al., 2012; Raczka et al., 2013). In order to estimate the fluxes, top-down studies typically require a meteorology model to link fluxes at the surface with measurements taken downwind. Using this link, one can estimate the fluxes even if the atmospheric measurements do not themselves directly measure the fluxes.

However, both the accuracy and effective resolution of the flux estimate hinge upon the accuracy of the meteorological model. Errors in the meteorological model may (or may not) bias estimated CO_2 fluxes depending upon the error characteristics and the space/timescales of interest.

More specifically, the effect of CO_2 transport errors on the estimated fluxes depends upon two important factors. First, the flux estimate becomes more uncertain as the CO_2 transport error variance (or standard deviation) increases. Top-down studies that use Bayesian statistics will explicitly account for these variances when estimating fluxes (e.g., Enting, 2002; Tarantola, 2005); before estimating the fluxes, the modeler first estimates the total variance due to an array of model or data errors – due to imperfect atmospheric transport or imperfect measurements, among many other sources of error (e.g., Gerbig et al., 2003; Michalak et al., 2005; Ciais et al., 2011).

Second, the flux estimate becomes more uncertain as the temporal and/or spatial error covariances increase. As the covariances increase, each CO_2 measurement effectively provides less and less independent information to constrain the surface fluxes. Furthermore, these temporally and/or spatially correlated errors can bias the flux estimate over a region or over the entire geographic area of interest (e.g., Stephens et al., 2007).

Quantification of this complex cause-and-effect between meteorological errors and errors in estimated CO_2 fluxes represents an ongoing research challenge, and a number of existing studies have characterized different aspects of these uncertainties. For example, a series of studies known as TransCom (Atmospheric Tracer Transport Model Intercomparison) represents one of the first coordinated projects on CO_2 transport uncertainties (Gurney et al., 2002; Baker et al., 2006). These early studies used 13 different global atmospheric models and compared differences in top-down CO_2 budgets due to atmospheric model differences. Subsequent to the

TransCom project, a number of studies have focused on the effects of changing vertical mixing and/or planetary boundary layer height (PBLH) (Gerbig et al., 2008; Williams et al., 2011; Kretschmer et al., 2012, 2014; Parazoo et al., 2012; Pino et al., 2012). In general, those papers found that uncertainties in PBLH can lead to biases of ∼3 ppm in modeled daytime CO_2. Another paper examined the effect of uncertain horizontal winds (Lin & Gerbig, 2005). The authors applied a particle-trajectory model at a measurement site in Wisconsin and found that uncertainties in the horizontal winds contributed ∼6 ppm (standard deviation) to the overall CO_2 transport uncertainty. In summary, a number of previous studies have either perturbed individual meteorological parameters or, in the case of TransCom, sampled transport uncertainties using 13 preselected atmospheric models.

The present study is particularly concerned with temporal and/or spatial error covariances in atmospheric CO_2 transport. To what extent do CO_2 transport errors covary in space and time? How large are these covariances relative to the magnitude of the surface CO_2 fluxes, and which meteorological factors drive large error covariances? These covariances are often difficult to characterize (e.g., Lin & Gerbig, 2005; Lauvaux et al., 2009) and are omitted from most existing top-down efforts.

We explore several facets of these questions using a global meteorology model ensemble and a meteorology data assimilation system – the Community Atmosphere Model (CAM) and a Local Ensemble Transform Kalman Filter (LETKF) (Hunt et al., 2007). Efforts by Liu et al. (2011) and Liu et al. (2012) extended this meteorological framework to model uncertainties in atmospheric CO_2.

This framework systematically estimates meteorology and CO_2 transport uncertainties to an extent not previously possible; CAM-LETKF explicitly represents the CO_2 transport uncertainties that remain after assimilating several hundred thousand meteorology observations at each 6 h model time step. To accomplish this task, CAM-LETKF uses an ensemble of weather forecasts

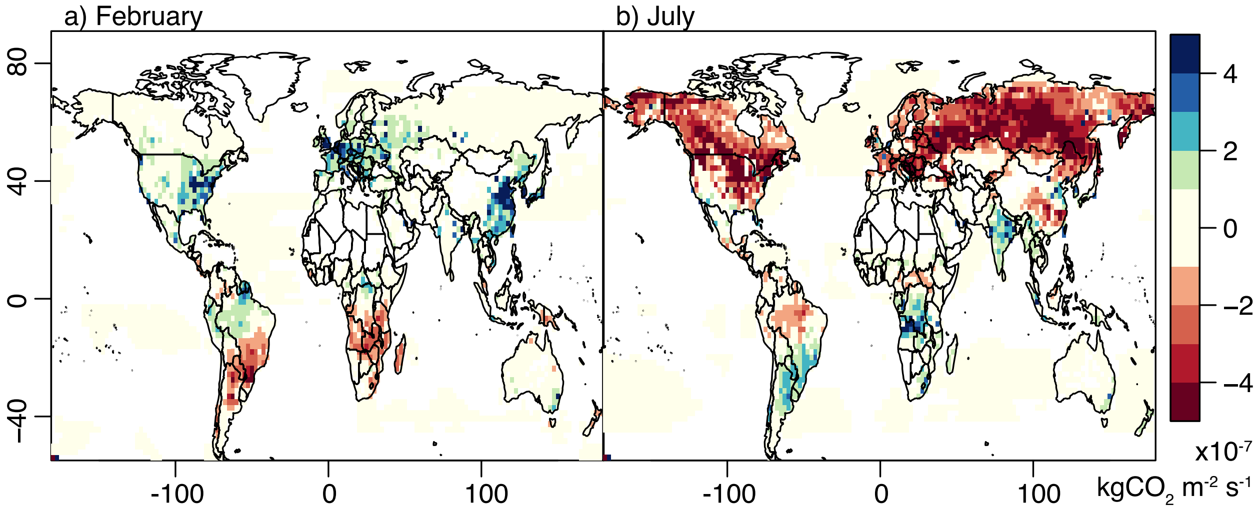


Figure 7.1: Average CarbonTracker CO_2 fluxes (version 2011oi) for (a) February and (b) July 2009. The fluxes include biosphere, ocean, fossil fuel, and biomass burning fluxes (`http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT2011_oi`).

and optimizes the ensemble to match available meteorological observations. Furthermore, CAM-LETKF adjusts the variance of the weather ensemble at each time step to match the modeling uncertainties implied by the meteorological observations.

Using this toolkit, we construct several case studies to understand both the possible magnitude and drivers of CO_2 transport error covariances – errors that persist over many time steps and/or across large regions. The next section describes CAM-LETKF and these case studies in greater detail.

7.2 Methods

7.2.1 The meteorology and CO_2 model

The first component of CAM-LETKF is the meteorological model. We simulate global meteorology using CAM and the Community Land Model (CLM, version 3.5) run in weather forecast mode (not climate mode) (Collins et al., 2006; Oleson et al., 2008; Chen et al., 2010). Model simulations in this study have a spatial resolution of 2.5° longitude by 1.9° latitude with 26 vertical model levels. In most regions, there are three vertical model levels within the lowest kilometer of the atmosphere. These model levels are centered at 929.6, 970.6, and 992.6 hPa over regions where the land / water surface is at sea level.

We save the global model output at 6 h time increments. Furthermore, we run the model for two time periods: January–February 2009 and May–July 2009. The first month of each run serves as an initial spin-up for the model–data assimilation system. The next section describes this assimilation in greater detail.

7.2.2 The meteorological model–data assimilation framework

The second component of CAM-LETKF is the data assimilation and model optimization framework. This framework serves two purposes. First, the LETKF optimizes modeled meteorology (CAM-CLM) to match available observations. Second, the LETKF uses an ensemble of model forecasts to represent model uncertainties that remain after data assimilation (Hunt et al., 2004, 2007). We define each ensemble member and the mean of the entire ensemble as follows:

$$x_i = \bar{x} + \mathbf{X}_i \qquad \text{where } i = 1\ldots k, \tag{7.1}$$

where x_i $(m \times 1)$ is a single model ensemble member, $\bar{x}$ $(m \times 1)$ is the mean of the model ensemble, and $\mathbf{X}_i$ $(m \times k)$ refers to the ith column of the matrix that defines the ensemble spread. In this paper, the variable m refers to the total number of model parameters – the model estimate for a variety of meteorological variables, concatenated across the globe and across all 6-hourly time steps in a given model run. Furthermore, we use $k = 64$ total ensemble members in this setup, as was done in Liu et al. (2011) and Liu et al. (2012).

Using this ensemble, CAM-LETKF steps through time in sequential 6 h intervals. First, the model ensemble at time t is optimized to match meteorological data (Hunt et al., 2007). To this end, we assimilate the same meteorological observations used in the National Centers for Environmental Prediction – Department of Energy reanalysis 2 (Kanamitsu et al., 2002): temperature (in situ and satellite), zonal wind (in situ and satellite), meridional wind (in situ and satellite), surface pressure (in situ), and specific humidity (in situ). At each 6 h model time step, we assimilate between $\sim$180 000 and 330 000 observations globally. At that juncture, the ensemble mean associated with time t, $\bar{x}(t)$, represents the model best guess and the ensemble members, $\bar{x}(t) + \mathbf{X}(t)$, collectively represent the uncertainties in the modeled meteorology (i.e., posterior variances and covariances). Second, we run 6 h CAM-CLM forecasts using these realizations as initial conditions – a total of 64 model forecasts. The 6 h cycle of data assimilation and model forecast then begins again.

This model ensemble, by design, is guaranteed to reflect actual uncertainties in modeled meteorology; at each 6 h model time step, we adjust the ensemble variance such that this variance matches against the model–data residuals (Li et al., 2009; Miyoshi, 2011). The Supplement describes this procedure, known as adaptive covariance inflation.

The model ensemble also accounts for both spatial and temporal covariances in modeled meteorological uncertainties; meteorological errors within one ensemble member can easily persist over many time steps. This continuity occurs because the optimized ensemble members from the one time step become the initial conditions for the weather forecast at the next time step. For example, if the PBL height in one ensemble member is lower than the ensemble average at a given time step, it will likely be lower than average at the next time step. Similarly, if the PBL height in one ensemble member is lower than average over one grid box, it will likely also be lower than the average over an adjacent grid box.

Certain meteorological uncertainties, however, may not always be captured by the assimilation system, particularly uncertainties that do not manifest in the model–data residuals. For example, CAM-LETKF will not fully characterize uncertainties due to different PBL schemes (e.g., Yonsei versus Mellor–Yamada–Janjic) or due to other structural model differences. Furthermore, LETKF cannot spatially resolve uncertainties that occur at sub-grid scale (e.g., turbulent eddies or numerical diffusion). For further technical detail on the LETKF and adaptive covariance inflation, refer to the Supplement, Hunt et al. (2004, 2007), Li et al. (2009), Liu et al. (2011), or Miyoshi (2011).

7.2.3 CO_2 transport error variances and covariances

The CAM-LETKF system described above estimates not only meteorological uncertainties but also uncertainties in CO_2 transport. In this study, CO_2 is a passive tracer that is not part of the data assimilation, so any uncertainties in CO_2 concentrations are solely due to uncertainties in atmospheric transport.

We drive all model simulations with a published CO_2 flux estimate from CarbonTracker (CT), version CT2011oi (Fig. 7.1; Peters et al., 2007, `http://carbontracker.noaa.gov`). CT is a commonly used global CO_2 flux estimate created by the US National Oceanic and Atmospheric Administration (NOAA). NOAA scientists optimize CT fluxes to match atmospheric CO_2 data, so the flux estimate is consistent with actual observations (Peters et al., 2007). The original CT fluxes have a temporal resolution of 3 h. We average these fluxes to a 6 h resolution for all of the CAM simulations in this study.

We subsequently estimate 6-hourly CO_2 transport uncertainties using this setup. These uncertainties are defined as the difference between the top and bottom of the 95 % confidence interval, computed from the 64 model realizations (e.g., Fig. 7.2). To make this estimate, we calculate the 2.5th and 97.5th percentiles of each row in $\mathbf{X}_{[CO_2]}$, where the subscript $[CO_2]$ refers to the atmospheric CO_2 concentrations estimated by the ensemble. The remainder of the methods section applies this CO_2 and meteorology modeling framework to two case studies.

7.2.4 Case study 1: the magnitude of temporally and spatially covarying atmospheric transport errors relative to a CO_2 flux estimate

This case study explores the importance of persistent, covarying transport errors and the magnitude of those errors relative to the CO_2 fluxes. In particular, we estimate uncertainties in monthly mean, afternoon, modeled CO_2 concentrations at a number of in situ atmospheric observation sites. In one case, we include temporal and/or spatial covariances in the atmospheric transport errors, and in another case we remove these covariances. We then compare these uncertainties against the modeled afternoon CO_2 boundary layer enhancement to understand the magnitude of these errors relative to the surface fluxes.

The uncertainty in monthly-averaged CO_2 concentrations serves as a measure of how transport errors correlate or covary across time. Uncorrelated transport errors will average out, to a large degree, over many model time steps, but temporal error covariances prevent the errors from averaging down over time. Furthermore, CO_2 budgets are often reported in month-long increments (e.g., Gourdji et al., 2012, and CT), so this time window is a relevant benchmark with respect to inverse modeling studies.

We calculate uncertainties in the monthly-averaged model output (including error covariances) via several steps. First, we select out the rows of $\mathbf{X}_{[CO_2]}$ that correspond to afternoon observations (13:00–19:00 LT) for a given month at an in situ CO_2 observation site. Second, we calculate the mean of each column in $\mathbf{X}_{[CO_2]}$. Each column corresponds to a different ensemble member. The resulting vector of length 64 is the difference between each ensemble member and the best estimate ($\bar{x}$), averaged at the monthly scale. Lastly, we use this vector to compute a confidence interval in monthly-averaged, modeled CO_2 (the 97.5th minus 2.5th percentiles).

We subsequently remove covariances in the CO_2 transport errors and recalculate uncertainties in the monthly-averaged CO_2 concentrations. As described in Sect. 7.2.2, errors in one ensemble member can persist over many steps and can persist across a large geographic region. However, we can remove these error covariances by randomly reshuffling the elements of each individual row in $\mathbf{X}_{[CO_2]}$. The variance in modeled concentrations in any row or at any given time step will remain the same. However, each column will no longer represent a single ensemble member. Rather, each column will represent a random assortment of different ensemble members, and the errors in each column will no longer covary from one time step to another or one geographic location to another.

We conduct this analysis at a representative selection of observation sites in North America, Asia, and Europe. This setup indicates how errors covary with time at the monthly scale. In addition, we also conduct the analysis using multiple observation sites; we estimate monthly-averaged uncertainties at the eco-region scale and include all observation sites that lie within the given eco-region. This latter approach indicates how errors covary spatially across multiple sites at the regional scale.

These monthly-averaged uncertainties can then be compared against the afternoon, modeled CO_2 increment from regional surface fluxes. To estimate this increment, we subtract modeled free troposphere, "clean air" concentrations at 600 hPa from concentrations modeled at the surface using CT fluxes. The concentrations at 600 hPa are not necessarily a perfect measure of clean air concentrations. Rather, this approach is an approximation similar to that used by inverse modeling studies in the literature (e.g., Gerbig et al., 2003; Gourdji et al., 2012).

In summary, case study one explores the magnitude of persistent atmospheric CO_2 transport errors or error covariances relative to the afternoon CO_2 signal from surface fluxes. The next case study, in contrast, explores the meteorological conditions under which these persistent CO_2 transport errors may be more likely to occur.

7.2.5 Case study 2: which meteorological factors may be associated with month-long transport biases?

We create a synthetic experiment to explore the meteorological conditions under which month-long model biases in atmospheric transport may occur. The spatial patterns in the CO_2 transport uncertainties are heavily influenced by spatial patterns in the CO_2 fluxes (Fig. 7.2). In other words, regions with large fluxes or large diurnal flux variability also show higher CO_2 transport uncertainties. As a result, it is difficult to disentangle the effect of different meteorological parameters on CO_2 transport uncertainties. Instead, we create a synthetic tracer with constant global emissions in both space and time. This experiment serves as a lens to explore the possible effects of different meteorological parameters independent of the spatiotemporal variability in CO_2 fluxes.

To this end, we initialize CAM-LETKF runs with zero atmospheric concentration of this synthetic tracer and then run CAM-LETKF forward for 1 month using constant global emissions (e.g., for both February and July 2009). Any uncertainties in the atmospheric distribution of this tracer are solely due to meteorological parameters, not due to the spatial distribution of the underlying fluxes.

Next, we calculate the coefficient of variation (CV) associated with the monthly-averaged surface concentrations. The CV is an inverted signal-to-noise ratio; it measures the uncertainty in modeled surface concentrations relative to the average surface concentration ($\frac{\sigma}{\mu}$). For example, an uncertainty of 1 ppm in modeled concentrations is most problematic if the signal from surface fluxes is weak, and a 1 ppm uncertainty is less problematic if the signal from surface sources and/or sinks is strong.

For this setup, the CV equals the standard deviation in the monthly-averaged surface concentrations divided by the monthly surface concentration averaged across all 64 realizations.

We then plot the tracer CV against monthly-averaged meteorological parameters and their associated uncertainties from CAM-LETKF. These relationships give insight into the meteorological conditions or meteorological uncertainties that are associated with month-long biases in the modeled synthetic tracer.

7.3 Results and discussion

7.3.1 Uncertainties in the 6-hourly modeled CO_2 concentrations

Before examining the two case studies in detail, we first provide context on the CO_2 transport uncertainties estimated with CT fluxes and CAM-LETKF. Figure 7.2a and b visually summarize the average 6-hourly CO_2 transport uncertainties in the model surface layer – the difference between the top and bottom of the 95 % confidence intervals. These figures show how CO_2 transport uncertainties vary across the globe – from 0.6 to 26 ppm, depending on location. Furthermore, the transport uncertainties in Fig. 7.2a and b show several distinctive features. The largest uncertainties are localized to regions where either the magnitude or the diurnal cycle of the CT fluxes is largest (e.g., the US Eastern Seaboard and southern Siberia during summertime, the Amazon, the Congo, and eastern China). CO_2 transport uncertainties in the eastern US and eastern Asia bleed, to a smaller degree, over the adjacent ocean where surface fluxes are small.

Figure 7.3 places these transport uncertainties in context of CO_2 data measured at two observation sites in the United States. These time series plots validate the model's capacity to simulate daily variations in CO_2 concentrations. Furthermore, the comparison illustrates the magnitude of the CO_2 transport uncertainties relative to the diurnal cycle in CO_2 concentrations. For example, the uncertainties at AMT in July are $\sim$ 30 % of the diurnal range in the CO_2 measurements. Overall, the model ensemble depicted in these plots usually encapsulates the hourly-averaged measurements. CT fluxes are estimated using these CO_2 observations and the TM5 transport model (Tracer Model, version 5) (Peters et al., 2007), so one might expect the CAM model to fit the CO_2 observations relatively well. In the instances when the model ensemble does not encapsulate the hourly-averaged CO_2 measurements, one of the many other non-transport error types could be to blame; the ensemble spread only encompasses transport errors and does not include measurement errors, errors due to finite-model resolution, or errors in the fluxes. Furthermore, these instances could be due to structural differences between CAM and TM5, including differences in model resolution. The Supplement provides more example CO_2 model–data comparisons, meteorology model validation, and data assimilation diagnostics.

7.3.2 CO_2 transport uncertainties at longer timescales

The uncertainty in monthly-averaged CO_2 concentrations provides one measure of how transport errors persist over time, as discussed in Sect. 7.2.4. Figure 7.2c and d display uncertainties in the month-long average surface concentrations for February and July 2009. In contrast to the 6-hourly uncertainties, these uncertainties are far more spatially distributed. This result implies that CO_2 transport errors covary over longer periods of time in remote regions, compared to regions with large surface fluxes. Observation sites that are far from large fluxes are therefore more likely to produce a biased CO_2 budget than sites near to large surface fluxes. These “remote” sites see a lower CO_2 signal from surface fluxes, and the transport errors at these locations generally covary over longer periods of time.

A number of factors may explain these relatively large error covariances in remote regions. CO_2 transport over remote or oceanic regions is likely dominated by synoptic-scale weather patterns that evolve over multi-day time periods. When CO_2 is transported across the oceans or remote areas from source/sink regions, atmospheric CO_2 transport errors would likely covary at timescales characteristic of this synoptic-scale air flow. Over large CO_2 source/sink regions, by contrast, atmospheric concentrations are likely influenced more strongly by processes that occur over smaller time periods – grid-scale winds or boundary layer mixing. In addition, sustained transport errors over regions of large biosphere flux would be more likely to cancel out at longer timescales – due to the diurnal cycle of biosphere CO_2 uptake and release.

In addition to remote and ocean regions, month-long transport uncertainties are also large across the entire Northern Hemisphere during February. A subsequent Sect. 7.3.4 explores possible reasons why these month-long biases occur.

7.3.3 Case study 1: the magnitude of temporally and spatially covarying atmospheric transport errors relative to a CO_2 flux estimate

We construct a case study to understand the importance of temporal and spatial error covariances relative to the magnitude of CO_2 surface fluxes. Figure 7.4 displays the results of this analysis for a selection of representative global CO_2 observation sites from Asia, Europe, and North America. The y axis of each bar plot indicates the difference between the top and bottom of the 95 % confidence interval in monthly mean modeled concentrations. We first consider the results when covariances in atmospheric CO_2 transport errors are included in the analysis (dark blue bars) and then compare those results to a setup in which we remove these error covariances (light blue bars).

At this selection of sites, uncertainties in the monthly mean afternoon concentrations range from 1.6 to 2.8 ppm (dark blue bars). These uncertainties are lower at marine sites like RYO and TTA (see definitions in Fig. 4) and are higher at continental sites located near large biospheric fluxes, sites like FSD and WBI. Note that this analysis only considers estimated uncertainties due to meteorology. The capabilities of the atmospheric observations would deteriorate if other errors were included, such as those due to imperfect measurements or due to finite-model resolution (e.g., Gerbig et al., 2003; Masarie et al., 2011).

We subsequently remove temporal covariances in the errors to identify the role that these covariances play in CO_2 transport uncertainties at the monthly scale. These results are displayed as light blue bars in Fig. 7.4. When we remove the covariances, the monthly-scale uncertainties are much smaller – by a factor of 5–20 at the individual observation sites. If CO_2 transport errors were temporally independent, then errors of opposite sign and different magnitude would cancel out to a degree when averaged over 1 month (light blue bars). Instead, the transport errors estimated by CAM-LETKF covary in time, and this covariance prevents the errors from averaging down (dark blue bars).

A multi-site comparison in Fig. 7.4 additionally indicates the role of spatial covariances in the transport errors; the figure shows the uncertainties in CO_2 concentrations when averaged across multiple observation sites within an eco-region. We compute the monthly-average afternoon concentration across multiple sites for a given ensemble member. We then estimate a confidence interval based upon the distribution of the 64 ensemble members.

The results indicate a large degree of spatial covariance in the atmospheric CO_2 errors. If the errors had no spatial covariance, these errors would average down as more and more observation sites were added to the analysis. However, the dark blue bars in Fig. 7.4 have a similar magnitude irrespective of whether the analysis was conducted on an individual site or on a collection of many sites from an eco-region; the errors must therefore covary in space. In contrast, the light blue bars (i.e., error covariances removed) do decrease in magnitude at the eco-region scale relative to individual observation sites. In that case, the errors do average out when more and more sites are included in the analysis.

Figure 7.5 places the results of case study one in the context of the surface fluxes. This figure displays the uncertainties in atmospheric CO_2 transport (the dark blue bars in Fig. 7.4) as a fraction of the mean afternoon CO_2 boundary layer enhancement. As discussed in Sect. 7.2.4, this enhancement approximates the CO_2 increment due to regional surface fluxes, and a similar CO_2 increment is used by a number of top-down studies to estimate the surface fluxes. At the individual observation sites, the uncertainty in atmospheric CO_2 constitutes 13–150 % of the average boundary layer CO_2 enhancement. This percentage is highest at marine sites like RYO and TTA that see a relatively small boundary layer enhancement, and the relative magnitude of the uncertainties is smallest at sites that see a very large enhancement due to large summertime vegetation fluxes (e.g., at the WBI site). The uncertainties due to atmospheric transport are substantial relative to the fluxes but only when we include covariances in transport error. When we remove these covariances, the uncertainty in monthly-average afternoon concentrations drops to only 2–22 % of the boundary layer enhancement.

The results of this analysis hold several implications for future atmospheric inverse models and/or top-down studies that optimize CO_2 fluxes. Most existing inverse models account for atmospheric CO_2 transport errors in their statistical setup. In a Bayesian synthesis or geostatistical inverse model, for example, this information is incorporated into a covariance matrix, and that covariance matrix is used as an input to the equation that optimizes the CO_2 fluxes (e.g., Enting, 2002; Michalak et al., 2004; Ciais et al., 2011). However, the majority of existing studies assume that this covariance matrix is diagonal (i.e., no error covariances), in part, because these temporal and spatial covariances are challenging to estimate (e.g., Lin & Gerbig, 2005; Lauvaux et al., 2009). The present study, in contrast, indicates that both temporal and spatial error covariances play an important role in monthly-scale errors in atmospheric transport.

Ignoring these error covariances could lead to numerous challenges. When we add more data at an observation site or add more sites the analysis, the actual errors do not average down to the extent that uncorrelated errors would. Rather, adding more data or more observation sites provides more limited gains in accuracy. As a result, an inverse model that overlooks the error covariances will estimate uncertainties in the CO_2 fluxes that are too small, and/or the inverse model may erroneously map atmospheric transport errors onto the surface fluxes (e.g., Stephens et al., 2007). Future inverse modeling studies could better account for these uncertainties by including off-diagonal terms in one of the covariance matrices used by the inverse model.

The next case study (Sect. 7.3.4) explores the meteorological factors that may be associated with these persistent atmospheric transport errors.

7.3.4 Case study 2: which meteorological factors are associated with month-long atmospheric transport biases?

In this case study, we use a synthetic tracer experiment (Sect. 7.2.5) to uncover possible drivers of atmospheric transport biases at month-long timescales. The previous section (Sect. 7.3.3) explored the importance of covariances in atmospheric CO_2 transport errors, and this section investigates the meteorological conditions associated with these persistent errors.

Figure 7.6 displays the CV for monthly-averaged surface concentrations of the synthetic tracer. The CV, a unitless quantity, does not just indicate where the uncertainties are largest. Rather, the CV indicates the magnitude of these uncertainties relative to the mean modeled tracer concentration. Arguably, this noise-to-signal ratio measures the influence of transport uncertainties more effectively than a simple standard deviation.

This coefficient shows a number of distinctive seasonal and spatial patterns. Like the uncertainties in monthly-averaged CO_2 (Fig. 7.2c, d), the CV in Fig. 7.6 is highest in terrestrial boreal and arctic regions of the Northern Hemisphere during winter. The CV is lowest over Europe, Australia, and the Amazon during all seasons.

The CV in Fig. 7.6 exhibits different spatial patterns over land and ocean regions, and these respective patterns correlate with different sets of meteorological variables. Over the oceans, for example, high CV values in Fig. 7.6a are clustered in zonal bands – along the Equator and along 40° S. In contrast, high CV values do not cluster into zonal bands to the same degree over terrestrial regions. Rather, CV values are often high when temperatures are low (e.g., over Canada or Russia in February).

We plot the synthetic tracer CV against numerous modeled meteorological parameters to further understand the possible drivers behind atmospheric transport uncertainties averaged over these monthly timescales. To this end, we examine correlations between the tracer CV and 60 different meteorological parameters, including the uncertainties in the meteorological variables. Figure 7.7 displays the two variables that correlate most strongly with the tracer CV over land regions and over ocean regions, respectively.

Over land regions, meteorological conditions that lead to high atmospheric stability and low energy are most closely associated with atmospheric transport errors. For example, a high tracer CV is associated with low temperatures ($R^2 = 0.45$) and low specific humidity ($R^2 = 0.40$). Similarly, a high tracer CV is correlated with low net solar flux ($R^2 = 0.35$), low planetary boundary layer height ($R^2 = 0.33$), and low vertical diffusion diffusivity ($R^2 = 0.31$). Note that many of these meteorological variables are closely related to one another, so the individual correlations listed above are all interrelated.

In addition, several of the meteorological variables exhibit a nonlinear relationship with the tracer CV, and the potential for bias in modeled atmospheric transport increases more quickly in stable atmospheric conditions. For example, the CV increases more quickly when planetary boundary heights are low.

In contrast to land regions, the tracer CV over the oceans is most closely associated with low zonal wind speeds ($R^2 = 0.29$, Fig. 7.7). Over land regions, that correlation is zero. Uncertainties in atmospheric transport over the oceans are also associated with low PBL heights ($R^2 = 0.25$). These two meteorological variables explain different patterns in the tracer CV; PBL heights and zonal wind speeds over the ocean are not correlated with one another ($R^2 = 0$), so these two parameters may indicate different processes underlying the atmospheric transport errors.

These differences between land and ocean regions may reflect differences in synoptic-scale circulation. Over the oceans, high CV values are clustered in zonal bands, and these clusters often occur at the transition between distinctive synoptic flow patterns. Modeled atmospheric tracer transport is more uncertain in these transition regions – at the transition between southern westerlies and southern trade winds and at the transition between the North Atlantic trade winds and the westerlies. Zonal winds over the continents are often more variable than over the oceans (Fig. S17 in the Supplement), and atmospheric transport uncertainties do not cluster into the same, distinctive, zonal bands.

The results of this synthetic tracer experiment hold a number of potential applications to top-down CO_2 flux estimation. The danger of obtaining a biased CO_2 budget is likely higher in regions with consistent low energy and limited vertical mixing. A number of existing studies indicate that uncertainties in PBLH and vertical mixing are closely tied to uncertainties in estimated trace gas transport or in estimated trace gas fluxes (e.g., Stephens et al., 2007; Williams et al., 2011; Miller et al., 2012b; Pino et al., 2012; Kretschmer et al., 2012). This study further suggests that sustained transport errors due to PBLH are more likely in regions or at times when PBL heights and mixing are consistently low. The meteorological model ensemble is not necessarily more uncertain in these regions (see Figs. S15–S16 in the Supplement). Rather, the extent to which meteorological uncertainties translate into tracer transport uncertainties appears to depend, at least in part, on the stability and net energy input associated with the boundary layer.

In summary, boundary layer energy and height explain some of the patterns in the estimated transport errors, but other patterns are associated with uncertainties in synoptic flow and are not related to a single meteorological parameter. In fact, over both terrestrial and oceanic regions, individual meteorological parameters only explain a maximum of 29–45 % of the variability in the tracer CV. This result stresses the utility of a meteorological model to calculate the variances and covariances in atmospheric transport errors rather than relying on a single, meteorological proxy.

Note that this study does not account for uncertainties in bottom-up, biogeochemical flux models due to uncertainties in driving meteorological variables. For example, process-based, biogeochemical models of CO_2 typically require estimates of meteorological variables like humidity, temperature, or precipitation to compute the surface fluxes. A number of existing studies have used atmospheric data and/or atmospheric models to explore the meteorological variables that drive CO_2 flux models. For example, Lin et al. (2011) explored how uncertainties in flux model drivers affected fluxes estimated for Canadian boreal forests. They found that uncertainties in downward shortwave radiation contributed to the largest uncertainties in the simulated fluxes. Similarly, numerous studies indicate that both air temperature and humidity are drivers of CO_2 fluxes (e.g., Law et al., 2002; Gourdji et al., 2012). These meteorological variables (e.g., downward shortwave radiation, temperature, and specific humidity) correlate with the persistent atmospheric transport uncertainties discussed earlier in this section. A future study could connect these uncertainties (in the biogeochemical model and in atmospheric transport) to gain an even broader picture of how meteorological uncertainties affect CO_2 flux modeling and ultimately top-down CO_2 flux estimates.

7.4 Conclusions

We use CAM-LETKF to explore the characteristics of correlated or covarying atmospheric CO_2 transport errors and the implications of those errors for CO_2 flux estimates. The first case study examines the relative magnitude of these errors at the monthly timescale. At this scale, error covariances play a critical role in the uncertainties in modeled atmospheric CO_2; we find that uncertainties increase by a factor of 5–20 at individual CO_2 observation sites when we include the error covariances in the analysis. These monthly-scale errors correspond to 13–150 % of the afternoon CO_2 boundary layer enhancement, depending on the site in question.

Existing top-down studies often overlook these covariances, and these results imply that atmospheric CO_2 measurements contain less information about the fluxes than is often assumed. As a result, existing inverse models may underestimate the uncertainties in estimated CO_2 fluxes and/or may be vulnerable to unforeseen biases in the estimated fluxes. Accounting for these correlated errors can be as simple as modifying one of the covariance matrix inputs in a Bayesian inverse model.

In a subsequent case study, we investigate the meteorological factors associated with month-long biases in atmospheric transport. The largest short-term CO_2 transport errors correlate strongly with the location of the largest surface fluxes, but month-long biases in atmospheric transport are not only localized to regions with large fluxes. Rather, these biases may be more likely to occur at observation sites that are far from large fluxes and in regions with high atmospheric stability and low net radiation. Over the oceans, biases in atmospheric transport are also associated with weak zonal winds. Existing top-down flux studies may be more likely to estimate inaccurate regional fluxes under those conditions. However, a large fraction of the estimated atmospheric transport errors cannot be described by a single meteorological parameter. This result indicates the utility of a meteorological modeling system, like CAM-LETKF, to estimate errors in atmospheric CO_2 transport. Through this framework, we can better understand the connections between uncertain atmospheric transport and uncertainties in CO_2 budgets estimated from atmospheric data.

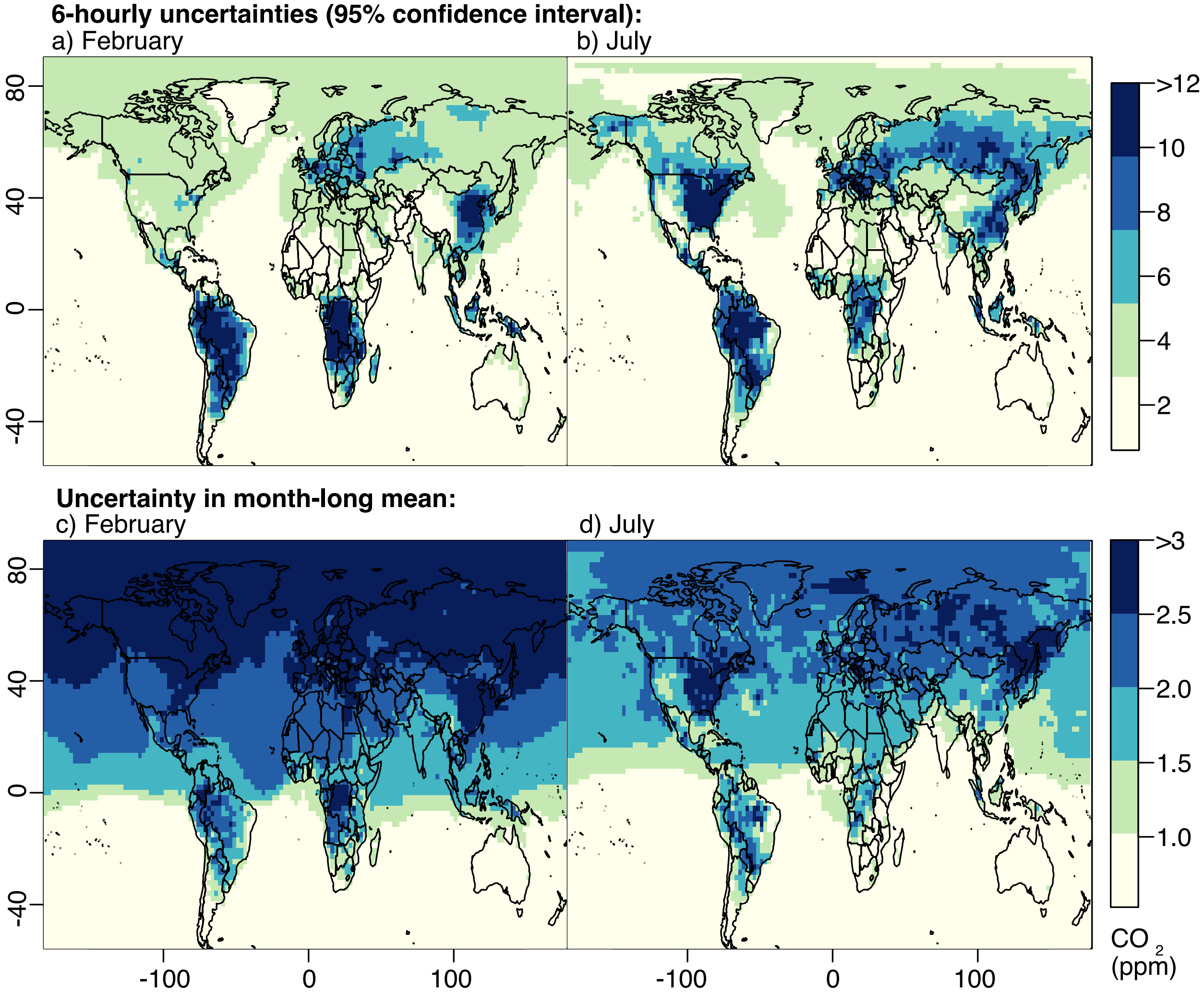


Figure 7.2: The top panels display average 6-hourly CO_2 transport uncertainties estimated by CAM-LETKF. The uncertainties (95 % confidence intervals) are for the surface model layer for (a) February and (b) July 2009. The bottom panels (c and d), in contrast, display the uncertainties in month-long averaged surface CO_2 concentrations. Note that these plots include model output from all 24 h of each day. The Supplement provides analogous figures for daytime- or nighttime-only model output.

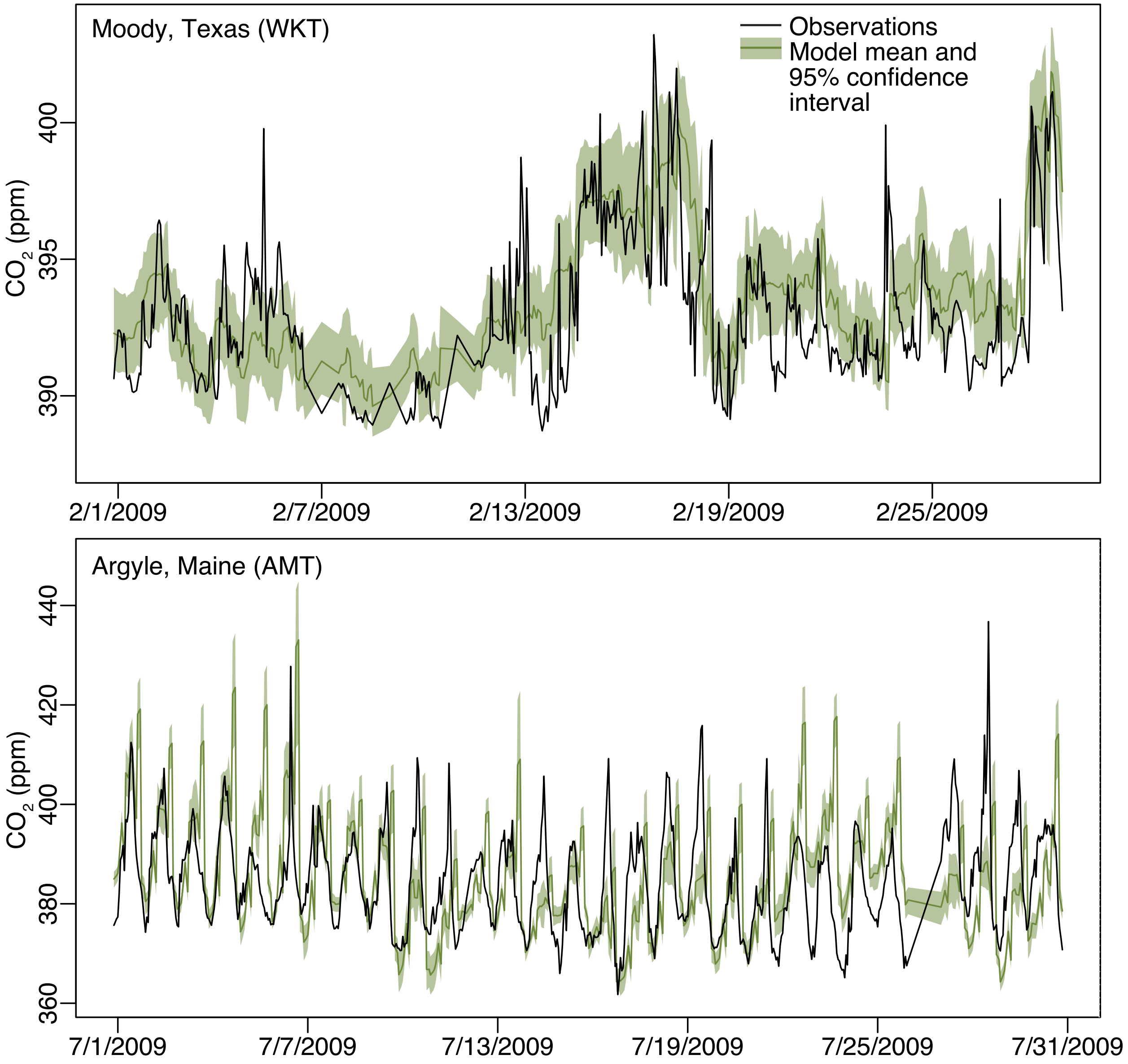


Figure 7.3: Hourly-averaged CO_2 measurements at (a) Moody, Texas, and (b) Argyle, Maine, compared against the CAM-LETKF model ensemble. Measurements are from the top inlet height at each location. In this figure, the model ensemble represents uncertainties due to atmospheric transport but not other errors (e.g., due to the fluxes and model resolution).

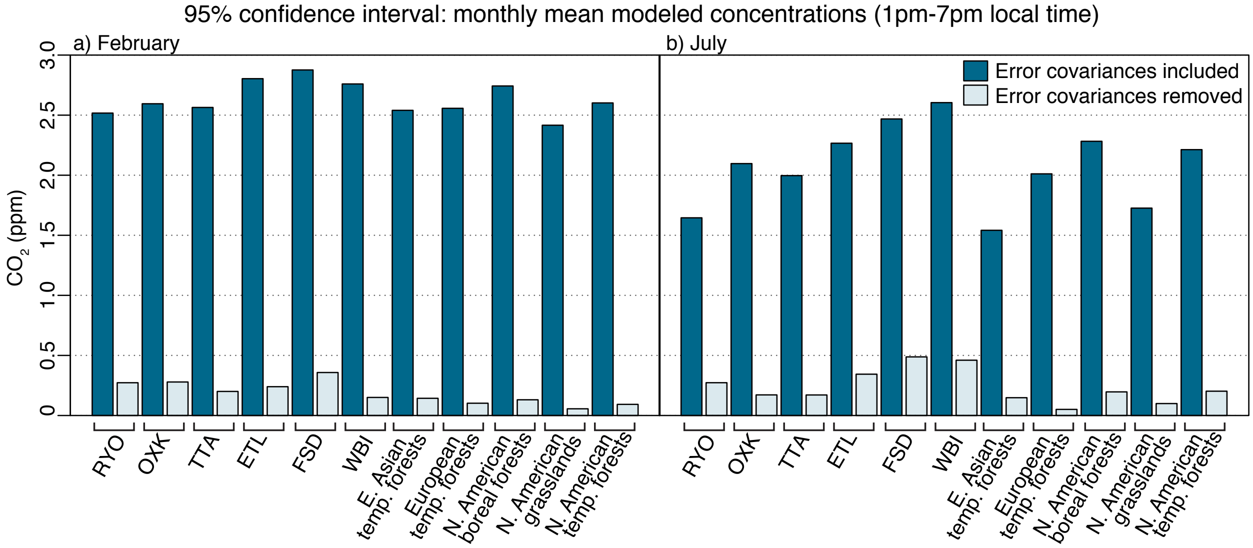


Figure 7.4: The uncertainties in monthly-averaged, afternoon atmospheric CO$_2$ (Sects. 7.2.4, 7.3.3) at a selection of representative, global CO$_2$ observation sites. Panels (a) and (b) show the results at each site for February and July 2009, respectively. Dark blue bars indicate the difference between the top and bottom of the 95 % confidence interval when we include error covariances. The light blue bars indicate the results when we remove these covariances in atmospheric transport errors. Observation sites in the figure include Ryori, Japan (RYO); Ochsenkopf, Germany (OXK); Talk Tower Angus, UK (TTA); East Trout Lake, Saskatchewan, Canada (ETL); Fraserdale, Ontario, Canada (FSD); and West Branch, Iowa, USA (WBI). For more information on these observation sites, refer to Table S1 in the Supplement.

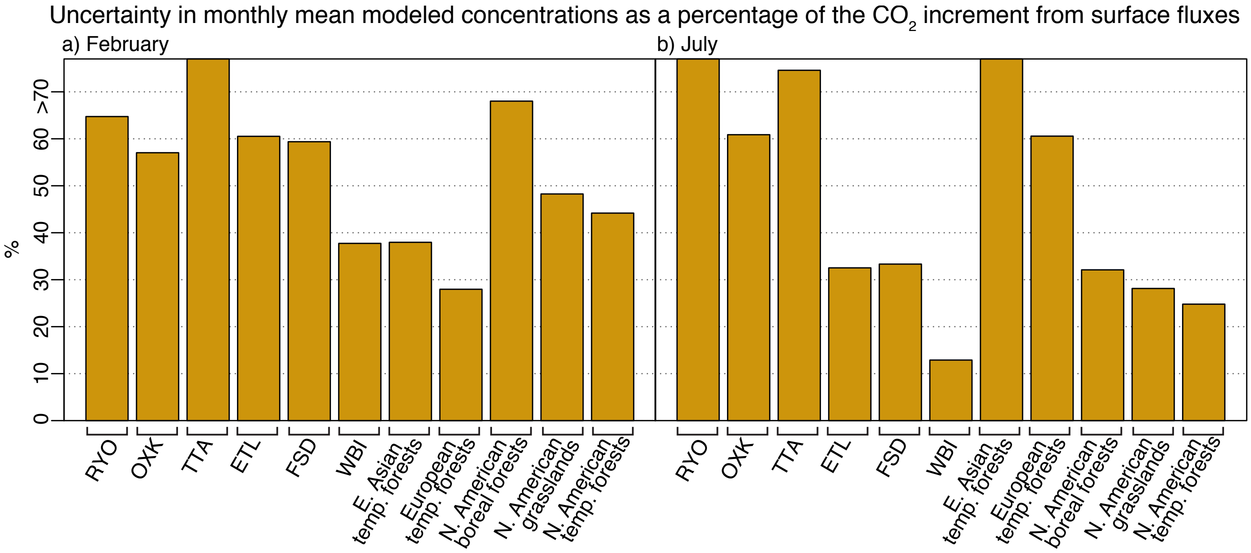


Figure 7.5: Uncertainty in monthly-averaged afternoon CO_2 concentrations as a percentage of the average afternoon CO_2 boundary layer enhancement. This figure places the uncertainties from Fig. 7.4 (dark blue bars) in context of the afternoon CO_2 increment from surface fluxes. Larger percentages indicate greater potential for bias in monthly CO_2 budgets estimated from atmospheric data.

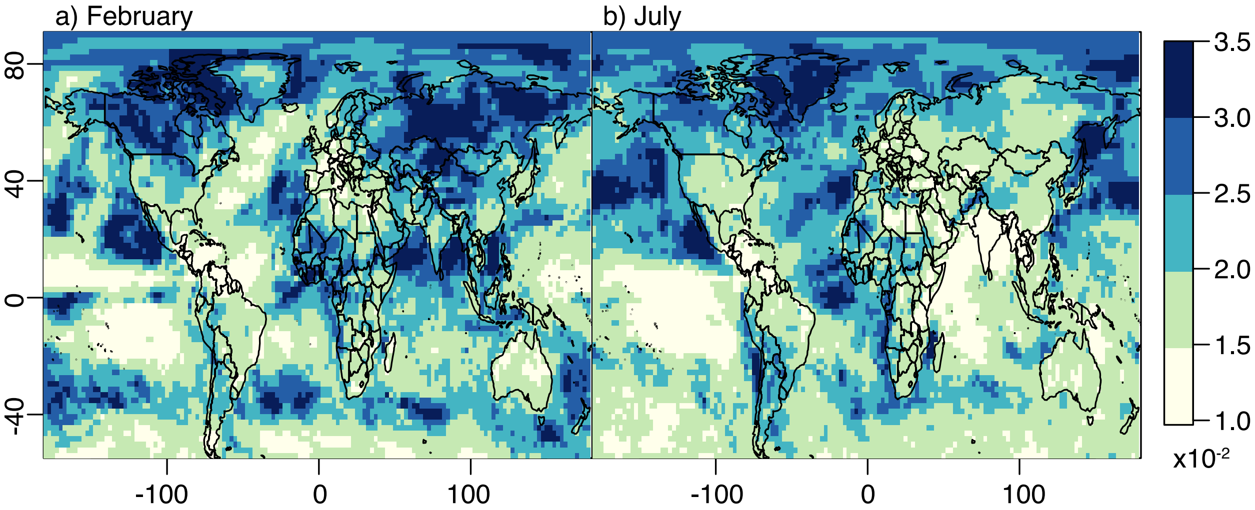


Figure 7.6: The coefficient of variation (CV, unitless) for the monthly-averaged model surface layer. The results plotted here are for the synthetic tracer simulation (Sects. 7.2.5, 7.3.4). In that simulation, the synthetic fluxes have a constant spatial distribution. The resulting CV (σ / μ) shows the distribution of month-long, surface-level transport uncertainties independent of the spatial distribution in the fluxes.

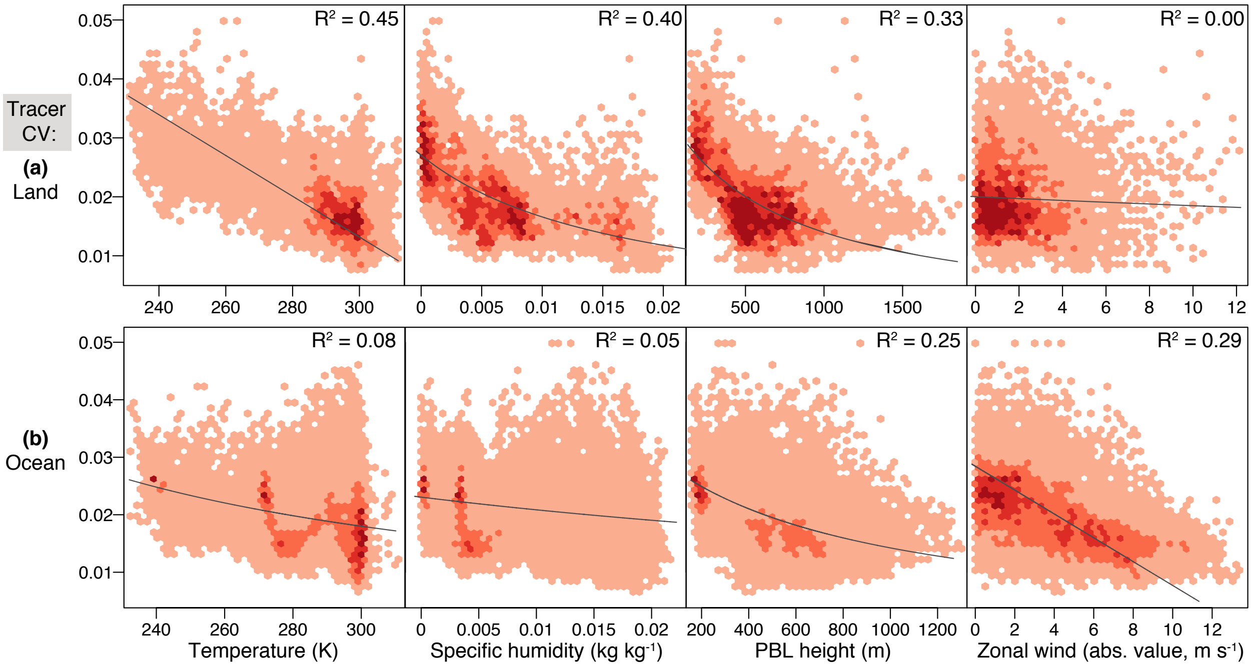


Figure 7.7: Each panel shows the relationship between the synthetic tracer CV (Fig. 7.6) and various monthly-averaged meteorological parameters estimated by CAM-LETKF. The top row (a) shows the results for terrestrial regions while the bottom row (b) displays the results for ocean/marine regions. Darker colors in each panel indicate a higher density of points. We test the correlation with 60 different parameters (Table S2 in the Supplement) and plot the two parameters that correlate most closely with the tracer CV over terrestrial and marine regions, respectively. In all cases, we fit both a standard major axis regression and nonlinear least squares ($\frac{1}{[\beta_1 \times \text{parameter} + \beta_2]}$) and plot the regression with the higher correlation coefficient.

8

Conclusion

In this thesis, we use atmospheric data to estimate the magnitude, spatial distribution, and seasonality of N_2O and CH_4 emissions over North America. We also leverage novel statistical techniques to infer information about the plausible source sectors or processes driving these emissions. We then analyze how those drivers differ from existing bottom-up estimates. Through these efforts, this research creates pathways to reconcile differences between bottom-up emissions estimates and the information contained within measured atmospheric concentrations.

We estimate anthropogenic N_2O and CH_4 emissions that greatly exceed most existing inventory estimates. In contrast, we estimate natural CH_4 sources that are smaller than most existing process-based biogeochemical models. My results indicate that total N_2O emissions in the US are likely 2.5 to 3.1 times higher than the commonly-used, bottom-up US Environmental Protection Agency (EPA) inventory. Furthermore, we find anthropogenic methane emissions that are approximately 1.5 times existing EPA estimates.

We further examine seasonal variability in the emissions. N_2O sources over the US are highly seasonal and peak early in the growing season. For example, emissions in June are approximately 4 – 7 times larger than emissions in winter months. In the case of methane, we estimate a seasonal cycle in wetland methane emissions that is broader than in most existing biogeochemical models. These existing models appear to shut off methane emissions prematurely during the fall season and may be too sensitive to surface temperatures, among possible explanations. US anthropogenic methane emissions, by contrast, show no significant seasonal cycle.

We also estimate a spatial distribution of N_2O and CH_4 emissions using atmospheric data, and these estimates often differ from the distribution in existing bottom-up inventories or models. We find the largest N_2O emissions in Iowa and California where fertilizer use is highest, not over Gulf Coast wetlands where the DLEM inventory predicts the largest emissions. These results imply that fertilizer application, not soil moisture/temperature, is the primary driver of N_2O release from soils.

Our methane estimates also have a different spatial distribution from existing inventories. We find that three states alone account for $\sim$25% of all US methane emissions – Texas, Oklahoma, and Kansas. Atmospheric methane measurements in those states correlate strongly with propane, a tracer of the oil and gas industries. To that end, this thesis indicates that methane leaks from the oil and gas industries far exceed US government estimates that are based upon limited measurements at individual well sites. In addition, we find that existing inventories likely underestimate methane emissions from ruminants and manure by a factor of two. In summary, our results indicate a very different magnitude and distribution of anthropogenic methane emissions in the US than previously estimated.

Over wetland regions, we find that most existing biogeochemical methane models contribute negative information on the spatial distribution of wetland fluxes; only 14-33% of existing biogeochemical models describe the distribution of fluxes more effectively than a spatial constant (as seen through atmospheric observations). The best models estimate wetland coverage using land cover maps, instead of using remote sensing inundation datasets. We also built a relatively simple methane flux model that shows better agreement with available atmospheric observations than the most sophisticated or complete process models. These results, using a model-data fusion, can help reconcile large existing uncertainties in methane fluxes from temperate and boreal wetlands.

New and forthcoming N_2O and CH_4 observations promise to create a wealth of opportunities for future work. These observations include more than 20 US, tower-based sites from a privately-funded effort by Earth Networks, new observation sites operated by Environment Canada, and new remote sensing efforts that include the TROPOMI satellite instrument. An important future challenge will be to leverage this wealth of data in a way that can provide information about individual source sectors or flux drivers, information that can be compared directly against the components of an emissions inventory. In March 2014, the Obama administration announced the Climate Action Plan for Methane and has recently developed targets for the oil and gas industries. These regulations will require independent means to evaluate methane emissions, particularly approaches that can provide reliable information on individual sources. This thesis is a stepping-stone toward that long-term goal.

A

Supplemental material for chapter 2

This supplement contains further explanation of the modeling and statistical methods and provides additional model validation.

A.1 Atmospheric modeling approach

A.1.1 Transport model overview

This study utilizes STILT, the Stochastic Time-Inverted Lagrangian Transport model, for all atmospheric transport simulations Lin et al. (2003). Model runs use an ensemble of 500 particles followed 10 days back in time. The methane increments computed from continental surface sources are added to the methane boundary condition, the concentration of methane in air masses before being influenced by emissions in North America.

The model equation can be written as Gerbig et al. (2003)

$$z = \mathbf{H}s + \varepsilon \tag{A.1}$$

where z (dimensions $n \times 1$) is the contribution of continental sources to the observation sites, and s ($m \times 1$) are the true, unknown methane emissions. Any estimate of the unknown emissions (s) is termed $\hat{s}$. The total methane concentration measured at the tower or aircraft is given by $z + b$, where b ($n \times 1$) is the boundary condition. The influence footprint $\mathbf{H}$ gives the concentration enhancement at the measurement site due to unit emission flux from each grid cell. The footprint has units of concentration per surface flux, or ppb per μmol m^{-2}s^{-1}. Each row of $\mathbf{H}$ ($n \times m$) is the footprint associated with an individual methane measurement. Finally, ε ($n \times 1$) describes model-data mismatch errors, all model or measurement errors that are unrelated to an imperfect emissions estimate. In other words, this mismatch remains the same irrespective of the emissions estimate ($\hat{s}$) used in the model. The mismatch includes, but is not limited to, errors in modeled transport, the methane boundary condition (b), and the methane measurements. Common inversion frameworks based on Gaussian statistics, including this one, assume that all model-data mismatch errors (ε) are random with a mean of zero and a covariance described by the $n \times n$ matrix $\mathbf{R}$.

STILT trajectories are driven by Weather Research and Forecasting model (WRF, version 2.1.2) meteorological fields Skamarock et al. (2005); Nehrkorn et al. (2010). Our WRF simulations consist of sequential 30-hour meteorological forecasts initiated daily from NARR (North American Regional Reanalysis). All simulations include convection using a Grell-Devenyi scheme Grell & Devenyi (2002). These wind fields use a nested resolution: 10-km over most of the continental United States and 40-km over other North American regions.

A.1.2 Methane boundary condition

The boundary condition (b) could be constructed either from interpolated measurements or from the output of a global chemical tracer model (such as Geos-Chem). We choose the former approach due to uncertainties associated with the global distribution of methane emissions.

We construct the boundary condition using a two stage process. First, we use an empirical methane boundary curtain over the Pacific Ocean as an initial guess for b. This western curtain consists of NOAA measurements near or over the Pacific Ocean, interpolated latitudinally, vertically, and daily using a curve-fitting procedure Thoning et al. (1989). The individual trajectories in every 500-trajectory STILT simulation typically end at different locations and reach the western boundary at different latitudes, times, or elevations. The mean statistics of the trajectory ensemble at the western curtain provide the initial value for b. Most of the STILT back-trajectories in this study reach the Pacific coastline less than 10 days after leaving the observation site (64%). 100% of trajectories originating from the WGC site, 83% of those originating at BAO, 60% at WBI, and 60% from WKT reach the Pacific Ocean. The Martha's Vineyard and Argyle, Maine, sites have the lowest fraction of trajectories reach the Pacific Ocean (37% and 32%, respectively), though many of the remaining trajectories never exit the continent during the 10 day span of the trajectory run.

Second, we use NOAA aircraft observations over North America in the free troposphere (>3000m) to fit the initial boundary estimate to regional conditions or airflow patterns. The adjustment is most relevant for observation sites farther from the western curtain (e.g., Massachusetts and Maine). We calculate, for different regions and seasons, the mean model-observation difference above 3000m using the initial boundary guess and EDGAR v4.2. Regions include the western, eastern, south-central, and north-central portions of the United States over winter, spring, summer, and fall. The purpose of this adjustment is twofold. First, it accounts for the inflow of "clean background" air that may enter a region outside the prevailing westerlies. Second, it accounts for the small amount of methane oxidation that may occurr en route between the western boundary curtain and the methane measurement sites. This aircraft-based adjustment has a mean of -2.7 ppb and a maximum magnitude of -7 ppb. An inversion using the initial western boundary curtain without the regional adjustment estimates methane budgets of 32.0 and 7.7

TgC yr^{-1} for the US and Texas-Oklahoma-Kansas, respectively, within about 5% of the final best estimate in the main article.

The boundary condition exhibits marked seasonality, with an average 40 ppb peak in winter. This peak reflects large-scale seasonal changes in Northern Hemisphere clean-air concentrations.

WRF-STILT does not explicitly model atmospheric oxidation processes. We fit the boundary condition to local or regional free troposphere values, eliminating the need to consider longer range oxidation chemistry. Furthermore, the footprint ($\mathbf{H}$) is greatest within 2–3 days of the associated measurement. Methane has a global-averaged lifetime of 7–11 years Prather et al. (2012); Ciais et al. (2013), implying methane decay of less than 1–1.5 ppb over these 2–3 day time scales.

A.1.3 Study time period

We choose 2007 – 2008 as the time frame of this study for two reasons. First, there are no daily methane measurements from US tower sites before mid-2006, with the exception of the Niwot, Colorado, sites. Weekly to monthly methane observations are available at some sites prior to 2006. Second, the WRF meteorology fields used here are only available through 2008. These fields are validated by Nehrkorn et al. 2010 Nehrkorn et al. (2010) and used in a number of existing studies Kort et al. (2010); Huntzinger et al. (2011); Gourdji et al. (2012); Miller et al. (2012b), but they have a limited time scope.

A.1.4 Observations

We use diverse methane measurements taken at tall towers or by aircraft during 2007-2008. Measurements include daily flask samples from the NOAA tall tower network (weekly at Argyle and Ponca City): Argyle, Maine (AMT, 45 °N, 69 °W, 107m above ground level (agl)); Erie, Colorado (BAO, 40 °N, 105°W, 300m agl); Park Falls, Wisconsin (LEF, 46°N, 90°W, 244m agl), Martha's

Vineyard, Massachusetts (MVY, 41°N, 71°W, 12m agl); Niwot Ridge and Niwot Forest, Colorado (NWF, NWR, 40°N, 105°W, 2,3,23m agl); Ponca City, Oklahoma (SGP, 37°N, 97°W, 60m agl); West Branch, Iowa (WBI, 42°N, 93°W, 379m agl); Walnut Grove, California (WGC, 38°N, 121°W, 91m agl), and Moody, Texas (WKT, 31°N, 97°W, 122, 457m agl) (see main article Fig. 2). The inverse model incorporates surface data and aircraft measurements up to 2500m agl;. observations at higher altitudes are less sensitive to surface emissions and are reserved for model validation and adjustment of the boundary condition. The flask and aircraft data are sampled only during the daytime hours, so this study is not affected by the large uncertainties associated with modeling the nocturnal boundary layer Matross et al. (2006); Gourdji et al. (2010).

A.2 Statistical methods

We utilize a geostatistical inverse modeling (GIM) framework to estimate monthly methane emissions (s) for 2007 and 2008 on a 1° x 1° latitude-longitude grid Kitanidis & Vomvoris (1983); Snodgrass & Kitanidis (1997); Michalak et al. (2004):

$$s = \mathbf{X}\beta + \mathcal{N}(0, \mathbf{Q}) \tag{A.2}$$

The GIM uses a deterministic model ($\mathbf{X}\beta$) for the prior estimate of emissions, similar to a multiple regression. Each column of $\mathbf{X}$ is a different spatial dataset, including a column for a constant component, and β is the vector of associated unknown coefficients. This differs from a Bayesian synthesis inversion, which typically uses a prior with a static, known magnitude Rodgers (2000); Tarantola (2005).

The GIM also has a stochastic component, described by a multivariate normal distribution $\mathcal{N}$ with a mean of zero and covariance matrix $\mathbf{Q}$. This component describes all emissions that do not fit the spatial pattern of the deterministic model. Unlike most Bayesian synthesis inversions, the GIM accounts for spatial and/or temporal correlation (i.e., covariance) in the stochastic component by including off-diagonal terms in **Q**.

The GIM framework allows the atmospheric observations to determine the spatial patterns of both the deterministic and stochastic components. Also, the formulation ensures that the prior has no overall bias, an important statistical assumption of most inversion frameworks. A number of existing studies have used this approach successfully for trace gas surface flux estimation Michalak et al. (2004); Mueller et al. (2008); Gourdji et al. (2008, 2012); Miller et al. (2012b).

The best estimate of emissions (s) is typically the minimum of the geostatistical cost function Michalak et al. (2004):

$$\begin{aligned} L_{s,\beta,\theta} = & \tfrac{1}{2}\ln|\mathbf{Q}| + \tfrac{1}{2}\ln|\mathbf{R}| + \tfrac{1}{2}(z-\mathbf{H}s)^T\mathbf{R}^{-1}(z-\mathbf{H}s) \\ & + \tfrac{1}{2}(s-\mathbf{X}\beta)^T\mathbf{Q}^{-1}(s-\mathbf{X}\beta) \end{aligned} \tag{A.3}$$

This cost function, based on Gaussian statistics, cannot preclude large negative emissions, and we use Lagrange multipliers to enforce nonnegativity Gill et al. (1981); Miller et al. (2014a). Large negative emissions would be unrealistic for methane given that the soil sink is only ∼4% of global methane loss Dutaur & Verchot (2007). Any soil sink over the US would be far smaller than the posterior uncertainties and therefore not detectable by the inversion framework with any degree of certainty. Lagrange multipliers, the method used to enforce nonnegativity, is iterative and produces a robust estimate of the posterior emissions subject to physical bounds. However, the resulting posterior uncertainties are generally too large and should be interpreted with caution. A recent paper discusses this method in detail and the impact on the final emissions estimate Miller et al. (2014a).

A.2.1 Covariance matrix estimation

Restricted maximum likelihood (REML) provides an objective way to estimate the structure and magnitude of the error covariance matrices in the inversion (**R** and **Q**); it guarantees that the actual inversion residuals match against those predicted in the covariance matrices Kitanidis (1986); Michalak et al. (2004).

REML estimates the parameters (θ) that define **R** and **Q** by minimizing a modified form of the cost function in Eq. A.3. In practice, it may be difficult to estimate the covariance matrix parameters (θ) using Eq. A.3 directly because this function also depends on the unknown values of the fluxes (s) and the drift coefficients (β). The restricted likelihood integrates over all possible values of s and β in Eq. A.3. The integration effectively removes s and β from the cost function, and the function is subsequently reformulated only in terms of the covariance matrices and several known pieces of information (see Kitanidis (1986) for a full derivation):

$$
\begin{aligned}
L_\theta &= \int_\beta \int_s L_{s,\beta,\theta}\, ds\, d\beta \\
&= \tfrac{1}{2}\ln|\mathbf{\Psi}| + \tfrac{1}{2}\ln|\mathbf{X}^T\mathbf{H}^T\mathbf{\Psi}^{-1}\mathbf{H}\mathbf{X}| + \tfrac{1}{2}z\mathbf{\Xi}z && \text{(A.4)} \\
\mathbf{\Psi} &= \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R} && \text{(A.5)} \\
\mathbf{\Xi} &= \mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}^T\mathbf{\Psi}^{-1}\mathbf{H}\mathbf{X})\mathbf{X}^T\mathbf{H}^T\mathbf{\Psi}^{-1} && \text{(A.6)}
\end{aligned}
$$

The optimal covariance matrix parameters are those that minimize the above equation, typically estimated with an iterative Gauss-Newton algorithm. Kitanidis (1983b) originally developed REML for geostatistics; many subsequent studies from geostatistics and other fields indicate that REML is one of the most accurate and unbiased methods for estimating errors and/or the structural parameters of a statistical model Robinson (1987); Kitanidis (1987); McGilchrist

(1989); Wilson (1989); Cheang & Reinsel (2000); Lark & Cullis (2004); Lark et al. (2006). Among other advantages, it ensures that the weighted sum of squares residuals from the inversion will follow the expected χ^2 distribution Kitanidis (983b).

In this study, we construct the covariance matrix $\mathbf{R}$ as a diagonal matrix. To estimate the diagonal elements of $\mathbf{R}$ ($\sigma^2_{\mathbf{R}}$, the model-data mismatch variance), we first calculate the variance of the model–measurement residuals for each measurement site when the STILT model is run with the EDGAR v. 3.2 FT2000 anthropogenic emissions inventory. Two top-down studies find that this inventory has the correct overall magnitude over the US Kort et al. (2008, 2010), and thus we use this version of EDGAR over others as a starting point in error estimation. We then use REML to estimate a single scaling factor to align the initial estimate with the variances suggested by the atmospheric data.

The estimation of $\mathbf{Q}$ follows slightly different form. The setup here estimates a constant value for the a priori variance (the diagonal elements). In other words, we assume there is little spatial or temporal variability in the variance in the the stochastic component of the emissions estimate, an assumption that makes sense given large emissions in disparate regions of the US and apparent absence of large seasonality in anthropogenic sources. Other parameters of $\mathbf{Q}$ to estimate include l, the decorrelation length parameter and t, the decorrelation time parameter (where $3l$ and $3t$ are the total approximate decorrelation length and time, respectively). REML would not converge on a decorrelation length for the off-diagonal elements of $\mathbf{Q}$. This may be due to geographic heterogeneity in the correlation lengths of the stochastic component. We set a decorrelation length parameter (l) at 100km, a compromise between emissions uncertainties that might be correlated over the distance of a large urban area and uncertainties in agricultural emissions that may be correlated over a larger regional scale. Test inversions with $l = 50$, 300, and 500 km provide a measure of the sensitivity of the estimated fluxes to the choice of the decorrelation parameter. Ultimately, the choice of l has little impact on the total US anthropogenic budget (less

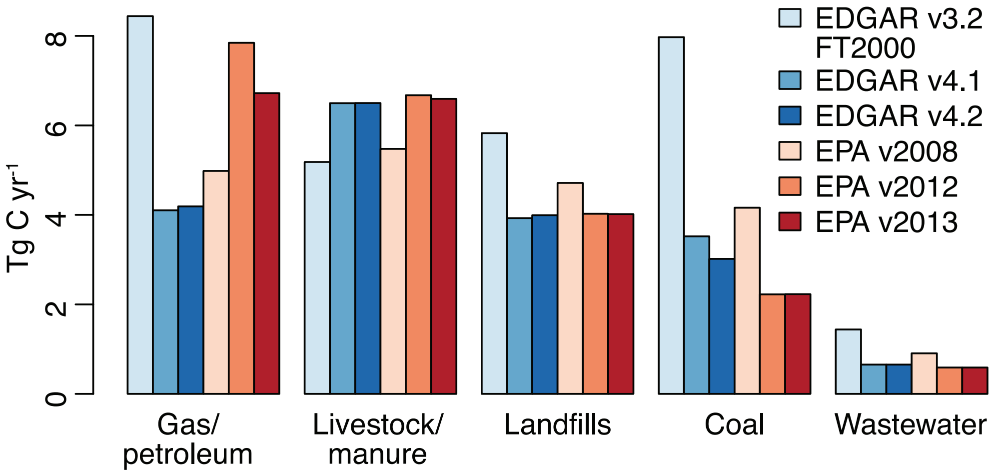


Figure A.1: US methane emissions by sector from several existing inventories. All methane budgets are from the year 2005 except EDGAR v3.2 FT2000 which estimates emissions from 2000.

than 1 TgC/yr).

Using REML, we estimate a variance in the stochastic component of 0.041 ± 0.001 μmol $\mathrm{m}^{-2}\mathrm{s}^{-1}$ (i.e., the square root of the diagonal elements of $\mathbf{Q}$). The decorrelation time and length parameters in the exponential covariance function are estimated at $t = 36 \pm 5$ days and $l = 100$ km, respectively.

A.2.2 The deterministic model

The GIM setup here adopts spatial activity datasets from the EDGAR inventory as predictors ($\mathbf{X}$) for the distribution of methane emissions (see Table 1 in the main article and the EDGAR web site [http://edgar.jrc.ec.europa.eu/]) and uses atmospheric data to estimate the associated emission factors (β). The emissions factors in existing inventories can be highly uncertain and have recently changed by up to 50% in EDGAR for sectors such as coal and natural gas European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL) (2010) (Fig. A.1).

We use a model selection method to assemble an optimal set of spatial activity datasets for the inversion. These methods will select as many predictors for use in **X** that can explain variability in the data but will prevent an over-fit or unreliable coefficient estimates Zucchini (2000); Konishi & Kitagawa (2008). We implement one of the most common methods, the Bayesian information criterion (BIC) Schwarz (1978); Mueller et al. (2010); Gourdji et al. (2012). The BIC numerically scores all possible combinations of activity datasets based on how well they improve goodness of fit (i.e., the log-likelihood of the model, similar to the weighted sum of squares) and applies an increasing penalty for model complexity. For each additional activity dataset, the penalty increases with the natural log of number of observations. The best candidate model (**X**) is the one with the lowest BIC score Gourdji et al. (2012):

$$\hat{\beta} = (\mathbf{X}^T\mathbf{H}^T\mathbf{\Psi}^{-1}\mathbf{H}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^T\mathbf{\Psi}^{-1}z \tag{A.7}$$

$$BIC = ln|\mathbf{\Psi}| + (z - \mathbf{H}\mathbf{X}\hat{\beta})^T\mathbf{\Psi}^{-1}(z - \mathbf{H}\mathbf{X}\hat{\beta}) + p\ ln(n) \tag{A.8}$$

where p is the number of predictors (number of columns in **X**), and n is the number of methane measurements.

The drift coefficients (β) in the model of the mean must be positive, since a spatial dataset should never contribute negatively to the posterior methane emissions. Hence we eliminate all candidate models from consideration that yield negative coefficients.

The BIC does not support hypothesis testing with p-values, but the difference in BIC scores provides a metric of confidence Kass & Raftery (1995). A score difference greater than 2 indicates notable evidence against the higher scoring model and a score increment greater than 10 indicates “very strong” evidence against that model.

Only two spatial datasets from EDGAR are identified by the BIC as important predictors of methane observations over the US:

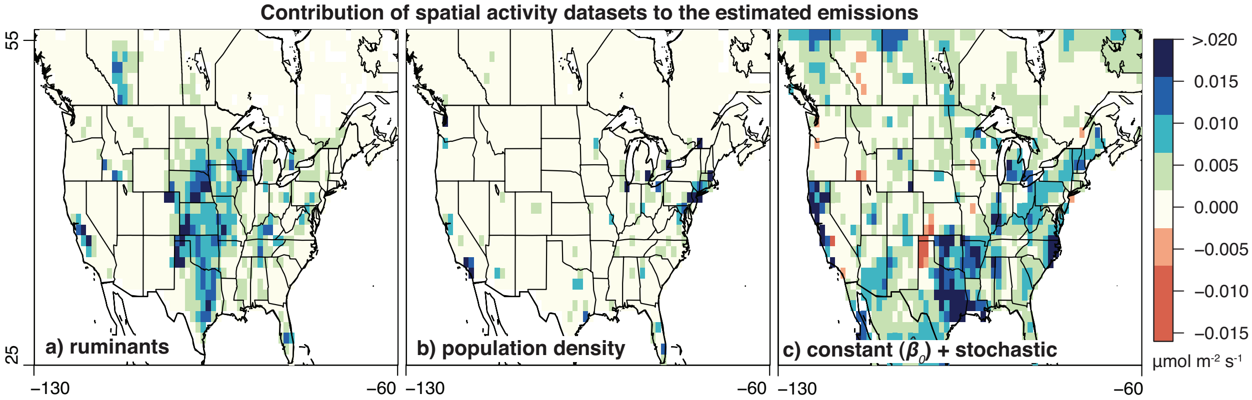


Figure A.2: The contribution of different activity datasets to the overall emissions estimate. The sum of the constant and stochastic components (c) includes all sources not described by the spatial activity datasets (the sum of a and b). The contribution of the activity datasets (a,b) is constant throughout the two years, and the stochastic component (c) changes by month (averaged in the plot here).

$$\beta_0 + \beta_1[\text{population density}] + \beta_2[\text{ruminant density}]$$

where β_0, β_1, and β_2 are the coefficients of the spatial activity data. The first term, β_0, represents the mean of all sources with spatial patterns other than population or ruminant density. Table 1 provides example BIC scores for this and other candidate models, including what are commonly considered the largest methane source sectors. The BIC scores strongly suggest that there is either insufficient data to include more than two activity datasets or that several existing activity datasets do not adequately describe the methane observations. In particular, the table indicates that the observation network is not sufficiently sensitive to coal sources and that the oil and gas production activity dataset from EDGAR do not accurately represent the spatial distribution of the methane emissions consistent with observations.

Fig. A.2 displays the methane budget from each of the spatial datasets in the deterministic model (e.g, scaled by the estimated coefficients, $\hat{\beta}$). It is important to note that population density serves as a proxy for a number of source sectors that are co-located with population (e.g., natural gas distribution, landfills, and wastewater treatment) at the 1^o spatial scale. Additionally, fuel

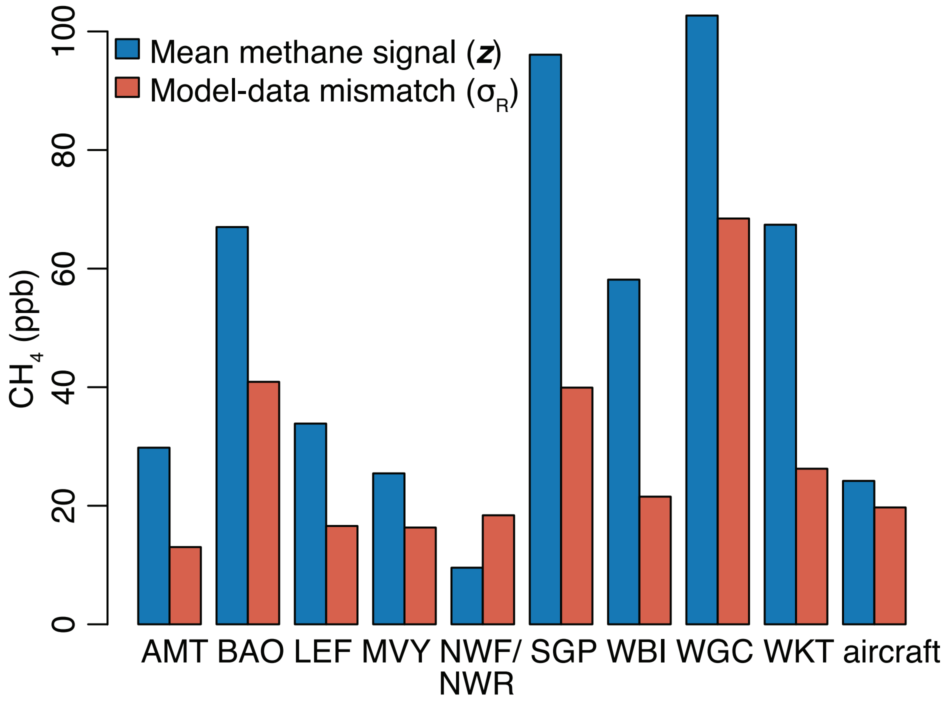


Figure A.3: A visualization of the estimated model-data mismatch across different methane measurement sites, compared against the two-year averaged contribution from anthropogenic emissions at each site (z, the concentration measurements minus the boundary condition and modeled wetland contribution).

extraction and animal husbandry are co-located over Texas and Oklahoma, so some emissions assigned to ruminant density (Fig. A.2a) may instead partially reflect fossil fuel industry sources.

A.3 Uncertainty analysis

A.3.1 General uncertainties in the model and measurements

Figure A.3 provides a visualization of the model-data mismatch errors estimated by REML ($\sigma_\mathbf{R}$, the standard deviation of ε). This mismatch includes random model and measurement errors unrelated to the emissions: errors in the wind fields, boundary layer height, methane boundary condition, and spatial/temporal aggregation, among other error sources. Model-data mismatch typically ranges from 40–70% of the total methane emissions signal seen at each tower, but the relative mismatch is higher at "clean air" sites like Niwot Ridge, Colorado (NWF/NWR). Absolute uncertainties are largest at measurement sites close to mountain ranges (e.g., BAO and WGC). This likely reflects difficulties in modeling wind fields near complex topography. Over Texas and Oklahoma, where the estimated anthropogenic methane emissions are among the largest in the US, the model-data mismatch is just under half the magnitude of the total average methane signal.

A.3.2 Uncertainties in the emissions estimate

The posterior covariance matrices (denoted **V**) provide a measure of uncertainty in the estimated emissions ($\hat{s}$) and estimated coefficients ($\hat{\beta}$) Michalak et al. (2004):

$$\begin{bmatrix} \mathbf{V}_{\hat{s}} & \mathbf{V}_{\hat{s}\hat{\beta}} \\ \mathbf{V}_{\hat{\beta}\hat{s}} & \mathbf{V}_{\hat{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} & \mathbf{Q}^{-1}\mathbf{X} \\ \mathbf{X}^T\mathbf{Q}^{-1} & \mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X} \end{bmatrix}^{-1} \tag{A.9}$$

Eq. A.9 is the inverse of the Hessian of the cost function (Eq. A.3). The posterior covariance matrix of $\hat{s}$, summed across different months and locations, produces the confidence intervals on the methane budgets listed throughout the paper. All uncertainties listed are 95% confidence intervals, unless otherwise noted.

The uncertainties in the emissions estimate vary depending on the temporal or spatial scales of interest. For example, the uncertainties can be larger than the estimated emissions at the 1° latitude–longitude spatial scale and monthly temporal scale. However, uncertainties decrease as the emissions estimate is averaged over larger regions and longer times (Fig. A.4). Intuitively, the uncertainties are higher at finer spatial/temporal scales because the atmospheric methane data have limited capacity to determine the precise location or time of grid-scale emissions. The

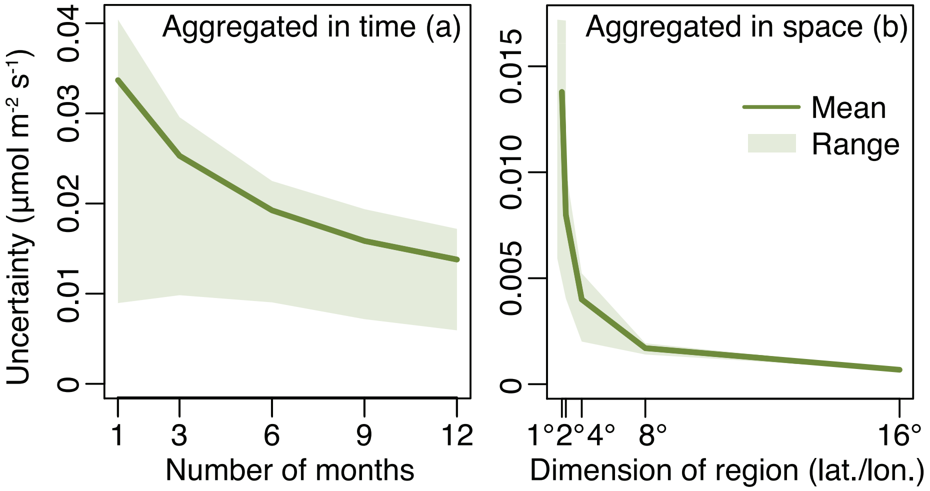


Figure A.4: The methane emissions estimate in this study has a 1° lat.–lon. spatial resolution and monthly temporal resolution. The uncertainty in the emissions estimate declines as the estimate is averaged to greater spatial and temporal scales. Panel (a) shows how uncertainties over the south-central US decrease as the grid-scale estimate is aggregated in time. Panel (b) displays how the uncertainties decrease as the annual-average estimate is averaged spatially. Uncertainties listed are standard deviations.

methane data, however, can indicate regional or national average emissions with greater confidence. Mathematically, the uncertainties per unit area decrease at aggregated spatial/temporal scales because the covariances in the posterior covariance matrix are often negative. Hence, at aggregated scales, the mean of the variances/covariances is usually smaller.

The covariance matrix $\mathbf{V}_{\hat{s}}$ encompasses most but not all uncertainties in the emissions estimate. It accounts for uncertainties in the drift coefficients ($\hat{\beta}$) and in the stochastic component of the emissions, and it accounts for uncertainty due to randomly-distributed model-data mismatch errors ($\varepsilon = \mathcal{N}(0, \mathbf{R})$, see Eq. A.1). However, existing statistical inversions cannot explicitly account for model-data mismatch errors that produce overall bias (i.e., if ε has a nonzero mean). Potential bias-type errors in WRF-STILT are discussed separately throughout the remainder of the supplement.

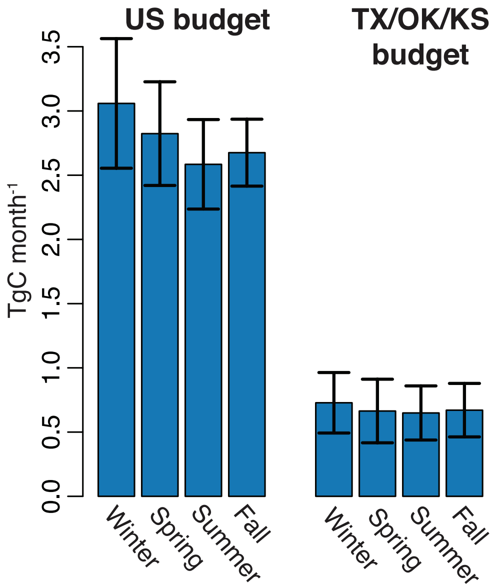


Figure A.5: Monthly anthropogenic methane budgets by season and associated 95% confidence intervals. The lack of strong seasonality implies minimal modeling errors due to wetland emissions, seasonal boundary condition biases, or seasonal wind field biases.

A.3.3 Wetland sources

We model the wetland contribution using the Kaplan wetland inventory Kaplan (2002); Pickett-Heaps et al. (2011), scaled in magnitude to match the observations as in Pickett-Heaps et al. (2011); Miller et al. (2014c). This signal (about 9 ppb in late summer, 2.0 Tg C TgC yr^{-1} for the continental US) is then removed from z to subtract the influence of wetland sources from the data. Wetlands make only small contributions to the signals at most of the observation sites in the US, and thus cannot be reliably constrained in the inversion. After this subtraction, the inversion produces optimized anthropogenic budgets with little seasonal variation (Fig. A.5). Since wetland emissions are strongly seasonal, this result indicates that our procedure does not produce wetland-related biases.

We also run a separate test inversion using the Dynamic Land Ecosystem Model (DLEM) for wetland subtraction instead of the Kaplan model Tian et al. (2010, 2011); Miller et al. (2014c) (also 2.0 TgC yr^{-1} for the continental US). This test inversion produces a US anthropogenic methane budget of 34.1 TgC yr^{-1} and a south-central US budget of 8.1 TgC yr^{-1}, very similar to the results using Kaplan wetland emissions. Hence we conclude that our results are independent of the source model used to account for wetland emissions.

Furthermore, wetland models predict only small to modest emissions over the largest source regions in our study. Two recent studies compare wetland methane fluxes for a number of biogeochemical models Melton et al. (2013a); Wania et al. (2013). None of the models put significant wetland emissions over Texas or Oklahoma, both relatively dry regions where our study found large methane emissions. Existing land maps indicate wetlands along the Mississippi River and Delta, but modeled wetland emissions in this region are nonetheless five times smaller than the anthropogenic sources estimated by our study in this area Melton et al. (2013a); Wania et al. (2013). The correlation between methane concentrations and propane in the south-central region additionally reinforce the attribution of high CH_4 fossil fuel extraction and processing.

A.3.4 Geological sources

Several studies report methane emissions from geological degassing, including ground seepage, geothermal emissions, and volcanic emissions Judd et al. (2002); Kvenvolden & Rogers (2005); Etiope et al. (2008). This study does not account for geological sources explicitly, but previous studies indicate that these fluxes would be small compared to the magnitude of US anthropogenic emissions. The estimated magnitude of this source ranges from 2.2 – 9% of total global emissions Judd et al. (2002); Kvenvolden & Rogers (2005); Etiope et al. (2008). One study estimates that terrestrial geological sources, in particular, contribute 1.1 – 2.8% of the global methane budget, and most emissions are attributable to volcanic activity and mud volcanoes Judd et al. (2002). A few mud volcanos exist along the California coastline, but these geological features are otherwise uncommon over the continental United States Dimitrov (2002). Based upon this information, we estimate at most a $\sim 5\%$ uncertainty in the emissions derived here due to geological degassing.

A.3.5 Uncertainties in sector-based emissions estimates

The section analyzes in greater depth the uncertainties in sector-based emissions estimates (e.g., for ruminants or the approximate fossil fuel budget). These uncertainties, listed in the main article, are calculated using the covariance matrices for $\hat{\beta}$ and $\hat{s}$ (Eq. A.9), summed to large spatial and annual temporal scales. The uncertainties on the sector-based budget estimates are large due to uncertainties in $\hat{\beta}$. The coefficients describe emissions with spatial patterns similar to the activity data, and co-located source sectors or errors in the activity datasets make the coefficient estimates ($\hat{\beta}$) less definitive. For example, ruminants and fossil fuel extraction have similar distributions over the south-central US, so some of the emissions assigned to ruminants in the deterministic model could instead be from the Texas and Oklahoma fossil fuel extraction sector. Similarly, if landfill emissions do not always coincide with population, then some landfill emissions may appear in the ruminant (Fig. A.2a), mean, or stochastic components (Fig. A.2c) instead of the population component (Fig. A.2b). Hence, the atmospheric methane data provides strong constraints on total emissions at the regional or national scale, but estimates by source sector often have larger confidence intervals.

We run a test inversion to investigate the possible effects of spatial correlation between ruminants and other source sectors. In this test inversion, we estimate different emissions factors ($\hat{\beta}$) for ruminants over four different regions of the United States: the western (CA, OR, WA, NV, ID, MT, AZ, UT, WY, NM, CO), the midwest (NE, SD, ND, MN, IA, MO, WI, IL, MI, IN), the south-central (TX, OK, KS, LA, AK), and the eastern United States (MS, TN, KY, AL, FL, GA, NC, SC, VA, WV, OH, PA, NY, MD, NJ, CT, VT, NY, NH, MA, RI, ME). Any differences in emissions factors by region represent one of three possibilities. First, emissions factors for agriculture may differ due to regional agricultural practices or climate. Second, the range may be caused by differences in the spatial distribution of the ruminant activity dataset from actual agricultural emissions. Third, the range may represent uncertainties caused by source sectors co-located with ruminants. Our agriculture emissions factors in the test case are lowest over the west and midwest (1.25 and 1.3 times EDGAR, respectively) and highest over the south-central US (2.6 times EDGAR). We repeat the calculation of south-central US non-agriculture and non-population emissions. In the main article, we estimate this budget at 3.7 ± 2.0 TgC yr^{-1}, a budget that could represent oil and gas emissions or unaccounted landfills. If we apply the western US ruminant emissions factor to Texas, Oklahoma, and Kansas, we obtain a higher estimate for this fossil fuel and/or landfill budget of 4.74 TgC/yr. Alternately, if we apply the south-central US ruminant emissions factor to Texas, Oklahoma, and Kansas, we obtain a lower estimate for the fossil fuel and/or landfill budget of 2.94 TgC/yr. These estimates are within the confidence intervals listed in the main article for ruminants and fossil fuel extraction and/or ruminants. Note that the different setup for $\mathbf{X}$ in this test case does not affect total posterior methane budgets in each region by more than 1%; changes in the configuration of $\mathbf{X}$ only affects emissions attribution by sector.

A.4 Model capability and validation

This section further validates the WRF-STILT model, the estimated methane budget, and discusses the geographic coverage of the methane observations.

A.4.1 Model transport and footprint validation

The WRF fields in this study are validated extensively by Nehrkorn et al. (2010), and several pertinent statistics are included here. The WRF simulations are set up specifically to conserve mass and do so by a factor of ten better than other meteorological products like the NCEP global analysis fields (FNL) Nehrkorn et al. (2010). Compared to US and Canadian radiosondes, horizontal winds in WRF exhibit a root mean squared error (RMSE) of 2.5 – 4 m/s with no change in error statistics at the top of the planetary boundary layer Nehrkorn et al. (2010).

Several previous inversion studies with STILT estimate emissions that are consistent across different meteorologies and compared with Eulerian models. This further validates model transport and indicates a lack of overall bias in the WRF-STILT footprints (**H**). Miller et al. (2012) use both WRF and the Regional Atmospheric Modeling System (RAMS) with STILT to estimate US nitrous oxide emissions; the US budgets match within 12 ± 6% Miller et al. (2012b). Furthermore, STILT studies of carbon monoxide and methane produce budgets comparable to top-down emissions estimates with the Geos-Chem model. Constraints on summertime US carbon monoxide emissions with RAMS-STILT and Geos-chem match to within 10% Hudman et al. (2008); Miller et al. (2008), and methane budgets for the Hudson Bay Lowland in Canada estimated with WRF-STILT and Geos-Chem are similar within 5% Pickett-Heaps et al. (2011); Miller et al. (2014c).

A.4.2 Validation of the methane boundary condition

Methane measurements from aircraft show good agreement with modeled concentrations, notably above 3000m where regional surface emissions have little influence (see Fig. 4 in the main article). At these altitudes, the mean measurement – posterior model difference is 2.8 ppb with a standard deviation of 18 ppb. These statistics reflect uncertainties in the modeled boundary condition but also reflect uncertainties in the modeled vertical gradient and in advection or convection of heterogeneous air masses in the upper free troposphere.

To test the effect of a 2.8 ppb uncertainty, we subtract this amount from the boundary condition and re-estimate the emissions. This test inversion produces US and south-central budgets of 35.4 ± 1.4 and 8.4 ± 1.0 TgC yr^{-1}, respectively. Given these uncertainties, the methane budgets presented in this study may be slightly low by 3 – 9 %.

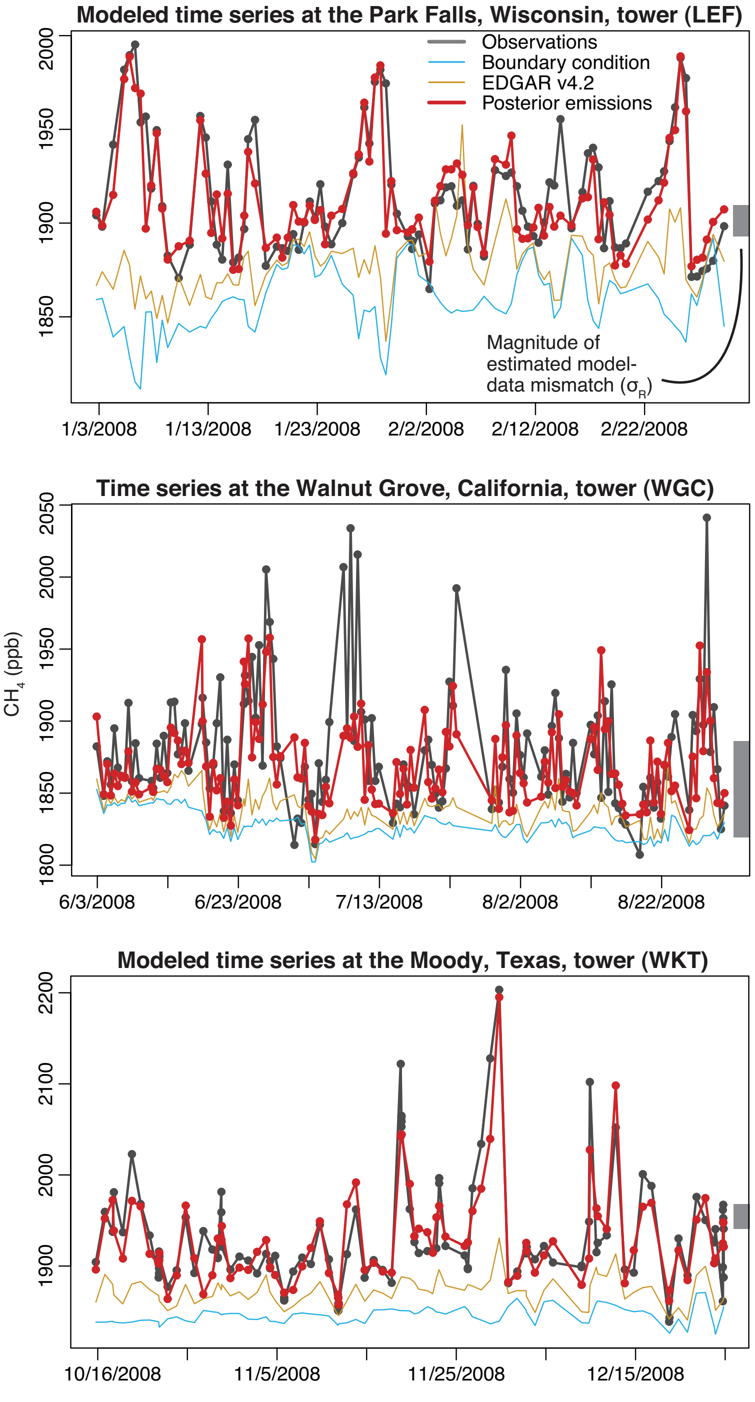
A.4.3 Comparison against aircraft and tower data

Regular methane observations from the NOAA aircraft monitoring network help validate the vertical model structure (i.e., planetary boundary layer height and convection). The comparison in Fig. 4 of the main article indicates several notable features of the model. First, the close match between model and observations in the free troposphere above 3000m confirms the suitability of the two-stage methane boundary condition, as discussed earlier. Second, the vertical structure in the model matches well against observations (to within 20 ppb at any aircraft sampling location).

Two additional figures provide further model-data comparison. First, Fig. A.6 compares modeled methane concentrations against time series of measurements at the NOAA tower locations. As discussed in the main article, the EDGAR inventory underestimates emissions in California (WGC tower) and Texas (WKT tower) more severely than in other geographic regions of the United States. Second, Figure A.7 compares all methane observations used in the inversion (from aircraft and tall tower locations) against modeled concentrations. Both figures (A.6 and A.7) highlight the improved data-model fit given by the posterior emissions estimate.

A.4.4 The utility of aircraft data in the inversion

To gauge the utility of aircraft data in the inversion, we run a test GIM using only observations from the ground sites. This inversion estimates a US methane budget of 37.4 ± 3.0 TgC yr^{-1} and Texas-Oklahoma-Kansas budget of 9.6 ± 1.3 TgC yr^{-1}. This test inversion produces emission fields that bleed into sparsely populated areas adjacent to large source regions (e.g., surrounding Texas and California). But modeled concentrations using this test emissions estimate are too high within the free troposphere compared to aircraft data. As a result, we infer that an estimated US


Modeled time series at the Park Falls, Wisconsin, tower (LEF)

Time series at the Walnut Grove, California, tower (WGC)

Modeled time series at the Moody, Texas, tower (WKT)

Figure A.6: A model-measurement comparison at several tower sites. The EDGAR v4.2 and posterior model plots include the boundary condition, wetland contribution modeled from the Kaplan inventory, and the anthropogenic contribution modeled from EDGAR and the posterior emissions, respectively.

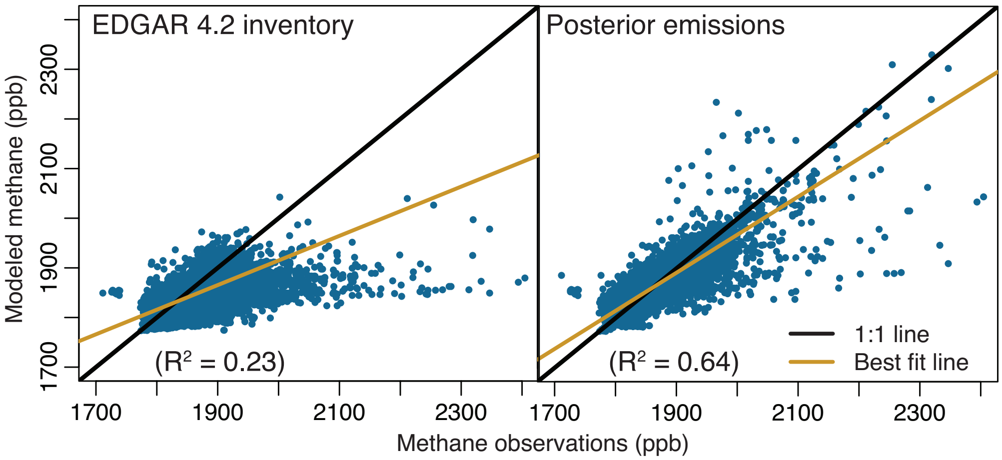


Figure A.7: A model-data comparison scatter plot for the posterior emissions estimate and EDGAR 4.2

budget from an inversion without aircraft data would likely be too high by 7–17%. The aircraft data bound the vertical re-distribution of surface emissions. Without this bound, the inversion may inaccurately estimate emissions that agree with surface methane measurement but nevertheless result in too much methane in the free troposphere and emission fields that spread too far across the landscape.

A.4.5 Geographic coverage

Figure A.8 visualizes the average footprint (**H**) of the methane measurement network in 2007 and 2008. The figure confirms that the observation network is sensitive to emissions over much of the central and western United States but insensitive to coal or urban-related emissions in the mid and southern Atlantic states. In particular, the sparsity of observations near West Virginia and Pennsylvania inhibits a strong constraint on East Coast coal emissions. US EPA estimates that coal constitutes 11% of all methane emissions, and approximately one third of all US coal production is in Appalachia US EPA (2013); Young (2012). Consequently, we estimate that a paucity of observations over Appalachia may contribute a 1 – 3% uncertainty in the total US

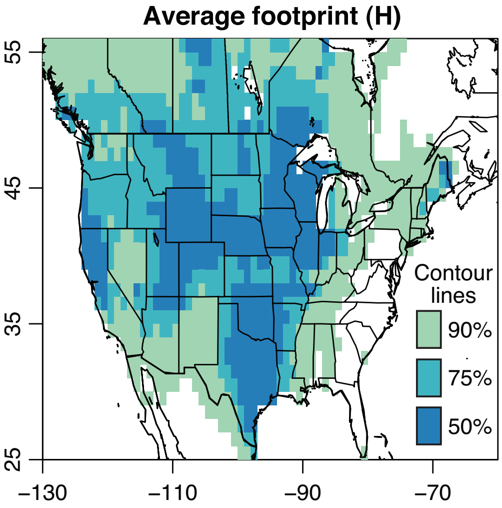


Figure A.8: The footprint of the 2007–2008 methane observation network (i.e., **H**, averaged by row). Contour lines represent the percentage of the summed observation network footprint that is encapsulated by the given area. In other words, this figure illustrates the extent to which emissions in different locations are "seen" by the observation network and provides a sense of the network's ability to constrain sources across different regions.

methane budget.

A.4.6 Comparison with the methane estimate from Katzenstein et al. (2003)

An steady increase in the number of US gas wells between 2001 and 2007–2008 may partially explain the differences between this study and Katzenstein et al. (2003). The total number of wells in Texas, for example, increased by ∼58% over this time period US Energy Information Administration (2013).

Several aspects of our inverse modeling study also allow a more extensive picture of emissions than available to previous studies. The NOAA/DOE observations from a diverse set of measurement platforms characterize the atmospheric distribution of methane over a multi-year period. We note that concentrations measured weekly at the NOAA Texas (WKT) tower from 2001 to 2003 average ∼80ppb higher than ground-level observations near the WKT site during the Katzenstein et al. study Katzenstein et al. (2003). A stationary front on the Texas - Oklahoma border and strong convection over Texas during the 5-day measurement period may have lofted methane plumes higher into the troposphere, beyond detection at the surface. The WRF-STILT model simulates the temporal and spatial variation in advection, convection, and boundary layer dynamics, consistent with meteorological data. This detailed characterization of the atmosphere accounts for methane plumes that are not uniformly mixed within the lower troposphere. Hence, the comprehensive NOAA/DOE methane measurements and our GIM provide perspective on national emissions and source sectors not possible in previous efforts.

B

Supplemental material for chapter 4

B.1 Detailed model explanation and validation

B.1.1 Meteorology overview

The WRF meteorological simulations used in this study have been constructed explicitly for trace gas modeling with STILT, and the same WRF fields have been used in a number of greenhouse gas modeling studies (e.g., Kort et al., 2010; Gourdji et al., 2010; Huntzinger et al., 2011; Gourdji et al., 2012; Miller et al., 2013, 2014c; Shiga et al., 2013). Among other features, these simulations conserve mass by a factor of ten better than other meteorological products like the NCEP global analysis fields (FNL) (Nehrkorn et al., 2010). Nehrkorn et al. (2010) and Hegarty et al. (2013) validate the WRF fields generated for STILT against meteorological observations and controlled tracer release experiments, respectively. The former study compares horizontal winds in WRF against radiosondes: WRF exhibits a root mean squared error (RMSE) of 2.5 – 4 m/s with no change in error statistics at the top of the planetary boundary layer. The latter study compares the performance of several particle trajectory models, including STILT, coupled with a variety of

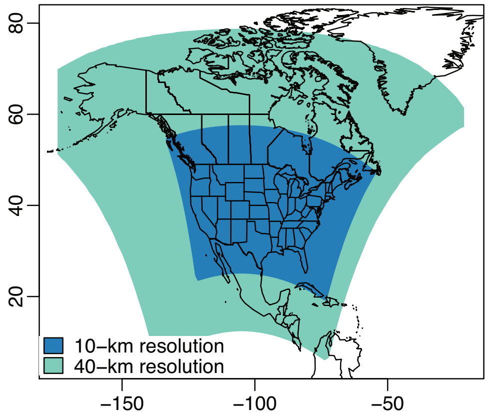


Figure B.1: The 10-km and 40-km resolution domains used in the WRF model runs.

wind fields. WRF fields produce the best results among the examined wind fields when tested against controlled tracer release experiments (Hegarty et al., 2013).

Section B.3 describes the land surface model used in WRF.

B.1.2 Nested meteorology setup

The WRF-fields used in this study are nested: 10-km resolution wind fields drive particle trajectories over most of the continental US and much of Canada while 40-km wind fields drive trajectories in more distant regions (Fig. B.1). This nested setup affords a detailed description of winds within 24–48 hours of the measurement location, and the 40-km resolution at greater distances allows substantial computational savings. This arrangement makes it computationally feasible to run an ensemble of 500 back-trajectories for each of ∼15,000 hourly methane measurements. Nested WRF fields have been employed in a number of STILT modeling studies (e.g. Zhao et al., 2009; Nehrkorn et al., 2010; Kort et al., 2010; Gourdji et al., 2010; Huntzinger et al., 2011; Pillai et al., 2011; Gourdji et al., 2012; Jeong et al., 2012; McKain et al., 2012; Pillai et al., 2012; Xiang et al., 2013b; Miller et al., 2013, 2014c; Hegarty et al., 2013; Newman et al., 2013; Shiga et al., 2013).

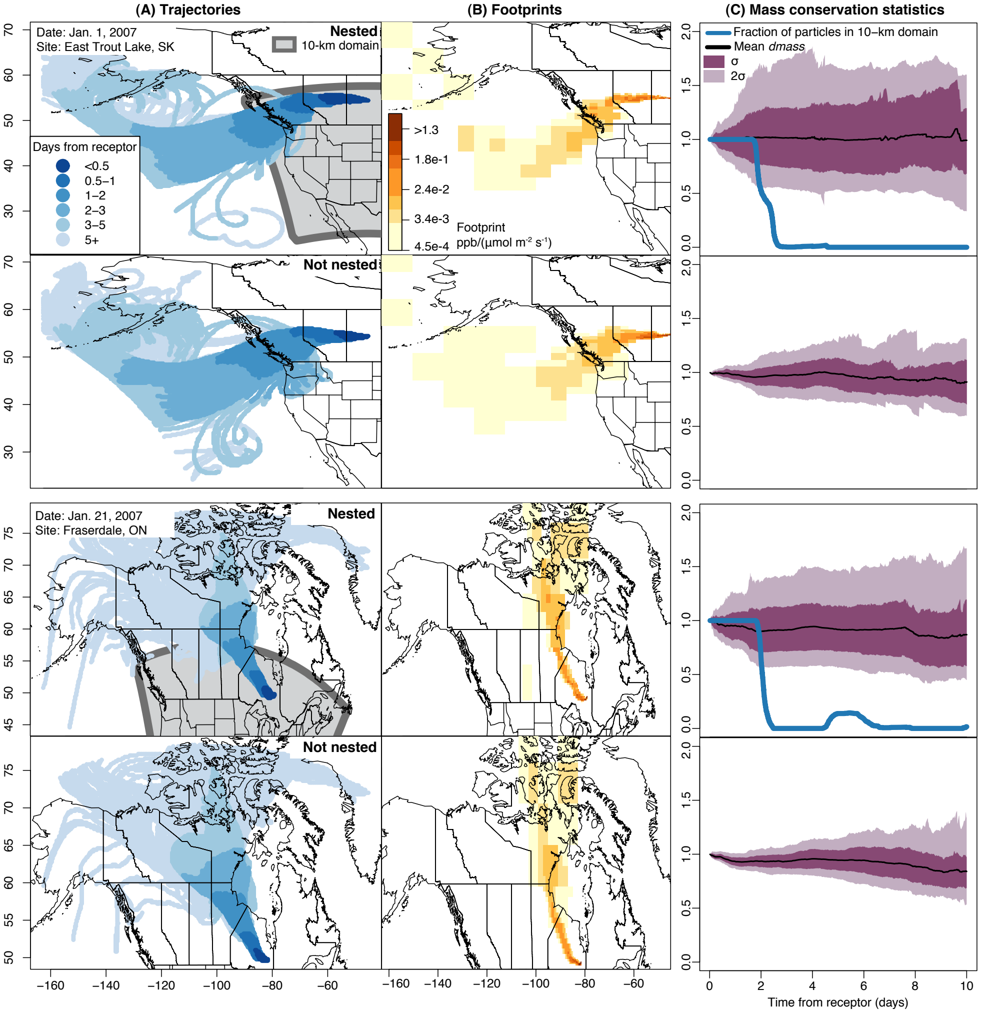


Figure B.2: A comparison of STILT trajectories run with nested-resolution WRF fields (odd rows of the figure) and non-nested 40-km fields (even rows). Column (A) shows the trajectories associated with single methane measurements at East Trout Lake, Saskatchewan, on Jan. 1, 2007, and at Fraserdale, Ontario, on Jan. 21, 2007. Each ensemble of trajectories contains 500 particles that run backward in time along the WRF wind fields. Column (B) displays the footprints associated with each set of trajectories (used to construct **H**). Only particles within the mixed layer have a non-zero footprint, so the footprints have a less dispersed spatial distribution compared to the trajectories. Column (C) plots *dmass*, a cumulative metric of mass conservation (refer to section B.1.2). Panels show both the mean and distribution of *dmass* at each time step. *dmass* values reach as high as 50 and as low as zero when STILT is paired with other meteorological products (section B.1.2).

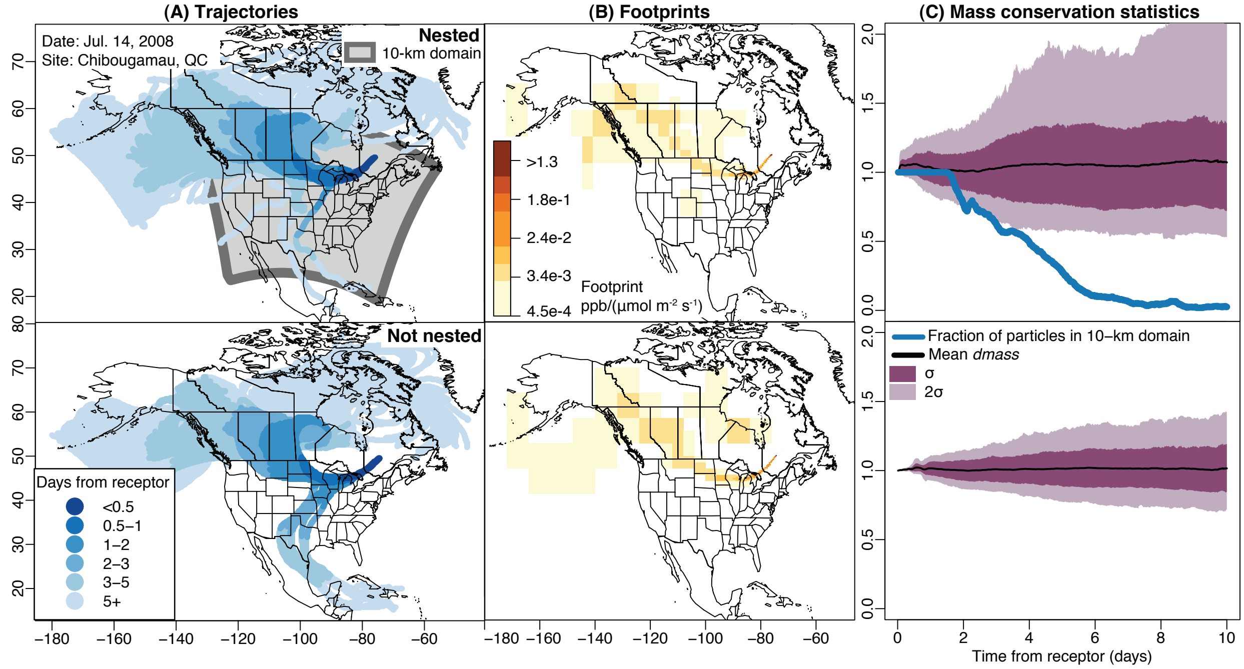


Figure B.3: This figure is identical to Fig. B.2 but compares trajectories from Chibougamau, Quebec, on July 14, 2008.

Figures B.2 and B.3 compare sample STILT trajectories created with the nested WRF fields against trajectories created with the 40-km fields only. The examples in these figures span winter to summer seasons and represent different synoptic transport patterns. Individual panels display the back-trajectories (Figs. B.2a and B.3a), the influence footprints (used to construct **H**, Figs. B.2b and B.3b), and a metric of mass conservation (*dmass*, Figs. B.2c and B.3c). For the nested WRF setup, transport is continuous and smooth at the 10-km/40-km wind field boundaries. Furthermore, synoptic-scale transport patterns are consistent in nested and non-nested model runs.

Both nested and non-nested WRF-STILT setups display favorable mass conservation statistics. The *dmass* parameter (Figs. B.2c and B.3c) provides an estimate of cumulative mass creation or loss in each back-trajectory. A value of one indicates perfect mass conservation while 0.5 indicates 50% loss, and 1.5 indicates 50% mass gain (Nehrkorn et al., 2010). STILT uses the *dmass* parameter to correct for mass violation in the footprints. In general, the 40-km wind fields produce less mass creation/loss compared to the 10-km wind fields. Hegarty et al. (2013) report that the 10-km resolution fields nonetheless reproduce atmospheric transport better than coarser-resolution winds. In the nested runs, mass conservation statistics are also continuous at the wind field boundaries. In comparison, other meteorological products like EDAS-40 (Eta Data Assimilation System) and FNL (Final) produce *dmass* values as high as 50 and as low as zero (Miller, 2007; Nehrkorn et al., 2010).

Note that we use nested wind fields exclusively for computing the back-trajectories. The Kaplan wetland flux model uses many inputs from WRF (e.g., soil moisture, soil temperature), and these inputs are drawn exclusively from the 40-km fields.

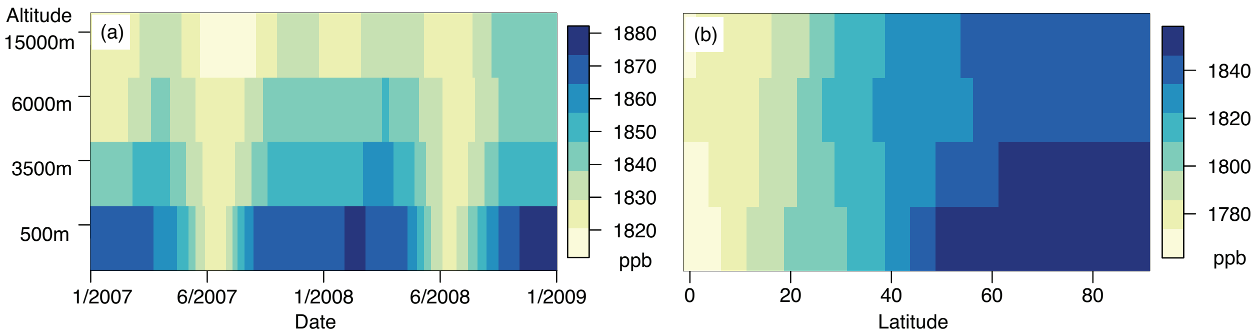


Figure B.4: Examples of the methane boundary condition (a) at 50°N and a variety of dates and altitudes and (b) on August 1, 2008, and a variety of latitudes and altitudes. Note the different scales on each panel.

B.1.3 Boundary condition

WRF-STILT only models methane sources over the North American continent and therefore requires a methane boundary condition (see sections 2.2 in the main article). The boundary condition represents the mixing ratio of methane in air masses before they reach the continent, from both westerly and northerly synoptic air flow.

This study uses an empirical boundary that interpolates a variety of trace gas measurements from ground-based sites and aircraft in the NOAA ESRL Global Monitoring Division's Cooperative Global Air Sampling Network. The interpolated boundary curtain includes methane measurements from Greenland, Alaska, California, Hawaii, and from ship tracks in the Pacific Ocean, among other sites. The measurements are interpolated latitudinally and vertically using the a curve-fitting procedure (Thoning et al., 1989). Fig. B.4 displays example cross-sections of the boundary condition by date (Fig. B.4a) and latitude (Fig. B.4b).

The estimated boundary value associated with each trajectory run depends on the ending latitudes, altitudes, and days of the trajectories.

For most model-data comparisons in the main article, we have subtracted the estimated boundary value from the measurements to clearly depict the effect of North American methane sources on the measurement sites (e.g., Figs. 2 and 4 in the main article). The full mixing ratio measurements show large seasonality (~40ppb) unrelated to emissions, due to changes northern hemisphere hydroxyl radical concentrations. A time series plot of full mixing ratio measurements would visually obscure the effect of North American methane fluxes on the measurements.

Existing studies validate this empirical boundary condition against methane data from aircraft in the North American free troposphere. Miller et al. (2013) compare the boundary condition against NOAA aircraft data over the continental US (which, in some cases, were also used to construct the boundary condition). They find that the boundary has small regional and/or seasonal biases of 0.75 to 7ppb compared to this aircraft data. Miller et al. (2013) further adjust the empirical boundary condition to remove these regional and/or seasonal biases compared to aircraft data. However, regular methane data from aircraft are sparse over Canada, so it is not feasible in this study to fit the methane boundary condition to regional free troposphere values.

B.1.4 Consistency between WRF-STILT and other atmospheric models

A number of existing STILT studies produce surface flux estimates that are comparable to studies with different transport models and/or meteorologies. These comparisons support model transport as represented by WRF-STILT and suggest a lack of overall bias in the WRF-STILT footprints (**H**). Miller et al. (2012b) use both WRF and the Regional Atmospheric Modeling System (RAMS) with STILT to estimate US nitrous oxide emissions; the US budgets match within 12 ± 6%. Furthermore, STILT studies of carbon monoxide and the present methane study both produce budgets comparable to top-down emissions estimates with the Geos-Chem model. Constraints on summertime US carbon monoxide emissions with RAMS-STILT and Geos-Chem match to within 10% (Hudman et al., 2008; Miller et al., 2008), and methane budgets for the Hudson Bay Lowlands (HBL) in this study match Pickett-Heaps et al. (2011) to within ~5%.

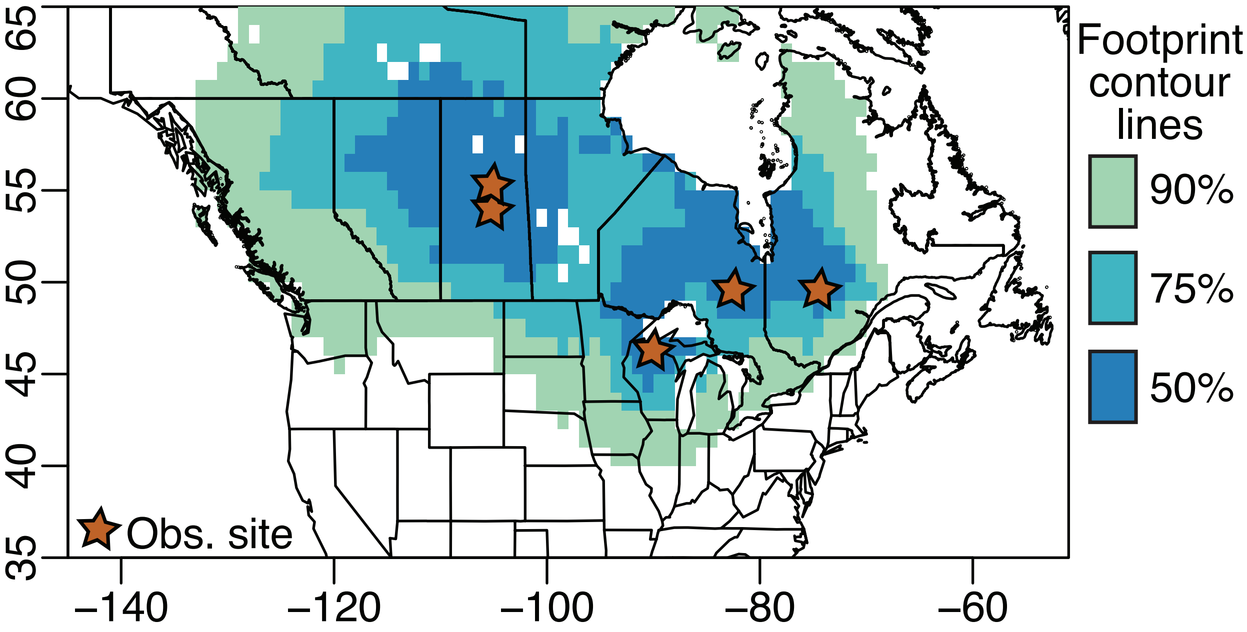


Figure B.5: This figure visualizes the sensitivity of the observations to emissions in different regions. The contour lines indicate the regions that encompass 50, 75, and 90% of the total sensitivity or footprint of the observations. In other words, the figure provides a measure of geographic coverage of the inversion.

B.1.5 Geographic coverage and data limitations

Fig. B.5 displays the summed footprints (**H**, units of mixing ratio per flux) for the two year (2007–2008) study period. The footprints indicate the sensitivity of the measurements to fluxes in different geographic regions. This figure displays the contour lines of the footprints: the regions that encompass 50, 75, and 90% of the total sensitivity. The footprints show broad coverage across Canada. Observations are particularly sensitive to fluxes in the Hudson Bay and Great Lakes regions as well as fluxes in Saskatchewan and eastern Alberta.

Three additional Canadian measurement sites (Lac Labiche, Alberta; Egbert, Ontario; and Sable Island, Nova Scotia) are excluded from the analysis. The Lac Labiche and Egbert sites lie relatively close to the ground (10m and 6m agl, respectively) and are therefore particularly sensitive to near-surface mixing or unresolved turbulent eddies. Sable Island lies far from methane sources and sees mostly marine background air.

Fig. B.5 indicates several gaps in the 2007–2008 observation network. The network has limited coverage across Arctic Canada, in eastern Quebec and maritime Canada, and over the southern half of Alberta.

B.2 Detailed statistical methodology

B.2.1 Covariance matrix structure and estimation

We estimate the parameters (θ) that define the covariance matrices (**R** and **Q**) using a method known as restricted maximum likelihood (REML). Maximum likelihood methods are standard estimation tools for statistical parameters, including variances and covariances (e.g., Devore, 2012, ch. 6.2). Restricted Maximum Likelihood makes it possible to estimate the variance and/or covariance when the mean, and any other drift coefficients (β), are unknown, by first removing the effect of these unknown variables (Corbeil & Searle, 1976). REML has been used for covariance parameter estimation in the solution of a variety of inverse problems (e.g., Kitanidis, 1995), including the estimation of atmospheric trace gas sources and sinks (e.g., Michalak et al., 2004; Gourdji et al., 2012).

REML will estimate the parameters (θ) that are most likely given both the measurements (z)

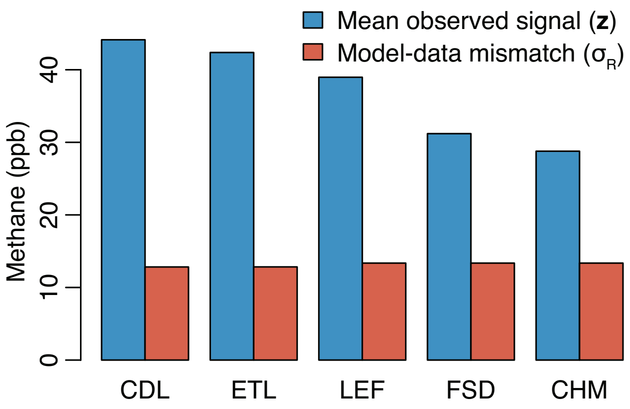


Figure B.6: Blue bars illustrate the mean methane signal from North American sources at each tower site (z, the measured mixing ratio minus the boundary condition). The red bars depict the estimated model-data mismatch errors (e.g., the square root of the diagonal of $\mathbf{R}$). These errors encompass uncertainties unrelated to the emissions (e.g., transport, methane boundary condition, etc.).

and atmospheric model ($\mathbf{H}$) (e.g., Michalak et al., 2004):

$$\begin{aligned} L_\theta &= -\ln \int_\beta \int_s p(s, \beta, \theta | z, \mathbf{H})\, ds\, d\beta \\ &= \tfrac{1}{2} \ln |\mathbf{\Psi}| + \tfrac{1}{2} \ln |\mathbf{X}^T \mathbf{H}^T \mathbf{\Psi}^{-1} \mathbf{H} \mathbf{X}| + \tfrac{1}{2} z^T \mathbf{\Xi} z \end{aligned} \tag{B.1}$$

where

$$\mathbf{\Psi} = \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R} \tag{B.2}$$

$$\mathbf{\Xi} = \mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}^T\mathbf{\Psi}^{-1}\mathbf{H}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^T\mathbf{\Psi}^{-1} \tag{B.3}$$

The best estimate of the covariance matrix parameters (θ) is obtained by minimizing the negative log likelihood of the observations (L_θ in Eq. B.1). This is equivalent to maximizing the probability $p(s, \beta, \theta | z, \mathbf{H})$ after integrating out the unknown fluxes (s) and coefficients (β).

For the inversion setup here, the model-mismatch covariance matrix ($\mathbf{R}$) is diagonal. This setup is consistent with a number of previous inversion studies that use tower-based trace gas data (e.g., Zhao et al., 2009; Bergamaschi et al., 2010; Jeong et al., 2013; Gourdji et al., 2012). We use REML to estimate a different model-data mismatch variance (σ_R^2) for the eastern (LEF, CHM, FSD) and western (CDL, ETL) tower sites. Furthermore, we model this variance to be constant across all seasons. Fig. B.6 compares the estimated model-data mismatch values at each measurement site against the average measured signal from North American sources (z).

The a priori covariance matrix ($\mathbf{Q}$) describes the differences between the deterministic model and the true fluxes. It requires a more complex structure for several reasons. First, the deterministic model may be a better (or worse) fit over wetland flux regions than in regions dominated by anthropogenic emissions. Second, the deterministic model fits wetland fluxes in eastern Canada better than in western Canada (see Fig. 4 in the main article). Finally, the deterministic model may describe wetland fluxes more adeptly in some seasons than in others (e.g., Fig. 4 in the main article). For these three reasons, the inversion problem requires a covariance function that is both spatially and temporally non-stationary. To this end, we model the diagonal elements of $\mathbf{Q}$, denoted σ_Q^2, as follows:

$$\begin{aligned}\sigma_Q \;=\; \alpha_1 \;+\; \alpha_2[\text{monthly Kaplan HBL budget}] +& \\ \alpha_3[\text{smooth functions over anthropogenic regions}]& \qquad \text{(B.4)}\end{aligned}$$

where α_1, α_2, and α_3 are constants estimated by REML. In other words, the covariance function has a component that is spatially and temporally constant (α_1), that varies seasonally with the magnitude of wetland fluxes (α_2[monthly Kaplan HBL budget]), and that varies spatially (α_3[smooth functions...]). Figure B.7 displays the annual mean σ_Q as estimated by REML.

The covariances decay exponentially with distance:

$$Q_{ij}(h_{ij}|\sigma_{Qi}, \sigma_{Qj}, l) = \sigma_{Qi}\sigma_{Qj} \exp\left(-\frac{h_{ij}}{l}\right) \qquad \text{(B.5)}$$

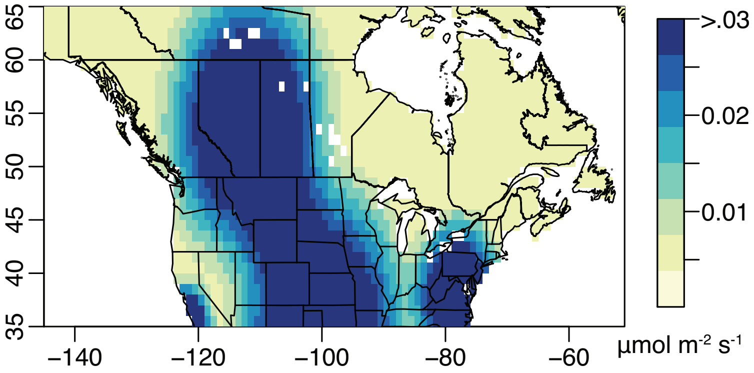


Figure B.7: The estimated standard deviation of the difference between the deterministic model and the true unknown fluxes ($\hat{\sigma}_Q$). This is synonymous with the standard deviation of the stochastic component, all fluxes not described by the deterministic model. We estimate these quantities using REML (section B.2.1).

where h_{ij} is the great circle distance between locations i and j, l is the decorrelation lengthscale, and σ_{Qi} and σ_{Qj} are the standard deviations given in Eq. B.4 at locations i and j.

REML estimates a total decorrelation length ($3l$) of 381 ± 36km. REML did not converge on temporal decorrelation length, and we omit temporal correlations in **Q**.

In summary, we implement REML to estimate six parameters (θ) that help define the covariance matrices (**R** and **Q**): α_1, α_2, α_3, l, and two values of σ_R.

B.2.2 The Bayesian Information Criterion

This study uses the Bayesian Information Criterion (BIC) to select predictors for the deterministic model. We calculate BIC scores using the following equations (as presented in Gourdji et al., 2012):

$$BIC \;=\; \ln|\mathbf{\Psi}| + z^T\mathbf{\Xi}z + p\ln n \tag{B.6}$$

The final BIC score for each candidate model is unit-less. This score increases with the magnitude of the combined covariance matrices ($\mathbf{\Psi}$), with the weighted sum of squared residuals

($z^T \Xi z$), and with the complexity of the deterministic model ($p \ln n$).

We add a further constraint that none of the environmental datasets in $\mathbf{X}$ can contribute negatively to the methane fluxes. In other words, it would be illogical for either an anthropogenic or wetland emissions model to have a net negative flux. Therefore, we check the estimated coefficients $\hat{\beta}$ on the selected model to ensure they are positive:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{H}^T \mathbf{\Psi}^{-1} \mathbf{H} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{H}^T \mathbf{\Psi}^{-1} z \tag{B.7}$$

This equation calculates the coefficients that are most likely given the atmospheric data z and $\mathbf{H}$, analogous to a weighted least squares regression.

B.2.3 Estimation of the posterior fluxes

A standard atmospheric inversion setup will minimize the cost function but is oblivious to any known physical bounds on the emissions. However, large negative methane fluxes would be unrealistic; methane has a small soil sink (∼4% of the global sink), and only ∼0.5% of the global atmospheric methane loss occurs in boreal soils (Dutaur & Verchot, 2007). Hence, we assume that all fluxes estimated by the inversion should be nonnegative. Of further concern, unrealistic negative fluxes can exaggerate source strength in adjacent, high emissions regions (Miller et al., 2014c). We implement nonnegativity in the inversion with Lagrange multipliers via an iterative trust region algorithm. The companion paper Miller et al. (2014c) describes this approach in detail.

B.2.4 Uncertainties in the flux estimate

The posterior error covariance matrices provides a measure of uncertainty in both the estimated fluxes ($\hat{s}$) and coefficients ($\hat{\beta}$) (e.g., Michalak et al., 2004):

$$\begin{bmatrix} \mathbf{V}_{\hat{s}} & \mathbf{V}_{\hat{s}\hat{\beta}} \\ \mathbf{V}_{\hat{\beta}\hat{s}} & \mathbf{V}_{\hat{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} & \mathbf{Q}^{-1}\mathbf{X} \\ \mathbf{X}^T\mathbf{Q}^{-1} & \mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X} \end{bmatrix}^{-1} \tag{B.8}$$

Michalak et al. (2004) and Miller et al. (2014c) discusses posterior covariances in greater detail, and the former paper includes an alternative formulation of Eq. B.8 that does not require calculating the inverse of **Q**.

In this paper, we use the posterior covariance matrices to calculate all uncertainty estimates. The variances and covariances of the estimated fluxes ($\hat{s}$) account for uncertainties in the drift coefficients and in the stochastic component of the emissions, and they account for uncertainty due to randomly-distributed model-data mismatch errors. Note that the uncertainties listed for the coefficients and posterior fluxes in this study are 2σ, unless otherwise noted.

Of note, existing statistical inversions cannot explicitly account for model-data mismatch errors that produce overall bias (e.g., biased mixed layer heights or biased boundary condition estimates). The consistency between STILT-based flux estimates and other model studies (Section B.1.4) implies an absence of large or flagrant systematic errors in WRF-STILT.

Furthermore, the posterior covariances do not account for any uncertainty in the choice of covariates for the deterministic model. If there were many plausible candidate models for **X**, all with similar BIC scores, this could be an important component of the uncertainty in the flux estimate. In this study, most candidate models have BIC scores that are more than 10 points above the optimal model. This indicates a ‘very strong’ level of evidence against those models (see section 3.2.2 in the main article and Kass & Raftery (1995)). A notable exception is the decision to

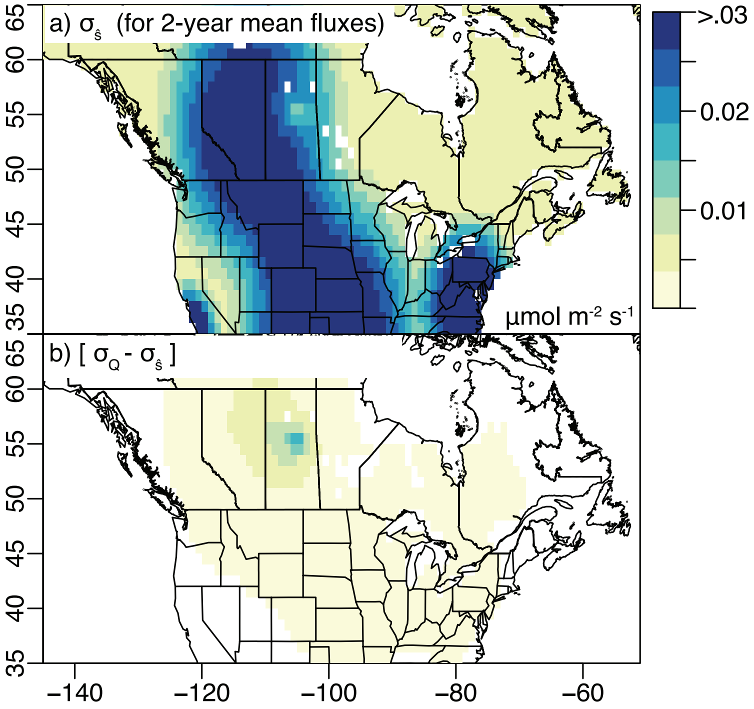


Figure B.8: (a) Uncertainty in the posterior estimate ($\sigma_{\hat{s}}$). (b) The reduction in uncertainty relative to the prior estimate ($\sigma_Q - \sigma_{\hat{s}}$). Note that these uncertainties decrease over increasing spatial scales. Mathematically, this decrease is due to negative covariances in the posterior covariance matrix.

exclude soil carbon from the deterministic model. The evidence against including soil carbon is 'positive' but not 'strong' or 'very strong.' In other words, the possibility of including soil carbon within **X** and its effect on the final flux estimate ($\hat{s}$) is a source of uncertainty not explicitly reflected in the posterior covariance matrices. The choice of covariate over Alberta represents an additional possible uncertainty. We use a set of smooth functions to describe anthropogenic emissions over Alberta (Section B.4). We could not find a spatial pattern that fit the atmospheric methane data more adeptly. However, a better spatial predictor of Alberta emissions could exist.

Figure B.8a displays a map of the posterior uncertainty, denoted $\sigma_{\hat{s}}$, and Fig. B.8b displays the corresponding decline from the prior uncertainties. It is important to note that the posterior uncertainties often decrease exponentially at increasing spatial scales (e.g. Miller et al., 2013). This decrease occurs because of negative posterior covariances among emissions points distributed geographically. In other words, the atmospheric data data may provide a weak constraint on the exact distribution of emissions between any two adjacent grid boxes. However, the data may nonetheless constrain regional and national methane budgets with high relative confidence. For example, the uncertainties in Fig. B.8a relative to the flux estimate can be 100%, but the relative uncertainty in the total Canadian methane budget (Fig. 7 of the main article) is far smaller.

The posterior uncertainties indicate key information about methane emissions in Alberta. These uncertainties at grid-scale are comparable to the magnitude of emissions. This result indicates that the precise spatial distribution of the posterior emissions over Alberta is highly uncertain. The inversion places large emissions in Northern Alberta and near the Rocky Mountain Front Range, but this assignment is poorly constrained by the available methane data. As explained above, the atmospheric methane data better constrain regional-scale totals. For example, the inversion estimates a budget for western Canada of 11.3 ± 2.4 (west of -100° lon.). The uncertainties in this regional-scale total are small relative to uncertainties at grid-scale.

The spatial distribution of emissions over Alberta remains an important, open question. Two efforts, in particular, could reduce these uncertainties in future estimates. First, expansion of the Environment Canada measurement network will provide stronger constraints on both the magnitude and spatial distribution of emissions. The 2014 network includes methane measurements at Abbotsford, British Columbia; Esther, Alberta; Fort McKay, Alberta, and Bratt's Lake, Saskatchewan (in addition to the sites used in this study). Second, top-down emissions estimates would benefit from accurate, detailed maps of spatial processes related to methane emissions. These maps could be used to construct a more capable deterministic model. The better the deterministic model describes the atmospheric methane measurements, the smaller the variance of the stochastic component, and the smaller the posterior variances and uncertainties.

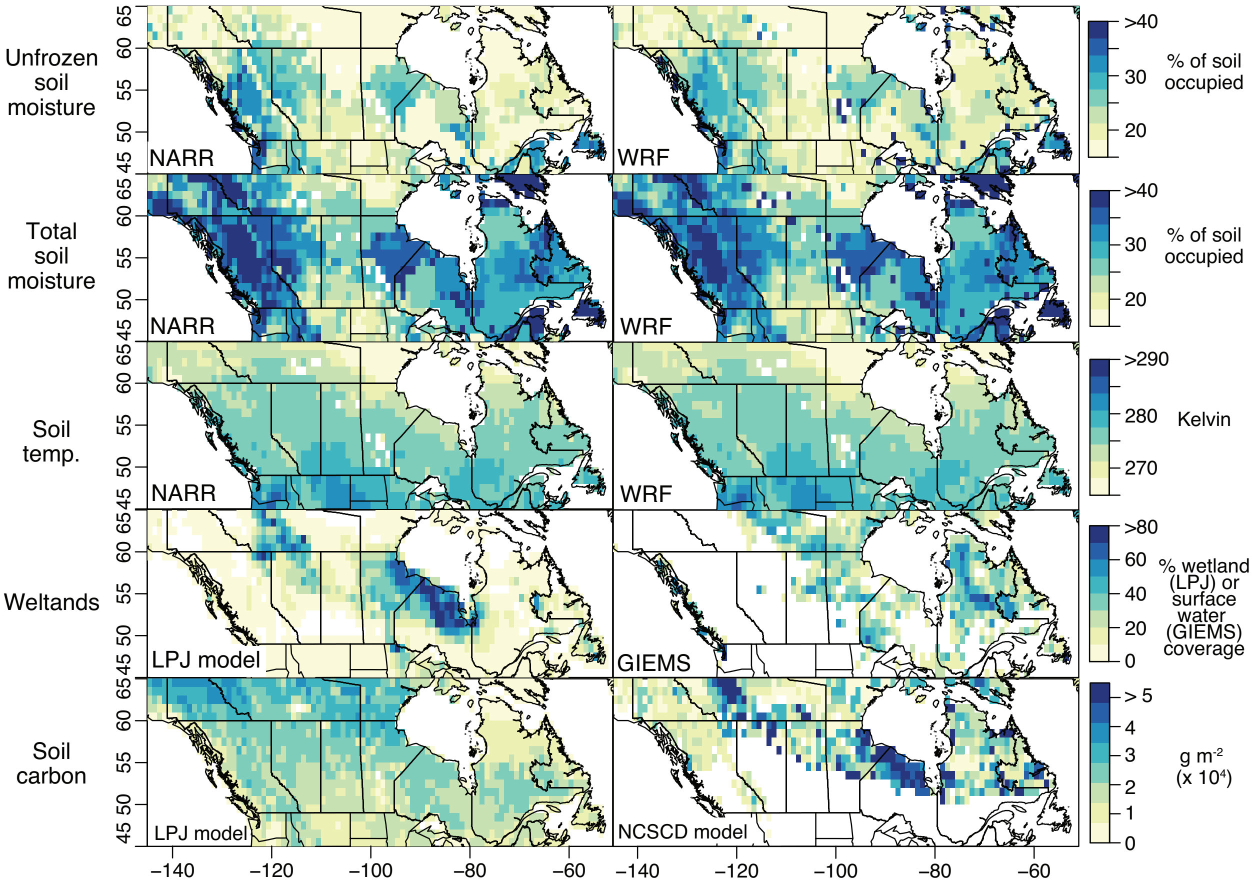


Figure B.9: Environmental datasets tested for the deterministic model (Table 1 in the main article). NARR and WRF plots are annual averages.

B.3 Detailed comparison of biogeochemical model drivers

This section provides visual comparison of the wetland-related environmental datasets that are used in Kaplan, DLEM, and the deterministic model.

Figure B.9 compares all of the wetland-related variables that are listed in Table 1 of the main article. The top panels display soil moisture and temperature estimates from WRF and NARR. These meteorologies are similar for two reasons; NARR serves as the initial condition for the WRF runs, and both NARR and WRF use the Noah land surface model. The Noah model is de-

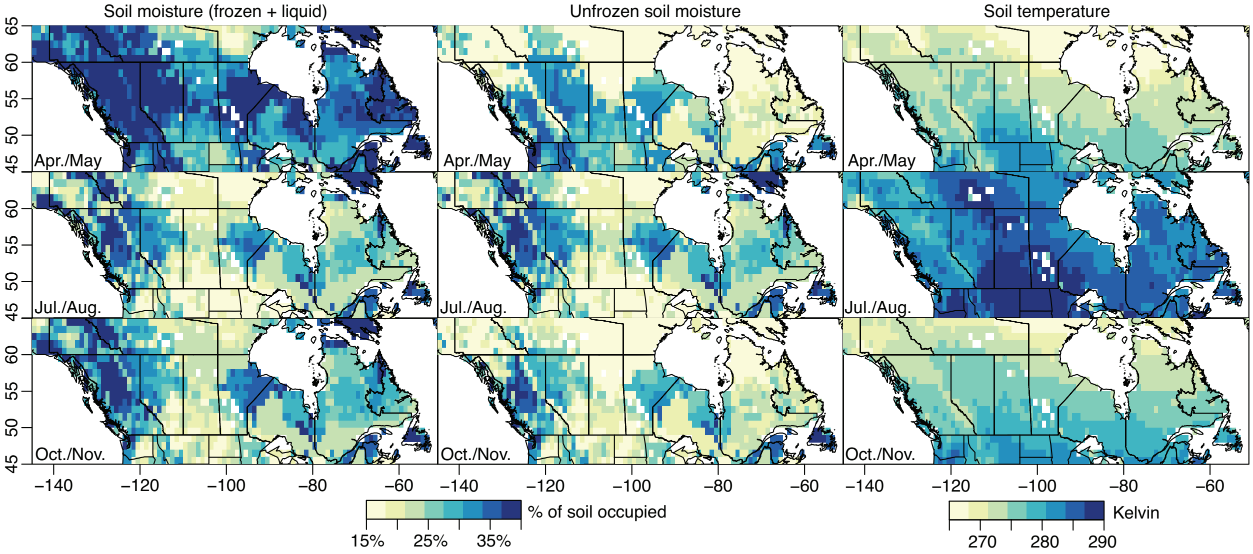


Figure B.10: Soil variables from NARR for selected months of the year (averaged over 2007 – 2008).

scribed in a multitude of publications, and a list of relevant references is available in Ek et al. (2003) and Niu et al. (2011). Fig. B.10 further illustrates the soil variables by season, in this case for NARR. Total soil moisture peaks in spring and declines into the summer. Temporal patterns in unfrozen soil moisture, by contrast, are dominated by the seasonal freeze-thaw cycle.

It is important to note that the NARR and WRF outputs used here are constrained by meteorological observations. For example, the NARR model-data assimilation framework adjusts land surface model output at each time step to match available meteorological observations, including temperature and precipitation. Mesinger et al. (2006) provide a succinct overview of NARR and visually compare precipitation measurements against NARR estimates. The WRF runs, in contrast, are re-initiated at regular intervals with initial conditions from NARR. The WRF runs do not assimilate meteorological observations directly, but this setup ensures that the WRF runs cannot stray far from the available meteorological data.

Subsequent panels of Fig. B.9 display the wetland and soil carbon estimates. Wetland estimates include output from the LPJ model and GIEMS. GIEMS is a monthly-resolution satellite product that estimates global surface water (Prigent et al., 2007; Papa et al., 2010). In this study, we use long term, summer-averaged surface water from GIEMS as a wetland proxy (1993–2007, averaged over July–September). As discussed in Melton et al. (2013b), surface water may not necessarily be an ideal proxy for wetlands. Among other issues, this proxy would underestimate wetlands that hold the majority of water in the subsurface. For this and other possible reasons, GIEMS and the LPJ model have different spatial distributions. For example, GIEMS estimates high surface water over northern Quebec while the LPJ wetland estimate is largely restricted to the HBL and Great Slave Lake regions.

The bottom panels of Fig. B.9 visualize the soil carbon estimates, and these estimates exhibit very different spatial distributions. The LPJ soil carbon estimate correlates closely with latitude. Note that this estimate is multiplied by a factor of 4.15 before it is used in the Kaplan model in order to match the LPJ inputs used in Pickett-Heaps et al. (2011) (refer to section 2.4.1 of the main article). The NCSCD soil carbon estimate (Tarnocai et al., 2009; Hugelius et al., 2013), in contrast, has a spatial distribution more similar to the LPJ wetland estimate. NCSCD notably only estimates soil carbon in permafrost regions.

The deterministic model, described in the main article (section 4.2), excludes any estimate of soil carbon. This feature may be reconcilable with existing methane models. One might expect wetland location to correlate with soil carbon, so both environmental datasets may not be strictly necessary in a wetland flux model. In the Northern Circumpolar Carbon Database (Tarnocai et al., 2009), for example, wet peatland soils (histels and histosols) have the highest carbon density of any soil type by a factor of 2 – 20. Other environmental factors may complicate this correlation between the wetland distribution and soil carbon; both the type of plant carbon and age since deposition can affect the rate of methane production (Chanton et al., 2008; Bridgham et al., 2013). However, existing soil carbon datasets disagree even on the continental-scale distribution of soil carbon, so it is doubtful that an existing dataset would capture more subtle features or ef-

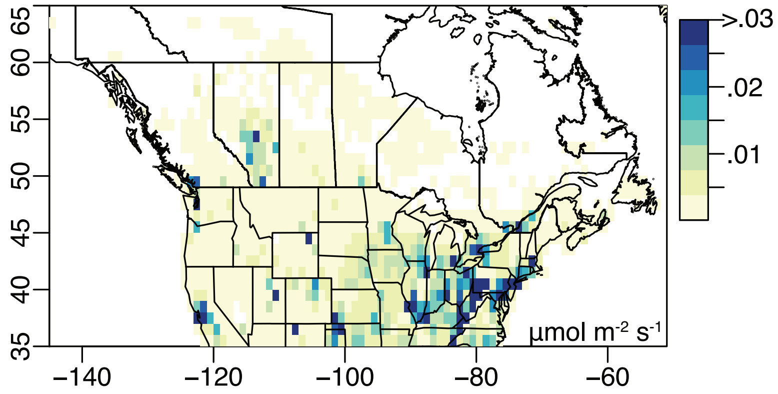


Figure B.11: The EDGAR v4.2 anthropogenic methane emissions inventory.

fects. Furthermore, several chamber studies argue that methane fluxes correlate more strongly with other environmental factors than with soil carbon (Prater et al., 2007; Olefeldt et al., 2013).

B.4 Anthropogenic methane estimates

Figure B.11 displays the EDGAR version 4.2 inventory over Canada. In the figure, the inventory is regridded to a 1° lat. by 1° lon. resolution for easier visual comparison with the inversion results in this study.

We use an alternative formulation of anthropogenic sources in the deterministic model. Specifically, we place smooth tricube functions over anthropogenic source regions (Fig B.12):

$$f(d_i, r) = \left(1 - \left(\frac{d_i}{r}\right)^3\right)^3 \tag{B.9}$$

$$d_i = \sqrt{(\text{latitude}_i - \text{latitude}_c)^2 + (\text{longitude}_i - \text{longitude}_c)^2} \tag{B.10}$$

where $f()$ is the tricube function. d_i is the latitude-longitude distance between location i and the center c of the function. Both the function center (c) and radius (r) are defined by the modeler. We define the distance d_i in terms of latitude and longitude instead of great circle distance in

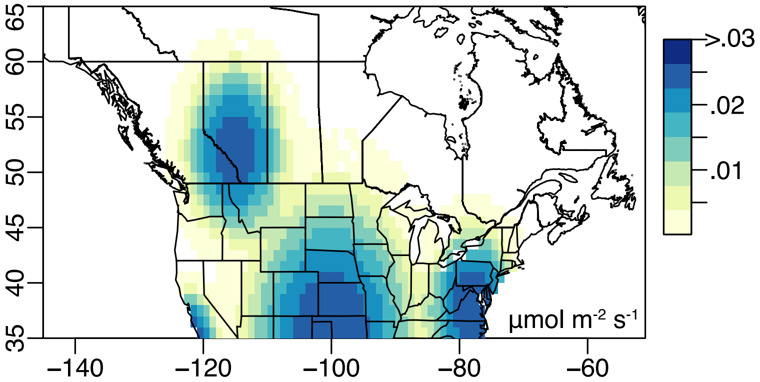


Figure B.12: The set of smooth, tri-cube functions used as a component of the deterministic model ($\mathbf{X}\beta$). In this plot, the functions are scaled by a coefficient estimated from the atmospheric methane data ($\hat{\beta}_1$, section 4.2 of the main article). Note that the Canadian observation network is not sensitive to methane emissions over most of the continental US (e.g., Fig. B.5). Instead, refer to Kort et al. (2008, 2010) or Miller et al. (2013) for top-down analysis of US methane emissions.

order to better fit the shape of anthropogenic source regions. Most large anthropogenic source regions in North America span longer north-south distances than east-west (e.g., the US Eastern Seaboard, California Central Valley, and resource extraction regions of Alberta). To this end, Eq. B.10 approximates anthropogenic source regions over North America with an ellipse that has been elongated in the north-south direction.

We test multiple tricube functions with different centers (c) and radii (r) and choose the combination of functions that produce the lowest BIC score. The functions chosen for use in the deterministic model are centered over Alberta (52° lat., -115° lon.), the southern Great Plains (35° lat., -100° lon.), the US East Coast (38° lat., -78° lon.), and California (35° lat., -122° lon.) (Fig. B.12).

B.5 Comparison with previous observational studies

The HBL methane budget estimated in this paper is higher than two previous observational studies: Roulet et al. (1992) and Worthy et al. (2000). Several factors may explain this difference. Roulet et al. (1992) extrapolated a flux estimate using ground-based measurement sites at the north and south ends of the HBL. The authors may have missed larger emissions in the central HBL, leading to a possible underestimate for the region. This point is also discussed in Pickett-Heaps et al. (2011). Worthy et al. (2000) used Fraserdale (FSD) observations and a box model with spatially-uniform flux to estimate an HBL budget. Their budget estimated using atmospheric data depends on the presumed distribution of the fluxes. For example, a modest source near the tower site or a much larger source in more distant regions could produce similar modeled concentrations at Fraserdale. Our study overcomes this distributional challenge by using multiple tower sites and by leveraging existing environmental datasets relevant to the distribution of wetland fluxes.

C

Supplemental material for chapter 5

This supplement provides more detail on the atmospheric observations, the wetland methane flux estimates, and the statistical methods used throughout the paper.

C.1 Atmospheric observation sites

Here we describe, in greater depth, the atmospheric methane observations collected across the US and Canada in 2007–2008. The methane analysis in the main article leverages 14,703 total methane observations. Of those measurements, 2,009 are from observation towers in Canada. These towers (from east to west) include Chibougamau, Quebec (CHM, 50°N, 74°W, 30m above ground level); Fraserdale, Ontario (FSD, 50°N, 83°W, 40m agl); East Trout Lake, Saskatchewan (ETL, 54°N, 104°W, 105m agl); and Candle Lake, Saskatchewan (CDL, 54°N, 105°W, 30m agl, 2007 only). These sites, operated by Environment Canada, measure methane continuously. In this study, as in Miller et al. (2014c), we use only afternoon averages of the methane data and WRF-STILT model output (1pm - 7pm local time); small scale heterogeneities in the continuous data caused by turbulent eddies and incomplete mixing make it difficult to model finer scale temporal patterns in the data. The 2,009 observations at these Canadian sites represent the total after averaging.

An additional 4,984 methane observations were collected from National Oceanic and Atmospheric Administration (NOAA) towers across the United States. These observations include daily flask samples from the NOAA tall tower network (weekly at Argyle and Ponca City): Argyle, Maine (AMT, 45 °N, 69 °W, 107m above ground level (agl)); Erie, Colorado (BAO, 40 °N, 105°W, 300m agl); Park Falls, Wisconsin (LEF, 46°N, 90°W, 244m agl), Martha's Vineyard, Massachusetts (MVY, 41°N, 71°W, 12m agl); Niwot Ridge and Niwot Forest, Colorado (NWF, NWR, 40°N, 105°W, 2,3,23m agl); Ponca City, Oklahoma (SGP, 37°N, 97°W, 60m agl); West Branch, Iowa (WBI, 42°N, 93°W, 379m agl); Walnut Grove, California (WGC, 38°N, 121°W, 91m agl), and Moody, Texas (WKT, 31°N, 97°W, 122, 457m agl).

A further 7710 methane measurements were obtained from flask samples on regular NOAA aircraft flights and from the START08 (Stratosphere-Troposphere Analyses of Regional Transport 2008) measurement campaign (Pan et al., 2010). As in Miller et al. (2013), we only use aircraft observations up to 2500m above ground level. Observations at higher altitudes are less sensitive to surface emissions and were instead used by Miller et al. (2013) to optimize the empirical methane boundary condition. In this study, we only use aircraft and tower-based observations collected during daytime hours.

We further screen the data for biomass burning influence at the Canadian sites and at Park Falls, Wisconsin. At the these sites, we remove all days with CO that peaks above 200 ppb, as was done in Miller et al. (2014c). When these sites see influence from distant anthropogenic emissions, CO is often elevated, but it rarely exceeds 200 ppb except during time periods with known fires (Miller et al., 2008).

C.2 WETCHIMP methane flux models

This section of the supplement details the WETCHIMP methane estimates from Melton et al. (2013a) and Wania et al. (2013). The seven methane estimates used in this study are shown in Fig. C.1. The wetland methane fluxes estimated by these models varies widely – both in magnitude and in spatial distribution. For example, the SDGVM model places large fluxes over the US Corn Belt relative to other regions while other models like Orchidee place large fluxes in Northern Canada that extent far into the Northwest Territories. For a more in-depth inter-comparison of these flux estimates, refer to Melton et al. (2013a) and Wania et al. (2013).

C.3 The synthetic data

In the main article, we use synthetic atmospheric methane data to explore the sensitivity of atmospheric observations to wetland fluxes (sections 5.2.3 and 5.3). This section describes in greater detail how we construct this synthetic data. The synthetic observations include contributions from anthropogenic sources, from wetlands, and from simulated model and measurement errors:

$$z_{\text{synthetic}} = \mathbf{H}(s_{\text{anthro}} + s_{\text{wetland}}) + \varepsilon \tag{C.1}$$

In this equation, $z_{\text{synthetic}}$ $(n \times 1)$ represents the synthetic observations generated for an observation site. The vector s_{anthro} $(m \times 1)$ represents emissions from anthropogenic sources and s_{wetland} $(m \times 1)$ represents wetland fluxes. The footprint or sensitivity matrix $\mathbf{H}$ $(n \times m)$, generated from WRF-STILT, models the impact of these emissions at the observation sites.

In this study, we use the a priori anthropogenic emissions estimates from Miller et al. (2013) and Miller et al. (2014c) for s_{anthro}. Those studies used activity data from the EDGAR inventory and a model selection framework to construct a prior anthropogenic emissions estimate. These

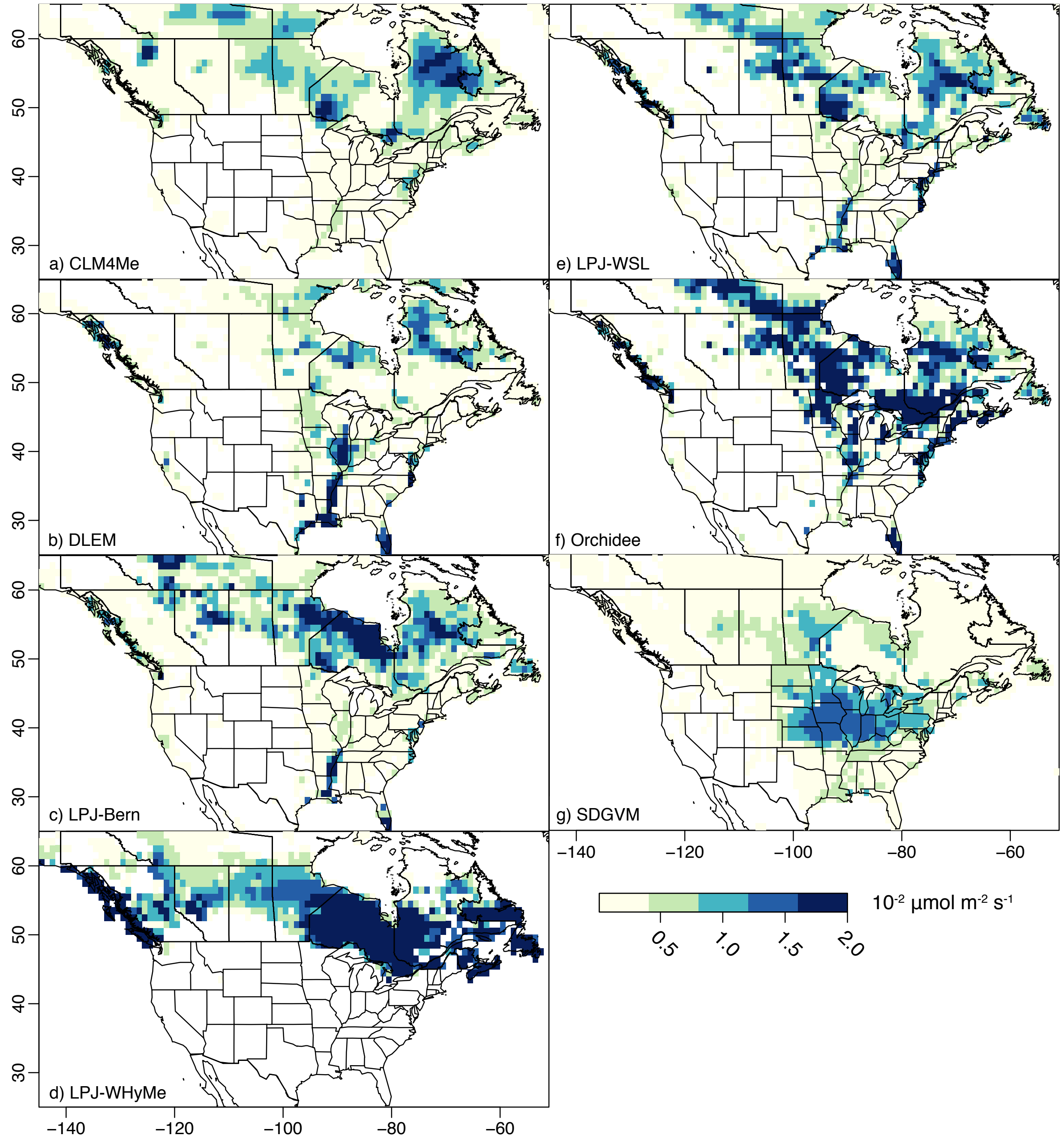


Figure C.1: Annual mean wetland methane fluxes from seven different WETCHIMP estimates (Melton et al., 2013a; Wania et al., 2013). The fluxes shown here are averaged over the 1993-2004 study period. Note that the fluxes shown above are averaged over the entire grid cell, not per m^2 of wetlands.

EDGAR activity datasets include economic or demographic data that may predict the spatial distribution of methane emissions (e.g., human or ruminant population maps).

The wetland fluxes (s_{wetland}) in Eq. C.1 are taken from the WETCHIMP methane flux models (experiment two in Melton et al. (2013a)). We use only four of the seven WETCHIMP models to generate synthetic data: CLM4Me, DLEM, LPJ-WSL, and SDGVM. These models have an overall magnitude that most closely matches the methane budgets estimated by three recent top-down studies over Canada's Hudson Bay Lowlands (Pickett-Heaps et al., 2011; Miller et al., 2014c; Wecht et al., 2014). The magnitude of these four models is likely the most realistic among the WETCHIMP flux estimates. The other WETCHIMP models, in contrast, predict much higher fluxes (Fig. 7.3).

As in Miller et al. (2013) and Miller et al. (2014c), the emissions (s_{anthro} and s_{wetland}) are regridded to a spatial resolution of 1° latitude by 1° longitude. The EDGAR activity data do not have any seasonality, so the anthropogenic emissions (s_{anthro}) are seasonally invariant. The WETCHIMP models have a monthly temporal resolution, as in Melton et al. (2013a). That study provides flux estimates for the years 1993-2004; we use the mean of these ten years for all analysis in this study.

The final term in equation C.1, ε ($n \times 1$), represents simulated errors in the measurements, in WRF-STILT, and in the fluxes (s_{anthro} and s_{wetland}). The magnitude and spatial/temporal structure of these errors were estimated in Miller et al. (2013) for the US and Miller et al. (2014c) for Canada. The remainder of this section details the specific calculations for simulating ε.

The errors in ε are distributed according to the covariance matrix $\mathbf{\Psi}$ ($n \times n$) (Eq. 5.1):

$$\varepsilon \sim \mathcal{N}(0, \mathbf{\Psi}) \tag{C.2}$$

$$\mathbf{\Psi} = \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R} \tag{C.3}$$

The variances and covariances within $\mathbf{\Psi}$ fall into two different categories. The first category are errors due to imperfect emissions, described by covariance matrix $\mathbf{Q}$ ($m \times m$). In atmospheric inversion studies, this matrix is typically termed the a priori covariance matrix. The diagonal elements of $\mathbf{Q}$ describe a set of variances – differences between the prior and the unknown true emissions over long spatial or temporal scales. The off-diagonal elements of $\mathbf{Q}$ describe any spatial and/or temporal covariances in these differences. In Eq. C.3, the footprint or sensitivity matrix ($\mathbf{H}$) projects $\mathbf{Q}$ from units of $(\text{flux})^2$ into units of parts per billion squared, $(\text{ppb})^2$.

We refer to the second type of errors as model-data mismatch errors, denoted by covariance matrix $\mathbf{R}$ ($n \times n$). These include all errors in the WRF-STILT model or the measurements that are unrelated to an imperfect flux estimate. Examples of model-data mismatch errors include measurement error, atmospheric transport error, or errors due to the spatial or temporal resolution of WRF-STILT. Over the United States, we use values for $\mathbf{R}$ and $\mathbf{Q}$ that were estimated by Miller et al. (2013) using WRF-STILT and the same atmospheric methane observations used in this study. Similarly, we use values for $\mathbf{R}$ and $\mathbf{Q}$ over Canada that were estimated in Miller et al. (2014c), a parallel inverse modeling study over that country.

In order to simulate ε, we next compute the Cholesky decomposition of $\mathbf{\Psi}$:

$$\mathbf{\Psi} = \mathbf{C}\mathbf{C}^T \tag{C.4}$$

The covariance matrix $\mathbf{\Psi}$ has units of $(\text{ppb})^2$, but its Cholesky decomposition ($\mathbf{C}$) has units of ppb, a fact that will become useful in the next step. With this decomposition in hand, we simulate a set of errors, ε (e.g., Fang et al., 2014; Shiga et al., 2014):

$$\varepsilon = \mathbf{C}u \tag{C.5}$$

$$u \sim \mathcal{N}(0, 1) \tag{C.6}$$

where u represents a set of randomly-generated numbers with a mean of zero and variance of one.

We simulate 1000 synthetic datasets for each experiment to adequately sample the random errors in ε. We then use the model selection framework to find the optimal candidate model for each of these datasets. The results presented in Fig. 5.3 are therefore the composite of thousands of model selection runs: one model selection run for each synthetic dataset. We use a branch and bound algorithm from Yadav et al. (2013) to improve the computational efficiency of these model selection runs. Furthermore, we estimate the coefficients (β) in Eq. 5.1 using Lagrange multipliers to ensure that none of the estimated coefficients have unrealistic negative values (e.g., Miller et al., 2014a).

C.4 Sensitivity of the observation network to surface fluxes

In this section, we describe the overall sensitivity of the observation network to methane fluxes. This sensitivity will play at least some role in network's ability to identify a signal from wetlands. The WRF-STILT model quantifies this sensitivity in terms of a footprint. Each row the matrix **H** is the footprint associated with a different atmospheric methane observation. In Fig. C.2, we plot these footprints, summed over all of 2007–2008.

This figure show several distinctive patterns. First, the US network has a higher sensitivity than the Canadian network. This pattern is due to the larger number of observation sites over the US. Second, the highest sensitivities are clustered in distinctive regions with multiple observation sites – Wisconsin, Texas/Oklahoma, and California, among other regions.

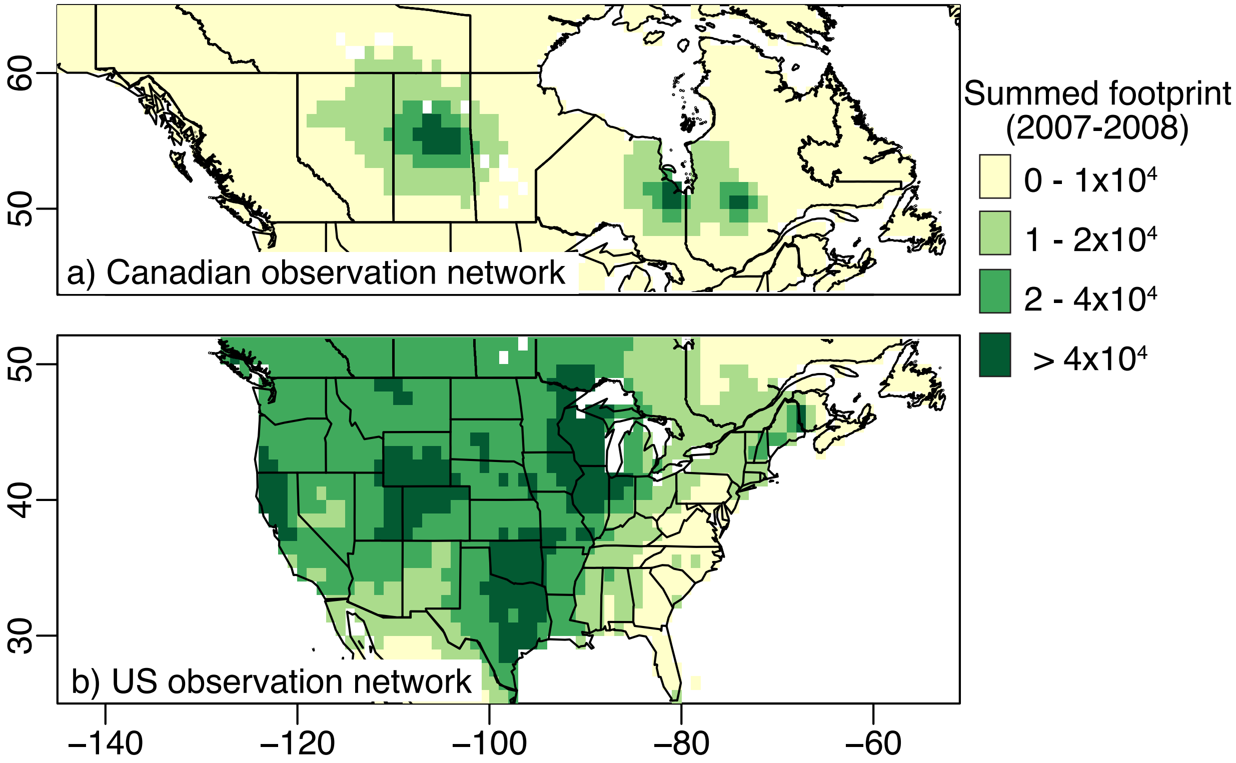


Figure C.2: Total, summed footprint from the (a) Canadian and (b) US observation networks. The observation sites incorporated into this figure are shown in Fig. 5.2. Each individual footprint (associated with an individual atmospheric observation) has units of concentration per unit flux (ppb per μmol m^{-2} s^{-1}). In this figure, we sum all footprints for 2007–2008.

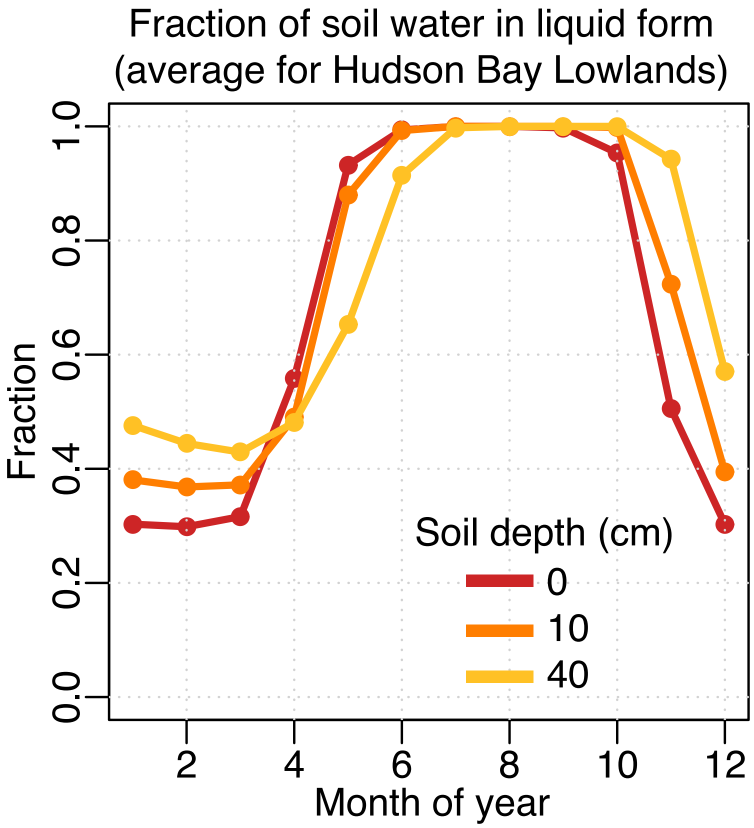


Figure C.3: This figure displays the fraction of soil water that is unfrozen for the HBL in different seasons and at different soil depths. Estimates are taken from NARR (Mesinger et al., 2006).

C.5 Soil freeze/thaw estimates from NARR

Figure C.3 shows the soil freeze/thaw cycle at different depths across the Hudson Bay Lowland. These estimates are taken from North American Regional Reanalysis (NARR) (Mesinger et al., 2006), and the values shown in Fig. C.3 are average values for the HBL in each month. The main article references this figure in a discussion of the methane flux seasonal cycle (Sect. 5.4.3).

D

Supplemental material for chapter 6

This supplement describes in greater detail the multiple-try Metropolis-Hastings algorithm and the Gibbs sampler implementation.

D.1 Existing Metropolis-Hastings implementations

Metropolis-Hastings algorithms can be used in a number of ways to enforce inequality constraints, and existing literature in hydrology (Michalak et al., 2004; Wang & Zabaras, 2006; Zanini & Kitanidis, 2009) and atmospheric sciences (Rigby et al., 2011; Burrows et al., 2013) implement different approaches. The algorithms implemented in the cited hydrology papers use Lagrange multipliers to enforce constraints on individual realizations. Such an approach is computationally attractive for larger-scale inverse problems, with hundreds or thousands of elements of the state space for which constraints must be enforced. The cited studies generally use the following approach:

1. Add noise to the observations and the prior. The vectors of random noise are generated using the covariance matrices (**R** and **Q**, respectively) and the last accepted realization.

2. Create a conditional, constrained realization of the emissions ($s_{cc,j}$) by minimizing the inversion cost function subject to the randomly-generated inputs above and enforcing constraints on the state vector using the method of Lagrange multipliers.

3. Calculate the likelihood of the candidate realization ($s_{cc,j}$) relative to the previous, accepted realization ($s_{cc,j-1}$). Based upon this likelihood, either accept or reject the realization, and begin again with step 1.

These steps are described in more detail in section D.2.

The cited existing atmospheric studies use a different approach (Rigby et al., 2011; Burrows et al., 2013). In these studies, a new realization ($s_{cc,j}$) is generated by adding a random quantity directly to the previous candidate realization ($s_{cc,j-1}$). If the candidate realization ($s_{cc,j}$) does not obey the inequality constraints, it is discarded. If the constraints are met in the realization, it is again accepted or rejected based on its relative probability compared to the last accepted realization. This implementation omits steps 1 and 2 above. In other words, the modeler is not required to minimize the cost function and can instead skip directly to step 3. This feature affords greater flexibility in the inversion setup (e.g., Rigby et al., 2011); the modeler can use a complex probability density function (pdf) for the inversion even if calculating its maximum is computationally-intractable. A larger number of multivariate distributions could be used for the prior pdf, not just a multivariate Gaussian distribution (implemented with Lagrange multipliers in the hydrology studies above). The fraction of realizations that are rejected, however, can be very high for two reasons. First, for large state vectors, a very high fraction of candidate realizations will not obey all bounds and will be rejected. Second, there are no formal steps to ensure that each proposed realization reproduces the atmospheric observations. Therefore, the fraction of realizations that are rejected will be high, even among those realizations that obey the inequality constraints.

The implementation used in the cited hydrology studies is therefore better for larger problems because it samples the posterior probability space more strategically. The implementation generates candidate realizations that explicitly sample the covariance matrices ($\mathbf{R}$ and $\mathbf{Q}$) and honor the observations (z). Refer to Chib & Greenberg (1995) or Bolstad (2012) for a general discussion on creating candidate realizations. In general, larger problems require strategic generation of candidate realizations; the efficiency and/or acceptance probability of Metropolis-Hastings decreases as the number of unknowns increase for problems with comparable implementation and covariance matrices (Gelman, 2004, ch. 11). This fact makes the approach used in the cited hydrology literature more suitable for larger problems like the methane case study in the main manuscript. However, this approach also restricts the inversion setup to pdfs with computationally-tractable maxima.

D.2 The multiple-try Metropolis-Hastings

The following section describes the multiple-try Metropolis-Hastings algorithm modified to accommodate inequality constraints. This algorithm is more computationally tractable for large problems than many traditional Metropolis-Hastings implementations.

This algorithm (Liu et al., 2000) first requires the generation of an unconstrained unconditional realization, denoted s_{uu}. The realization for step j, denoted $s_{uu,j}$, is created by applying a modification to $s_{uu,j-1}$. The modification to the previous realization is provided by what is known as the jumping distribution $T()$. This distribution should create new realizations that are sufficiently different from the previous one such that the algorithm effectively samples the entire probability space. However, the jumping distribution should avoid creating subsequent realizations that are so different such that $s_{uu,j}$ gets rejected by the algorithm (e.g., Chib & Greenberg, 1995).

The jumping distribution used here requires taking the Cholesky decomposition of **Q**:

$$\mathbf{Q} = \mathbf{C}\mathbf{C}^T \tag{D.1}$$

The distribution $T()$ can be chosen in any number of ways (e.g., Chib & Greenberg, 1995), but we generate new unconditional unconstrained realizations as follows (where u is a random vector with distribution $\mathcal{N}(0,1)$):

$$\begin{aligned} s_{uu,0} &= \mathbf{C}u \\ s_{uu,j} &= \phi s_{uu,j-1} \; + \; \sqrt{1-\phi^2}\,\mathbf{C}u \end{aligned} \tag{D.2}$$

In this case, we set $\phi = 0.9$, though any value greater than zero and less than one is acceptable (e.g., Michalak et al., 2004). The multiple-try Metropolis-Hasting with inequality constraints has the following steps:

1. Draw k trial proposals for $s_{uu,j}$ from the jumping distribution described by Eq. D.2.
2. Compute a conditional constrained realization ($s_{cc,j}$) for each of the trial proposals by minimizing the posterior negative log-likelihood via Lagrange multipliers:

   $$\begin{aligned} L_{s,\beta} \;=\; & \tfrac{1}{2}(z+v-\mathbf{H}s^*_{cc,j})^T\mathbf{R}^{-1}(z+v-\mathbf{H}s^*_{cc,j})+ \\ & \tfrac{1}{2}(s^*_{cc,j}-s^*_{uu,j})^T\mathbf{G}(s^*_{cc,j}-s^*_{uu,j}) \end{aligned} \tag{D.3}$$

   where v is a random vector with covariance **R**. In this case, the asterisk ($*$) indicates that the candidate is one of k trial proposals for the realization.

3. Compute the weighting function for each trial proposal:

$$w(s^*_{cc,j}|s_{cc,j-1}) = \frac{p''(s^*_{cc,j}|z, \mathbf{H}, \mathbf{X})}{T(s^*_{cc,j}|s_{cc,j-1})} \tag{D.4}$$

   where $p''(s^*_{cc,j}|z, \mathbf{H}, \mathbf{X})$ indicates the posterior probability of $s^*_{cc,j}$, and $T(s^*_{cc,j}|s_{cc,j-1})$ is the jumping probability of $s^*_{cc,j}$ given $s_{cc,j-1}$. The posterior probability and approximate jumping probability can be calculated as follows (Michalak et al., 2004):

$$\begin{aligned} p''(s^*_{cc,j}|z, \mathbf{H}, \mathbf{X}) \;&\propto\; \exp[-\tfrac{1}{2}(z - \mathbf{H}s^*_{cc,j})^T\mathbf{R}^{-1}(z - \mathbf{H}s^*_{cc,j}) \\ &\quad - \tfrac{1}{2}{s^*_{cc,j}}^T\mathbf{G}s^*_{cc,j}] \qquad \text{(D.5)} \\ T(s^*_{cc,j}|s_{cc,j-1}) \;&\propto\; \exp[-\tfrac{1}{2}(s^*_{uu,j} - \phi s_{uu,j-1})^T \\ &\quad \frac{\mathbf{Q}^{-1}}{(1-\phi^2)}(s^*_{uu,j} - \phi s_{uu,j-1})] \qquad \text{(D.6)} \end{aligned}$$

4. Select $s_{cc,j}$ from the trial proposals by individually, randomly drawing each element from $s^*_{cc,j}$ with probability proportional to the weighting function $w(s^*_{cc,j}|s_{cc,j-1})$. Select the corresponding elements of $s^*_{uu,j}$ to construct $s_{uu,j}$.

5. Create $(k-1)$ new trial proposals for $s_{cc,j-1}$. To do this, draw samples from the jumping distribution $T(s^*_{uu,j-1}|s_{uu,j})$ (i.e., $s^*_{uu,j-1} = \phi s_{uu,j} + \sqrt{1-\phi^2}\mathbf{C}u$). Calculate the trial conditional constrained realizations $s^*_{cc,j-1}$ using the procedure outlined in step 2. Set trail proposal k to $s_{cc,j-1}$. Finally, calculate the weighting function for each trial conditional constrained realization, $w(s^*_{cc,j-1}|s_{cc,j})$.

6. Calculate the acceptance/rejection probability (Liu et al., 2000):

$$\xi = \min\left\{1, \frac{\sum^k w(s^*_{cc,j}|s_{cc,j-1})}{\sum^k w(s^*_{cc,j-1}|s_{cc,j})}\right\} \tag{D.7}$$

Accept $s_{cc,j}$ if $\xi > \mathcal{U}(0,1)$. Otherwise, set $s_{cc,j} = s_{cc,j-1}$.

Repeat steps 1 – 6 until a sufficient number of realizations have been generated to sample across the entire posterior probability space. Note that unlike the Gibbs sampler, this multiple-try Metropolis-Hastings algorithm does not require discarding realizations from an initial spin-up period. For this application, we choose $k = 8$. Larger values for k can lead to greater acceptance rates but higher computational cost. Liu et al. (2000) note that an acceptance rate of $0.4-0.5$ is ideal for a multiple-try Metropolis-Hastings algorithm.

D.3 The Gibbs sampler implementation

The Gibbs sampler requires generating the element-wise conditional probability density, the probability of any individual element in s given an estimate of all other elements in s. This conditional density is denoted $p(s_i|s, z)$ where i is one of m elements in s. The equations for $p(s_i|s, z)$ can be found in Michalak (2008) for the inversion setup discussed in this paper.

The Gibbs sampler has the following steps:

1. Make an initial guess for s_1 where the subscript '1' denotes the first realization of s.

2. Obtain a new realization, s_j, from the previous realization, s_{j-1}. To do this, successively generate a conditional probability for each element in s, and draw a random sample from

each one:

$$\begin{aligned}
p(s_{1,j}) &= p(s_{1,j}|s_{2,j-1}, ..., s_{m,j-1}) \\
p(s_{2,j}) &= p(s_{2,j}|s_{1,j}, s_{3,j-1}, ..., s_{m,j-1}) \\
p(s_{i,j}) &= p(s_{i,j}|s_{1,j}, ..., s_{i-1,j}, s_{i+1,j-1}, ..., s_{m,j-1}) \\
p(s_{m,j}) &= p(s_{m,j}|s_{1,j}, ..., s_{m-1,j})
\end{aligned} \tag{D.8}$$

3. Update j to $j+1$ and continue generating realizations.

Create a large number of realizations (in this case 1200) to fully sample across the posterior probability space. The initial realizations are usually discarded as a “spin-up” period (in this case, the first 200).

In this implementation $p(s_i|s,z)$ is Gaussian. To enforce the inequality constraints, Michalak (2008) draws a random sample from $p(s_i|s,z)$ until the random draw falls within the bounds. This draw becomes the estimate for $s_{i,j}$. The approach is equivalent to using a truncated Gaussian as a the prior probability density function, but this implementation avoids the computational challenge of directly computing a multivariate, truncated probability distribution.

This study uses a modified approach for the methane case study. If the random sample from $p(s_i|s,z)$ is positive, it becomes the estimate for $s_{i,j}$. If the random sample is negative, set $s_{i,j} = 0$. This approach is equivalent to sampling from a truncated normal distribution with an added Dirac delta function. The method adapted here increases the probability of estimating zero emissions for a given flux or emissions location.

E

Supplemental material for chapter 7

E.1 The meteorological model-data assimilation framework

This section of the supplement describes the Local Ensemble Kalman Filter (LETKF) in greater detail. Many of the equations listed below are abbreviated versions of those detailed in Hunt et al. (2004), Desroziers et al. (2005), Hunt et al. (2007), Li et al. (2009), and Miyoshi (2011). For a mathematical derivation of either the meteorology optimization or covariance matrix estimation within LETKF, refer to any of those studies.

The model–data assimilation system (abbreviated CAM–LETKF) can be summarized in a number of steps. First, we create an initial condition for modeled meteorology, in this case using NCEP–DOE AMIP-II reanalysis (Kanamitsu et al., 2002). We generate a set of small perturbations to the initial conditions and use these perturbations to create a set of k initial conditions that are all slightly different. In this case, we set $k = 64$ (as in Liu et al., 2011, 2012). This choice represents a compromise between thorough statistical sampling and computational considerations: a very large k will exhaustively sample the model uncertainties. However, k CAM–CLM realizations require $4k$ computer cores, so a very large k would also become computationally prohibitive.

Second, we run a 6-hour weather forecast using CAM-CLM for each of the k model initial conditions. The spread of this model ensemble represents our prior uncertainty in the modeled meteorology:

$$x_i = \bar{x} + \mathbf{X}_i \quad \text{where } i = 1....k \tag{E.1}$$

where x_i ($m \times 1$) is a single model realization, $\bar{x}$ ($m \times 1$) is the mean of the model ensemble, and $\mathbf{X}_i$ ($m \times k$) refers to the i^{th} column of the matrix that defines the model ensemble spread. In the main article (e.g., Eq. 1), we defined these variables to refer to all model time steps, collectively. In the supplement, by contrast, we will instead define these variables to refer to the model–data assimilation at a single, 6 hourly time step. In other words, m and n now refer to the model outputs and number of weather observations, respectively, associated with a single model–data assimilation cycle. This redefinition of the variables facilitates a discussion of time-stepping in the remainder of this section.

Third, we calculate a set of k weights such that the weighted average of the realizations best matches the meteorological observations:

$$\bar{x}^a = \bar{x}^b + \mathbf{X}^b \hat{w} \tag{E.2}$$

The superscript b refers to the model state before assimilation and a the model state after data assimilation. The $k \times 1$ vector of weights ($\hat{w}$) are estimated by minimizing a statistical cost function with respect to the meteorological observations (Hunt et al., 2007):

$$J(w) = (k-1)w^T w + \left(z - H(\bar{x}^b + \mathbf{X}^b w)\right)^T \mathbf{R}^{-1} \left(z - H(\bar{x}^b + \mathbf{X}^b w)\right) \tag{E.3}$$

In the above equation, z ($n \times 1$) represents the meteorological observations, and $H()$ is a function or operator that maps the model output to the observations. For example, the function $H()$ may convert the model units to the measurement units or may interpolate the model output to an observation site that lies between multiple model grid boxes. Lastly, the diagonal matrix $\mathbf{R}$ ($n \times n$) represents the nugget variance, variance in the model–data residuals that is due to measurement errors or meteorological processes too small in scale to be captured by CAM–CLM.

The following equation further elucidates the role of the $\mathbf{R}$ covariance matrix in the analysis (e.g., Hunt et al., 2007):

$$\mathbf{z} \;=\; H(x) + \mathcal{N}(0, \mathbf{R}) \tag{E.4}$$

In the above equation, x refers to the true value of the meteorological parameters averaged to the resolution of the model grid. This vector is an unknown quantity, and the estimated values ($\bar{x}^a$) are only a best guess of this unknown quantity. Any differences between the true, unknown values and the measurements must be due to one of two causes: either measurement error or errors due to the finite model resolution. These errors should follow a normal distribution with a mean of zero and a covariance matrix $\mathbf{R}$. We estimate the elements of this matrix using actual model-data residuals, as described in greater detail below.

In order to estimate the weights (w), we use a localization procedure. In practice, we never compute the weights simultaneously for the entire global model output. Rather, we estimate a different set of weights for each model grid box using model-measurement pairs within a certain radius (in this case, within 1500km). As such, the matrices in Eqs. E.2 and E.3 represent a subset of the global model output, and the dimensions n and m are small relative to the total number of global observations and model grid boxes, respectively.

As part of this localization procedure, we interpolate the gridded model output to the observation locations and times; we use these model-measurement pairs to compute each set of weights. We further taper the influence of model-observation pairs on the optimization depending on their distance from the grid box in question (using a Blackman window function as described by Oppenheim & Schafer (1989) and Liu et al. (2012)). Hence, model-measurement pairs located within the model grid box of interest will influence the optimization much more strongly than model-observation pairs located 1000km away. A radius of 1500km for the Blackman window function is comparable to values used throughout the meteorological literature. For example, Liu et al. (2011) and Liu et al. (2012) also used a 1500km radius. Furthermore, Miyoshi (2011) set a 1825 km radius of influence, Miyoshi & Kunii (2012) used a 1460km radius, and Szunyogh et al. (2008) used an 800km radius.

This localization approach ensures coherence among adjacent grid boxes and ensures that the optimization is not an over-fit to the data. For example, if we estimated the weights using only model-measurement pairs in the grid box of interest, several problems could arise. First, there may not be many relevant observations that are sensitive to that specific grid box, particularly over the open ocean or near the poles. In those circumstances, the estimated weights could be inaccurate. Second, that approach could produce vastly different weights in adjacent grid boxes, a result that is unlikely to be physically realistic. For example, the estimated weights for one model grid box over eastern North Dakota should look somewhat similar to the weights for a grid box over western North Dakota. If the two sets of weights were completely unrelated, one could argue that the optimization would be an over-fit. A localization radius of 1500km precludes these potential problems.

The weights ($\hat{w}$), estimated using this localization procedure, will have the following covariance matrix ($k \times k$) (Hunt et al., 2007):

$$\tilde{\mathbf{P}}^a = (k-1)\mathbf{I} + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b \tag{E.5}$$

$$\text{where} \quad H(\bar{x}^b + \mathbf{X}^b w) \approx \bar{y}^b + \mathbf{Y}^b w \tag{E.6}$$

Fourth, we generate 64 realizations that collectively represent our posterior uncertainty in the meteorology. Like the best estimate ($\bar{x}^a$), these posterior realizations are also a linear combination of the prior model realizations (Hunt et al., 2007):

$$x_i^a = \bar{x}^a + \mathbf{X}^b \left([(k-1)\tilde{\mathbf{P}}^a]^{\frac{1}{2}} \right)_i \tag{E.7}$$

where '$\frac{1}{2}$' denotes the symmetric square root of the covariance matrix. The subscript i on the right hand side of the equation refers to individual columns of the matrix.

Fifth, and finally, we adjust the overall model ensemble spread to match the model uncertainties implied by the meteorological observations. We refer to this process as adaptive covariance inflation (e.g., Li et al., 2009; Miyoshi, 2011). Note that this step is new since previous CAM–LETKF studies by Liu et al. (2011) and Liu et al. (2012).

Adaptive inflation operates on the following principle: the ensemble variance and nugget variance should match against the actual model-data residuals (e.g., Li et al., 2009):

$$E\left[\left(z - H(\bar{x}^b) \right) \left(z - H(\bar{x}^b) \right)^T \right] = \mathbf{HPH}^T + \mathbf{R} \tag{E.8}$$

$$\text{where} \quad \mathbf{P} = (k-1)^{-1} \mathbf{X}^b (\mathbf{X}^b)^T \tag{E.9}$$

In that equation, E denotes the expected value, and the matrix $\mathbf{H}$ ($n \times m$) is a linearization of the function $H()$. In practice, however, these covariance matrices can diverge from the actual residuals (Refer to Miyoshi (2011) for more detail.). Therefore, we estimate a scaling factor (α) for the diagonal elements of the covariance matrix $\mathbf{P}$ ($m \times m$). This scaling factor can be estimated by manipulating Eq. E.8 as in Li et al. (2009) and Miyoshi (2011):

$$\alpha = \frac{tr\left[\left(z - H(\bar{x}^b)\right)\left(z - H(\bar{x}^b)\right)^T \circ \mathbf{R}^{-1}\right] - n}{tr\left[\mathbf{HPH}^T \circ \mathbf{R}^{-1}\right]} \tag{E.10}$$

In this equation, tr refers to the matrix trace, and the symbol $\circ$ indicates element-wise multiplication. The result of Eq. E.10 is then weighted against the scaling factor from the previous model time step to produce a final scaling factor estimate (refer to Li et al., 2009; Miyoshi, 2011).

To date, the use of a single scaling factor (α) per grid box has been a standard practice in ensemble Kalman filters applied to weather models (e.g., Szunyogh et al., 2008; Liu et al., 2011, 2012; Miyoshi & Kunii, 2012; Kang et al., 2012), and we do the same here. When we estimate a single scaling factor per box, we leverage more observations to make a more stable inflation estimate. Otherwise, adaptive inflation can become challenging to implement; adaptive inflation performs poorly when observations are sparse (e.g., Miyoshi, 2011).

We also estimate the nugget variance ($\sigma^2_{\mathbf{R},j}$) for a given observation type (j) using the model output and observations (Desroziers et al., 2005; Li et al., 2009):

$$\sigma^2_{\mathbf{R},j} = \frac{\left(z_j - H(\bar{x}^a)\right)^T \left(z_j - H(\bar{x}^b)\right)}{n_j} \tag{E.11}$$

As with α, the result in Eq. E.11 is also weighted against the estimated variance from the previous time step to produce a final variance estimate (Li et al., 2009). Unlike the localized LETKF calculations, we estimate a single nugget variance for the entire globe (for each meteorological observation type). In other words, in Eq. E.11, the inputs represent global values for observation type j, not a localized implementation as in previous equations.

After these steps, the model-assimilation cycle begins again with another 6-hour CAM–CLM forecast. The posterior ensemble members (x_i^a) become the initial conditions for this next CAM-CLM forecast.

In our study, we utilize biospheric, oceanic, biomass burning, and fossil fuel CO_2 fluxes from CarbonTracker (CT), and we do not change these fluxes in response to any parameters within CAM-CLM. The original CT fluxes have a temporal resolution of 3 h. We average these fluxes to a 6 h resolution for all of the CAM-CLM simulations in this study, the length of each model time step.

Furthermore, we use CT as the initial condition for global atmospheric CO_2 mixing ratios on 1 January and 1 May 2009. Each CAM ensemble member uses the same initial condition for atmospheric CO_2, so any subsequent differences in CO_2 among the model realizations are due entirely to meteorological uncertainties.

E.2 CAM–LETKF performance metrics

The paragraphs that follow discuss two different metrics of CAM–LETKF performance: large-scale meteorology model–data comparisons and a more in-depth view of the estimated variances (i.e., the variance inflation and the nugget variance).

First, we examine the meteorology model–data residuals for the model best-guess ($\bar{x}^a$). Figure E.1 displays the root mean squared model–measurement error (RMSE, $\sqrt{(1/n)\sum(\mathbf{y} - H(\bar{x}^a))^2}$), broken down by time and by observation type. Each point plotted in Fig. E.1 is the RMSE computed from all available global observations. This RMSE appears comparable in magnitude to several existing weather reanalysis products. For example, these statistics are similar to CAM-LETKF simulations by Liu et al. (2011), though simulations in that paper cover a much shorter time period. Furthermore, the temperature, pressure, and wind errors reported here are in the

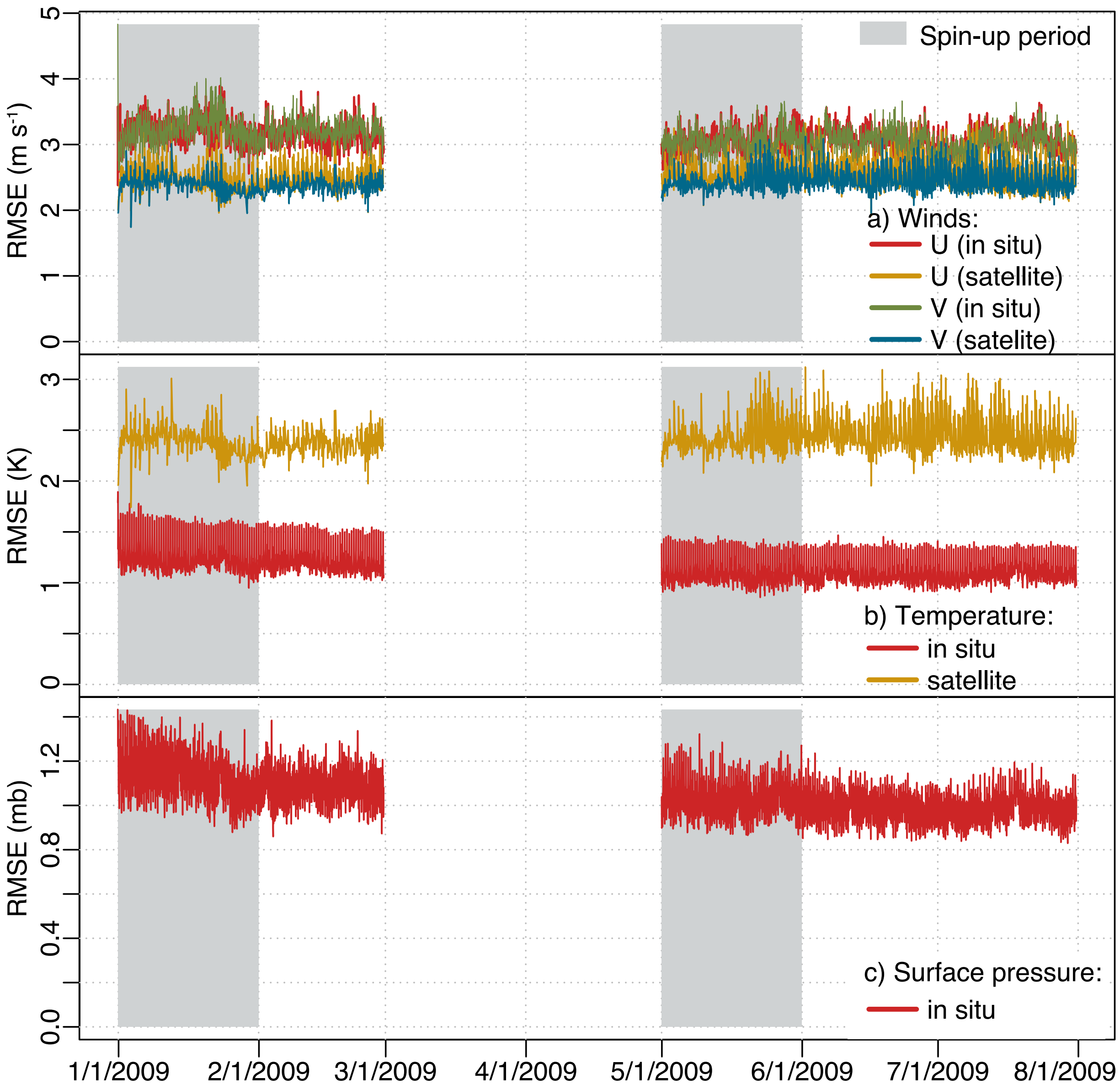


Figure E.1: Root mean squared errors for the CAM–LETKF best estimate compared against various meteorological observations (RMSE, $\sqrt{(1/n)\sum(\mathbf{y} - H(\bar{x}^a))^2}$).

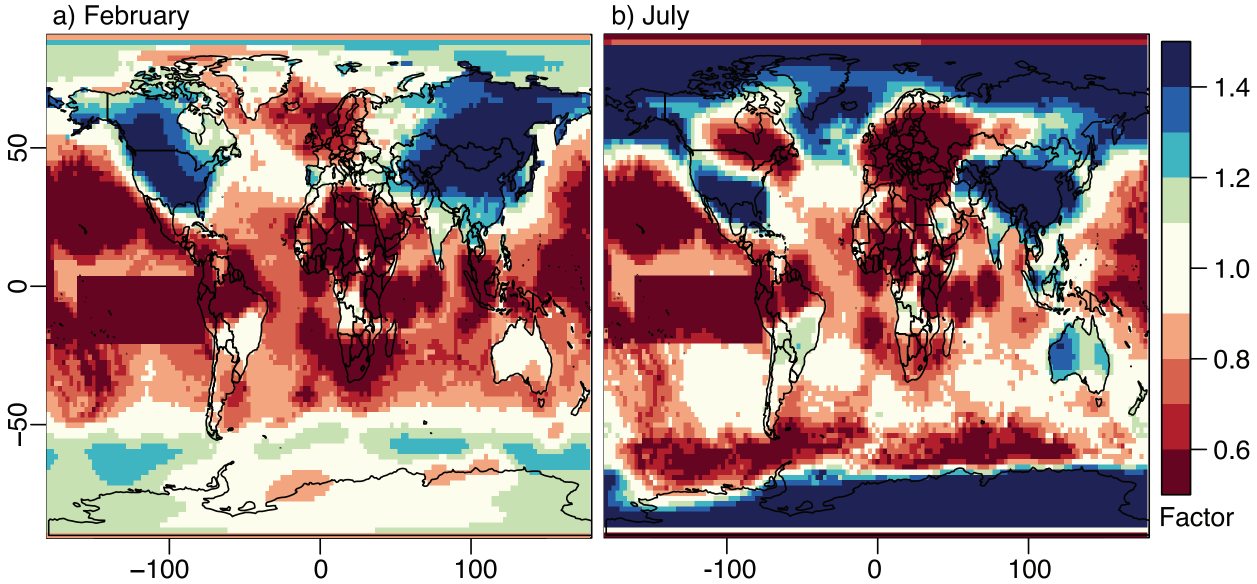


Figure E.2: The variance inflation factors for the CAM model surface layer, averaged over each 6 hourly estimation period in February and July, 2009. In this study, we estimate a different inflation factor for each model grid box and each 6 hourly estimation period. More specifically, we estimate a single inflation factor for all model parameters (e.g., wind, temperature, surface pressure, and specific humidity) in each box.

range of those listed for North American Regional Reanalysis (NARR) and ERA-Interim reanalysis (Mesinger et al., 2006; Dee et al., 2011).

The remainder of this section discusses the estimated covariance matrix parameters. Fig. E.2 displays a map of the average variance inflation factors (α) in the model surface layer for February and July, 2009, and Fig. E.3 shows how the average variance inflation factor changes over time through five months of CAM–LETKF simulations.

These figures show several notable patterns, three of which we discuss in more detail. First, the inflation factors in Fig. E.2 are highest over North America, Asia, and Australia, regions with relatively abundant meteorological observations. A number of previous studies confirm this positive relationship between data density and covariance inflation (e.g., Anderson, 2009; Miyoshi, 2011; Miyoshi & Kunii, 2012). Furthermore, Miyoshi (2011) points out that a high inflation fac-

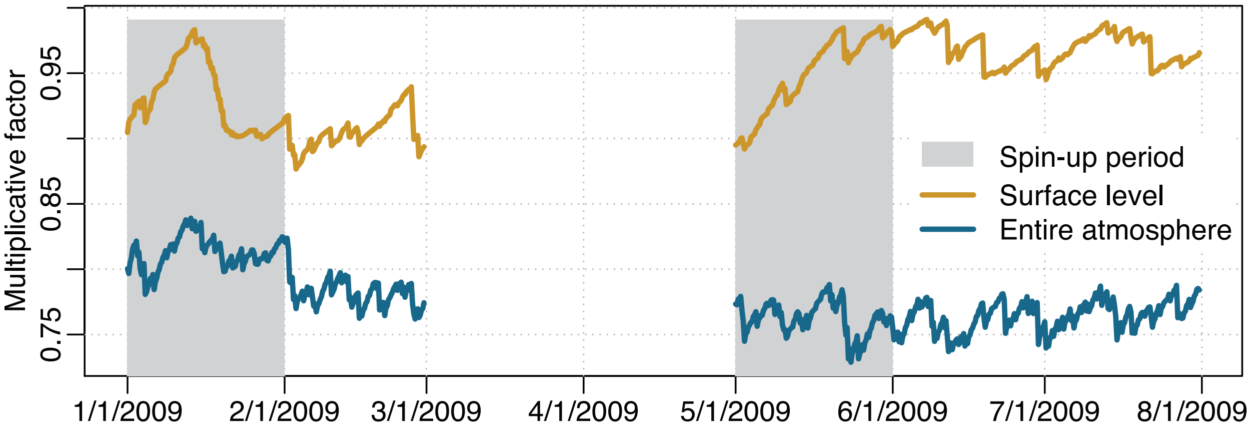


Figure E.3: Time series of the average variance inflation factors, both at the surface and for all vertical model levels. The inflation factors show some variability during the model spin-up periods, then stabilize to relatively constant values.

tor in observation-rich regions may cause the ensemble spread to be too large downwind. This explanation may account for the adjacent regions of high inflation (over continents) and low inflation (over the oceans) in Fig. E.2.

Second, the global average of the inflation factors is less than one (Fig. E.3). Even though the inflation factors, on average, decrease the ensemble variance, the global ensemble variance remains relatively constant over time. For example, the average 6 hourly model ensemble spread at meteorology observation sites is comparable in February, June, and July: ~1.5 m s^{-1} for zonal and meridional wind (standard deviation), ~0.7 K for surface temperature, and ~1.1 mb for surface pressure. This consistency, in spite of the small inflation average, may be due to the nonlinear nature of the meteorological model – differences among individual ensemble members can escalate or intensify over the 6-hour meteorology forecast.

A third notable feature is the low inflation values over eastern, tropical Pacific Ocean. These low values are intentional by design; we set a very low initial estimate for the variance inflation in this region of the globe. Higher inflation values cause the ensemble variance to increase rapidly in this region and lead to unphysical temperature estimates near the tropopause. This issue is due to an enigmatic temperature instability in the meteorological model. In the forecast stage of the CAM model, the ensemble's temperature spread in this region can increase rapidly if the initial conditions (i.e., the posterior ensemble from the previous time step) have a sufficiently large spread.

Normally, one might expect the adaptive inflation to correct for this issue; the adaptive inflation adjusts the variance of the meteorology model ensemble to match the actual model-data residuals. In theory, this procedure should prevent the ensemble spread from exploding (given sufficient data). However, the inflation factor, by design, cannot change suddenly from one time step to another. The adaptive inflation procedure uses the previous time step as the prior inflation estimate, and that prior estimate has a finite uncertainty (in this case, a prior standard deviation of 0.03 – similar to the values used by Miyoshi (2011)). Because of this prior uncertainty, the adaptive inflation factor must evolve slowly over many days if it changes at all. In most cases, this property is desirable because it prevents a small number of observations from making dramatic changes to the evolution of the model-data system. However, in the case of this temperature instability, the instability in the model develops over 4-5 model time steps, much faster than the response time of the adaptive inflation factor.

The adaptive inflation procedure requires an initial inflation estimate for the first time step of the model run (i.e., an initial condition). The adaptive procedure then updates this estimate at the each model time step (e.g., Eq. E.10). For this initial estimate or initial condition, we set a small value (0.4) for the equatorial western Pacific. During the one-month model spin-up period, the estimated inflation value evolves substantially from the initial estimate in most regions of the globe (e.g., Fig. E.3). Over this region of the Pacific, however, the estimated inflation factor does not evolve or change very much; either this initial estimate is consistent with the actual model-data residuals, or the meteorological data (and the adaptive inflation procedure) are not very informative over the region. In either case, this small initial condition prevents the ensemble spread from becoming unstable over the region.

In addition to the covariance inflation, the nugget variance also remains consistent over time. Fig. E.4 shows the square root of the nugget variance for each observation type and at each model time period. Note that we estimate different values of the nugget variances by observation type and time, but the estimated variances are spatially constant across the globe. These estimates remain consistent over time, except for the initial January spin-up period, during which the estimate slowly evolves from the initial guess.

E.3 Uncertainties in atmospheric CO_2 transport

This section of the supplement provides more detailed plots of the CO_2 transport uncertainties shown in Fig. 7.2 of the main article. In particular, the plots in this section (Figs. E.5 – E.8) visualize the transport uncertainties for different time slices of the day and show how CO_2 transport uncertainties differ between day and nighttime. The first two figures (Figs. E.5 and E.6) display the mean 6 hourly CO_2 transport uncertainties for February and July, 2009, a setup analogous to Figs. 7.2a-b in the main manuscript. Conversely, Figs. E.7 and E.8 exhibit the uncertainties in the month-long mean CO_2 concentrations, analogous to Figs. 7.2c-d in the main article.

In general, the 6 hourly uncertainties vary widely depending on the local time with higher uncertainties at night (Figs. E.5 and E.6). Note that case study one in the main article (sections 7.2.4 and 7.3.3) only uses model output associated with local afternoon CO_2 measurements. In contrast to these 6 hourly uncertainties, the uncertainties in monthly-mean concentrations do not vary as much by time of day (Figs. E.7 and E.8). For example, over North America and northern Eurasia in February, the CO_2 uncertainties are equally high during all times of day. However, a diurnal cycle in the month-long uncertainties is apparent over some regions – equatorial Africa, South America, and over Northern Hemisphere land regions in summer.

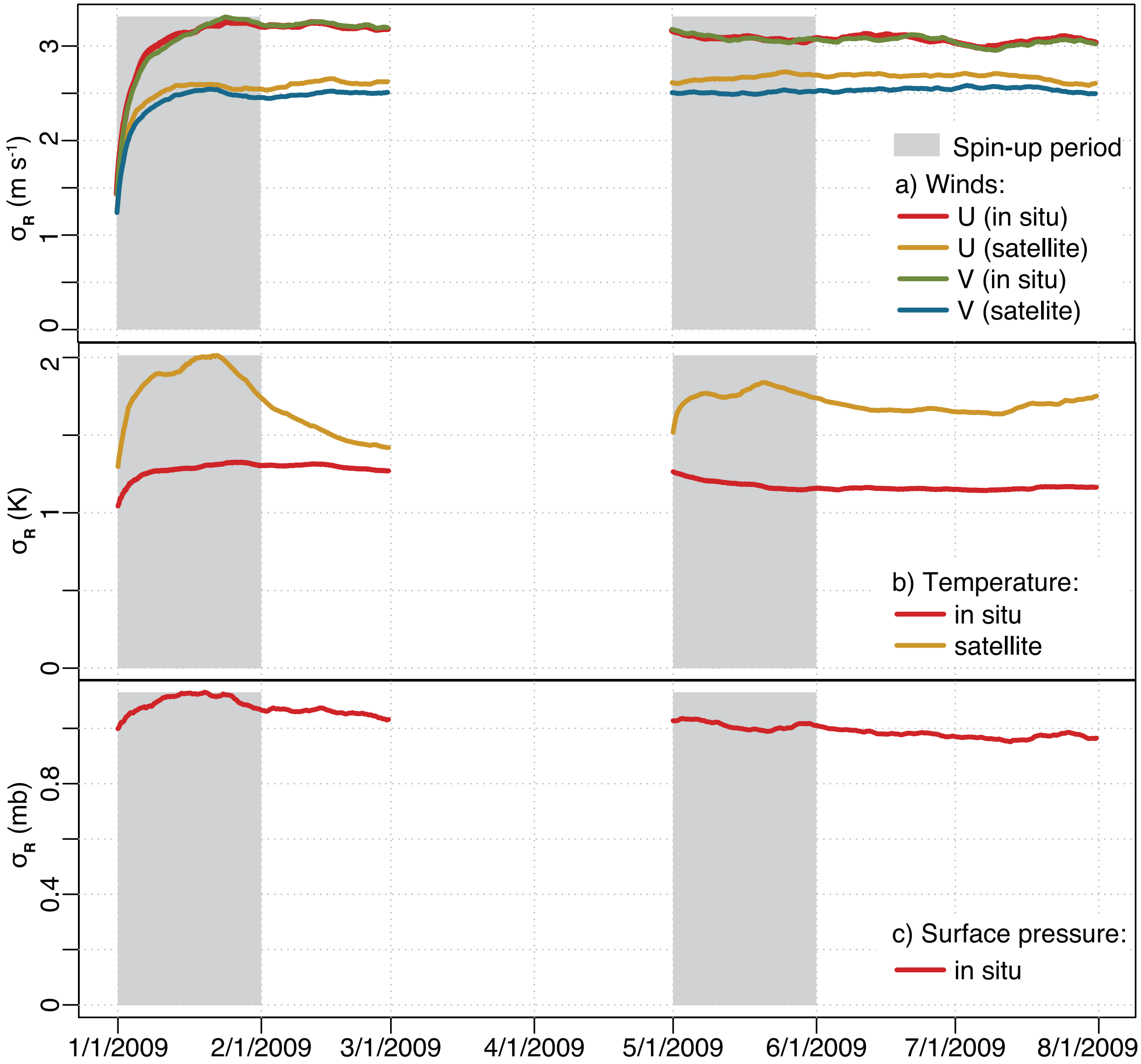


Figure E.4: The square root of the nugget variance (σ_R^2) estimated within CAM–LETKF.

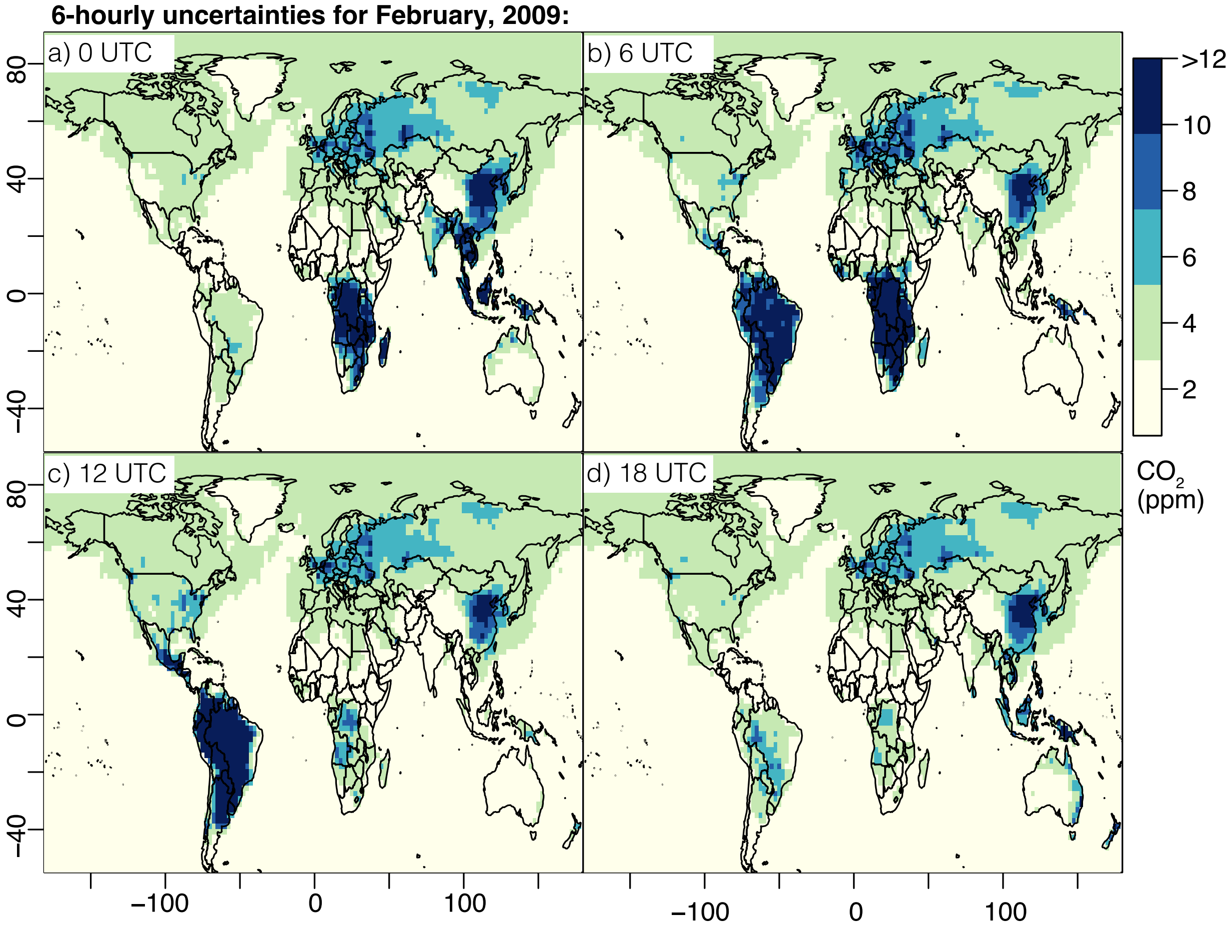


Figure E.5: This figure displays the average 6 hourly CO_2 transport uncertainties in the model surface layer for a) 0 UTC, b) 6 UTC, c) 12 UTC, and d) 18 UTC. This figure is similar to Fig. 7.2a in the main manuscript except the uncertainties (95% confidence interval) shown here are disaggregated by time of day.

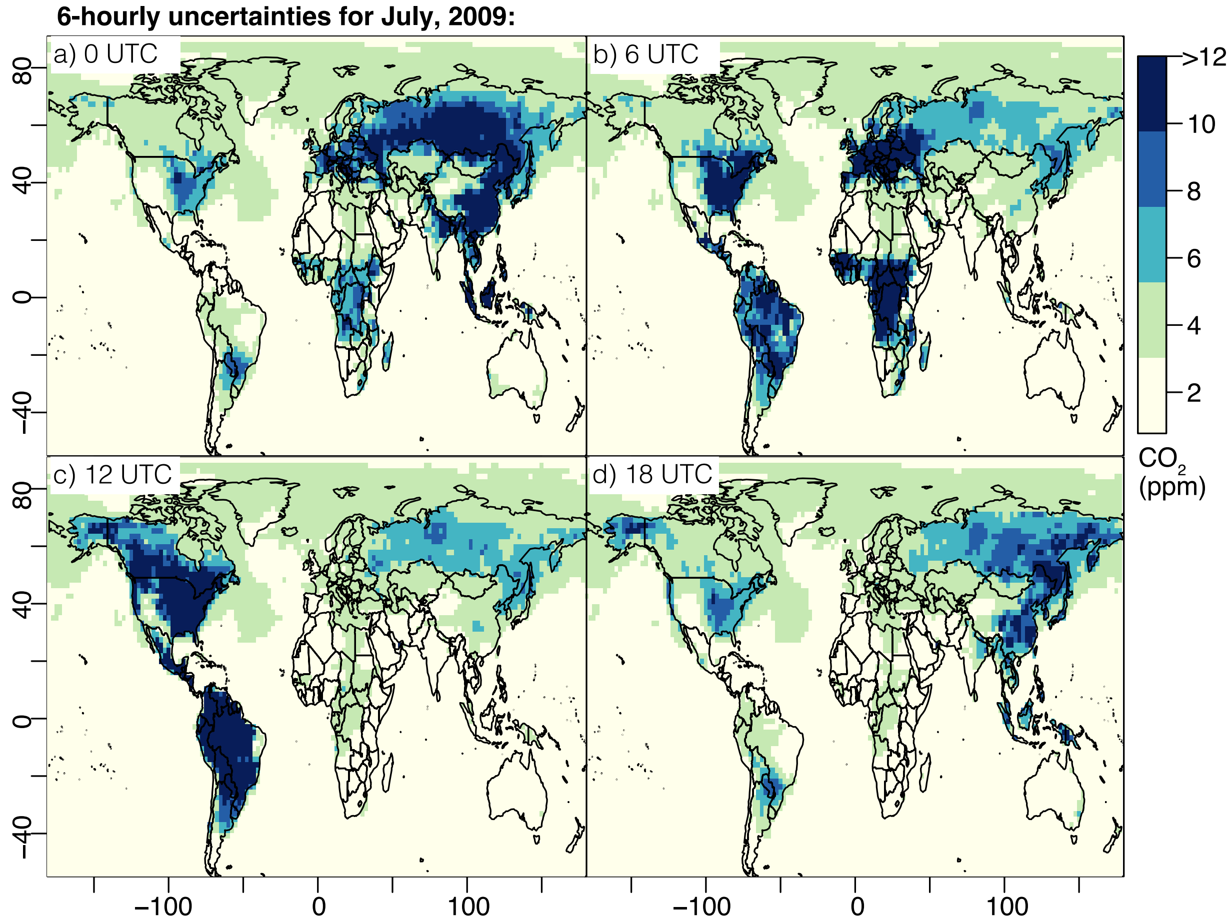


Figure E.6: The CO_2 transport uncertainties for July, 2009, analogous to Fig. E.5 but for a different time period.

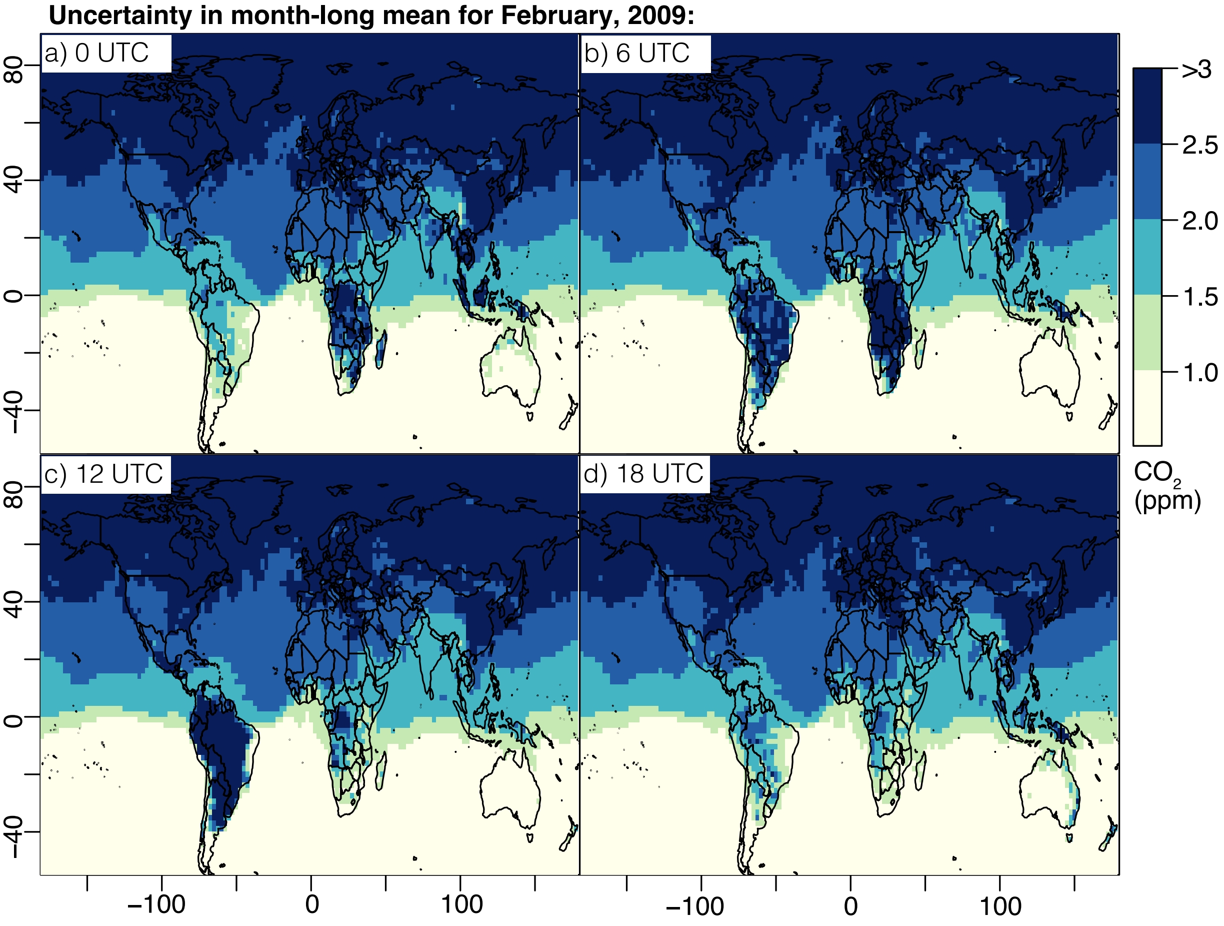


Figure E.7: Uncertainties (95% confidence interval) in the monthly-averaged surface CO_2 concentrations for February, 2009. This figure is similar to Figs. 7.2c in the main article except the uncertainties are broken down by time of day for a) 0 UTC, b) 6 UTC, c) 12 UTC, and d) 18 UTC.

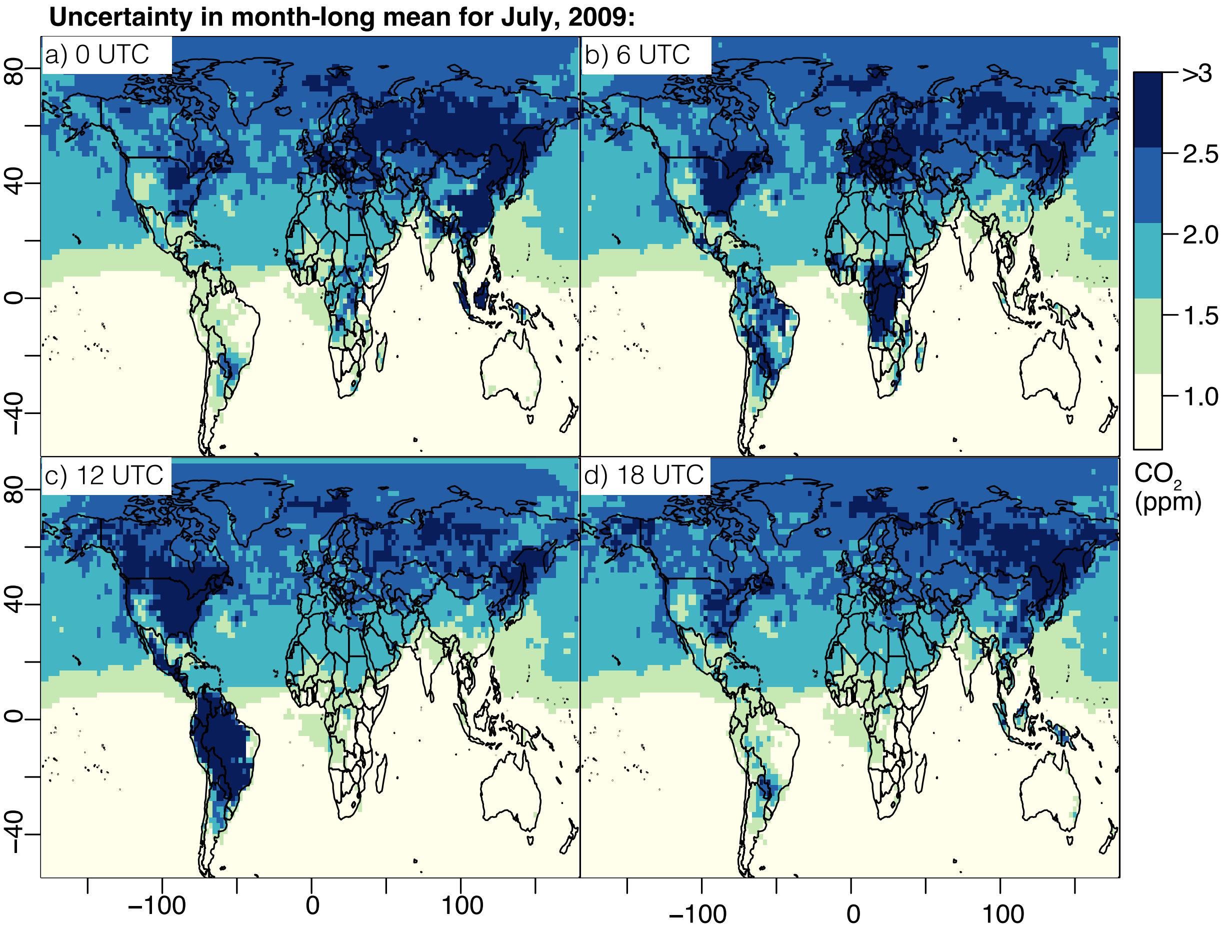


Figure E.8: Uncertainties in the monthly-averaged surface CO_2 concentrations for July, 2009.

These transport uncertainties are in the range of the uncertainties estimated in a number of previous studies. For example, the spatial patterns in the 6 hourly uncertainties are similar to those modeled by Liu et al. (2011) using CAM-LETKF and temperature-scaled CO_2 fluxes from TRANSCOM 3. In addition, a number of previous studies focused on the effects of perturbing individual meteorological parameters at specific observation sites or for individual aircraft campaigns (e.g., Gerbig et al., 2003, 2008; Lin & Gerbig, 2005; Kretschmer et al., 2012). Our 6 hourly transport uncertainties, though very different in both scope and scale, are comparable in magnitude to the individual parameter uncertainties estimated by Gerbig et al. (2003), Gerbig et al. (2008), and Kretschmer et al. (2012) but are less than the uncertainties in Lin & Gerbig (2005). Furthermore, our estimated 6 hourly transport uncertainties also appear similar to or slightly smaller than the model–data mismatch errors estimated at individual observation sites in several inversion studies (e.g., Peters et al., 2007; Schuh et al., 2010; Gourdji et al., 2012). Model–data mismatch includes not only transport errors but also any model or data errors unrelated to an imperfect initial flux estimate. This result may reflect the fact that atmospheric transport often dominates model-data mismatch errors.

E.4 CO_2 observation sites

This section lists the geographic locations of the measurement sites used for case study one in the main article (Figs. 7.4-7.5). Note that this case study does not use actual data from these observation sites, only model output generated for these locations. These observation site locations are available from the GAWSIS Station Information System (MeteoSwiss Federal Office of Meteorology and Climatology, 2014). The sites below are grouped by ecoregion, and the regions used here are defined by Olson et al. (2001).

Table E.1: A list of observation sites used for the analysis in Figs. 7.4-7.5. All measurement sites are towers unless otherwise noted.

Site	Code	Lat. (°N)	Lon. (°)	Alt. (m)
East Asian temperate broadleaf and mixed forests				
Anmyeon-do, Korea	AMY	36.5	126.3	47
Gosan, Korea	GSN	33.3	126.2	72
Kisai, Japan	KIS	36.1	139.6	13
Mikawa-Ichinomiya, Japan	MKW	34.9	137.4	50
Mount Dodaira, Japan	DDR	36.0	139.2	840
Ryori, Japan	RYO	39.0	141.8	260
European temperate broadleaf and mixed forests				
Bialystok, Poland	BIK	53.1	23.0	183
Cesar, Netherlands	CBW	52.0	4.9	-2
Diabla Gora, Poland	DIG	54.2	22.1	157
Gif sur Yvette, France	GIF	48.7	2.1	167
Hegyhatsal, Hungary	HUN	47.0	16.7	248
Heidelberg, Germany	HEI	49.4	8.7	116
Hohenpeissenberg, Germany	HPB	47.8	11.02	985
Kollumerwaard, Netherlands	KMW	53.3	6.3	0
Mace Head, Ireland	MHD	53.3	-9.9	8
Moussala, Bulgaria	BEO	42.2	23.6	2925
Neuglobsow, Germany	NGL	53.2	13.0	65
Norunda, Sweden	NOR	60.0	17.3	70
Ochsenkopf, Germany	OXK	50.0	11.8	1185

Table E.1: (Continued)

Orleans, France	TRN	48.0	2.1	131
Puy de Dome, France	PUY	45.8	3.0	1465
Schauinsland, Germany	SSL	47.92	7.92	1205
Tall Tower Angus, United Kingdom	TTA	56.6	3.0	400
North American temperate broadleaf and mixed forests				
Argyle, Maine, US	AMT	45.0	68.7	50
Beech Island, South Carolina, US	SCT	33.4	81.8	115
Egbert, Ontario, Canada	EGB	44.2	-79.8	253
Park Falls, Wisconsin, US (tower/aircraft)	LEF	46.0	90.3	472
Shenandoah National Park, Virginia	SNP	38.6	78.4	1008
Worchester, Massachusetts, US (aircraft)	NHA	43.0	-70.6	0
North American boreal forests / taiga				
Candle Lake, Sask., Canada	CDL	53.9	-104.7	489
Chibougmau, Quebec, Canada	CHM	49.7	74.3	393
East Trout Lake, Sask., Canada (tower/aircraft)	ETL	54.4	-105.0	492
Fraserdale, Ontario, Canada	FSD	49.9	-81.6	210
Poker Flats, Alaska, US (aircraft)	PFA	65.1	-147.3	210
North American temperate grasslands, savannas and shrublands				
Beaver Crossing, Nebraska (aircraft)	BNE	40.8	-97.3	466
Bondville, Illinois, US (aircraft)	AAO	40.1	-88.4	230
Boulder, Colorado, US	BAO	40.1	-105.0	1584
Briggsdale, Colorado (aircraft)	CAR	40.4	-104.3	1740

Table E.1: (Continued)

Dahlen, North Dakota (aircraft)	DND	47.5	-99.2	472
Lac La Biche, Alberta, Canada	LLB	55.0	-112.5	540
Southern Great Plains, Oklahoma, US (tower/aircraft)	SGP	36.78	-97.5	314
Moody, Texas, US	WKT	31.3	97.3	251
Walnut Grove, California, US	WGC	38.3	121.5	0
West Branch, Iowa, US (tower and aircraft)	WBI	41.7	91.4	242

E.5 CO_2 model-data comparisons

In this portion of the supplement, we show several CO_2 model and data time series from different types of observation sites (Figs. E.9 – E.14). These plots illustrate the capacity of CAM–LETKF (paired with CarbonTracker fluxes) to reproduce hourly-averaged CO_2 observations. Furthermore, the plots provide greater context on the CO_2 ensemble spread. The top panel of each figure illustrates the ensemble mean and ensemble spread. The bottom panel shows the modeled CO_2 boundary layer enhancement – modeled CO_2 at the observation site minus modeled concentrations at 600 hPa. This enhancement approximates the CO_2 contribution rom regional surface fluxes. This increment is used for case study one in the main paper (section 7.2.4). In general, the modeled contribution of regional fluxes is largest during summer where biosphere uptake is strongest (e.g., LEF and AMT).

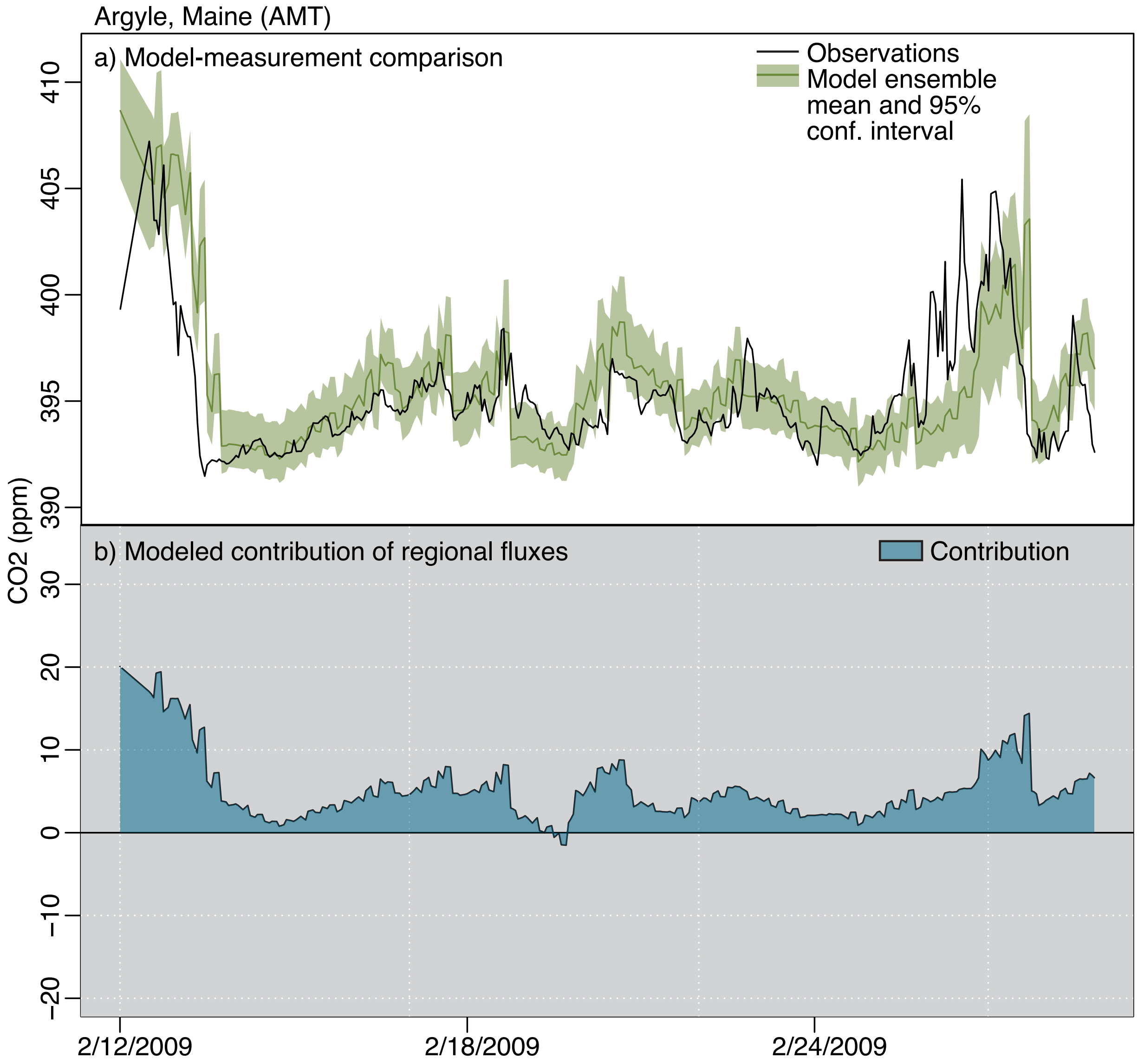


Figure E.9: Panel (a) displays the hourly-averaged CO_2 measurements at Argyle tower, Maine, and the modeled CO_2 time series using CAM–LETKF and CarbonTracker fluxes. Panel (b) shows the estimated contribution of regional CO_2 fluxes at the observation site. Here, we define this contribution as modeled CO_2 at the surface minus modeled CO_2 at 600 hPa.

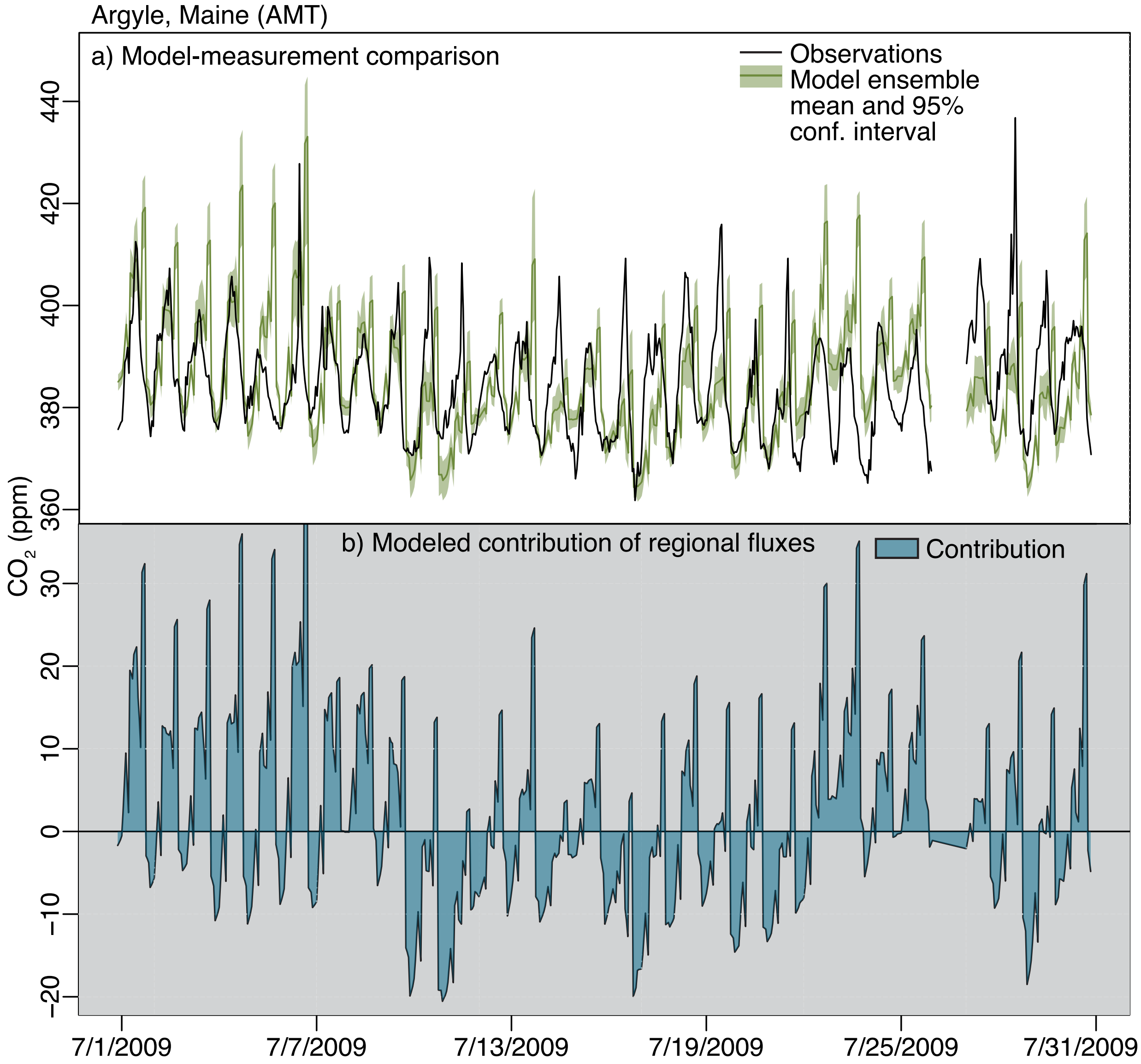


Figure E.10: This figure is analogous to Fig. E.9 but the Argyle tower in July 2009. Note that the top panel of each time-series plot (Figs. E.9a– E.14a) has a different y-axis, but the bottom panels (Figs. E.9b– E.14b) all have the same y–axis.

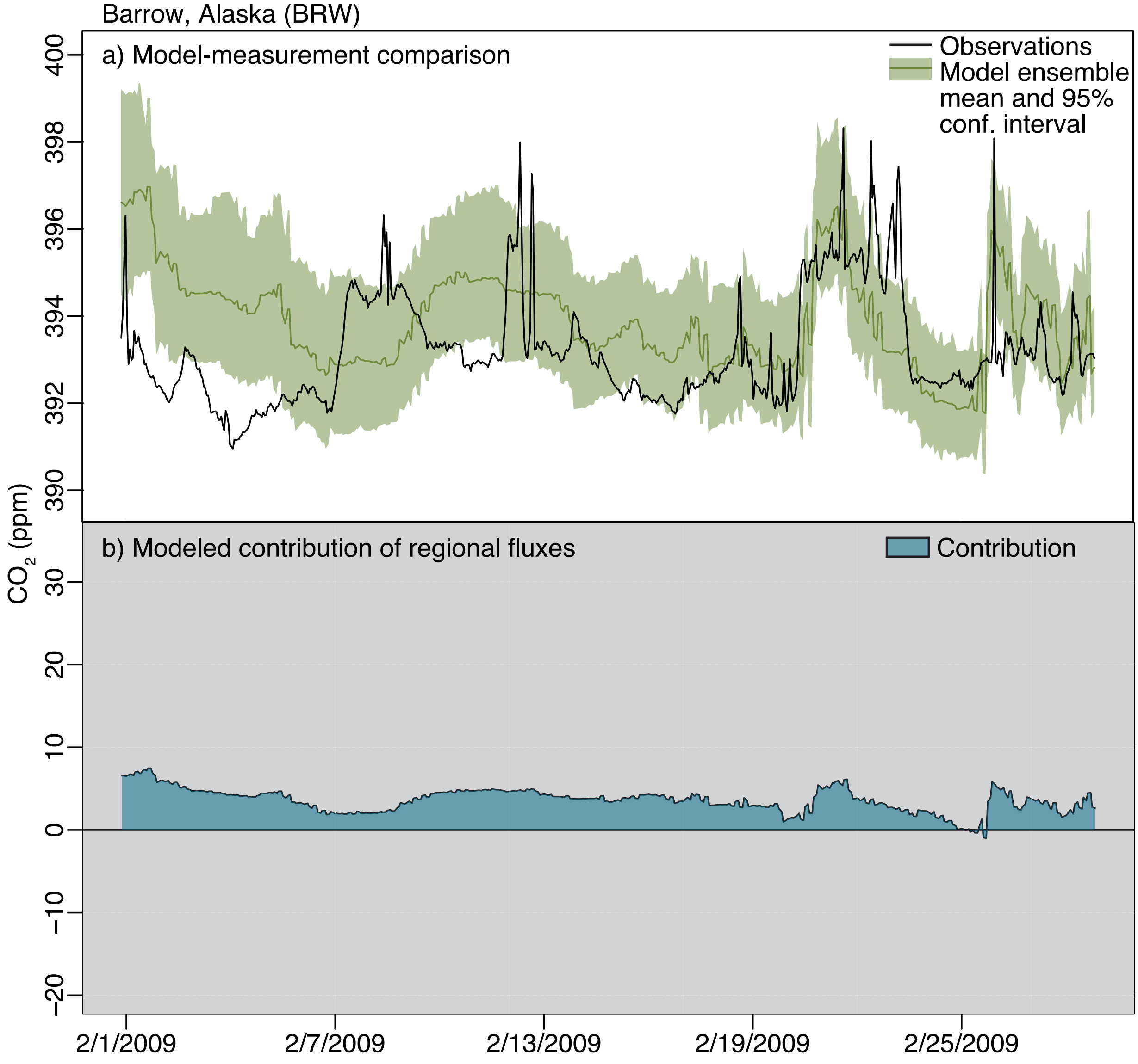


Figure E.11: This figure is analogous to Fig. E.9 but the Barrow, Alaska, in February 2009.

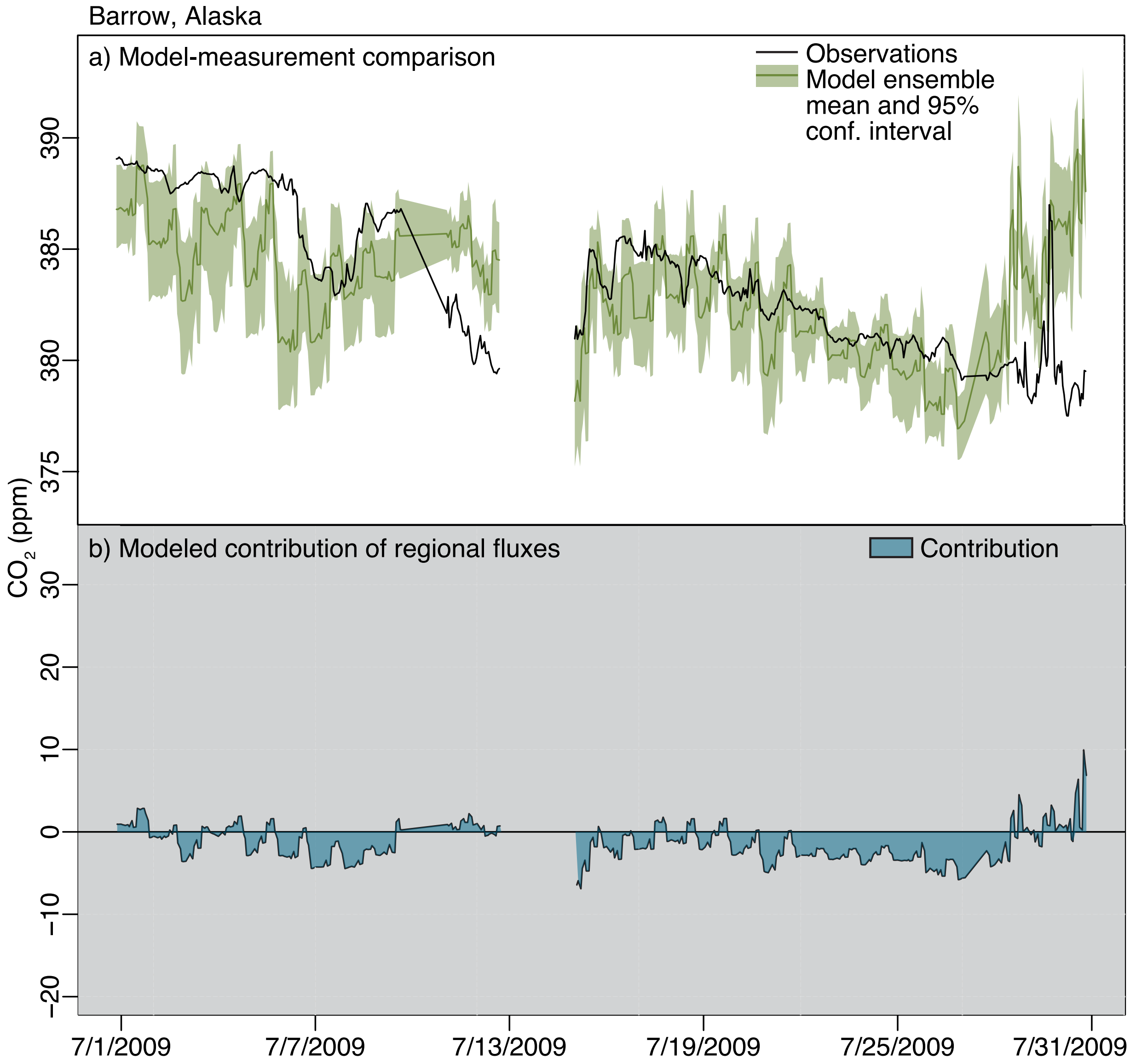


Figure E.12: This figure is analogous to Fig. E.9 but the Barrow, Alaska, in July 2009.

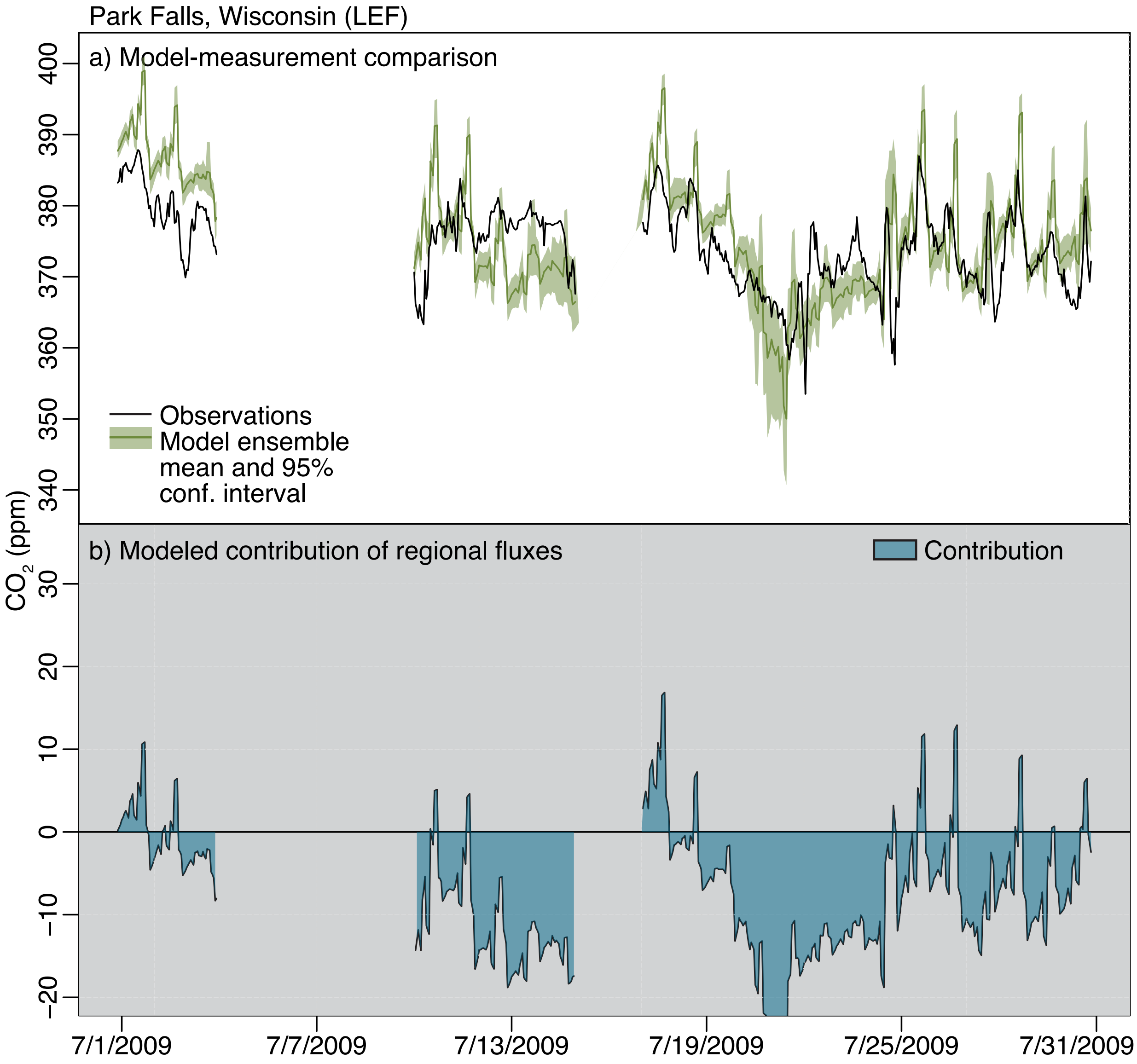


Figure E.13: This figure is analogous to Fig. E.9 but for Park Falls, Wisconsin, in July 2009.

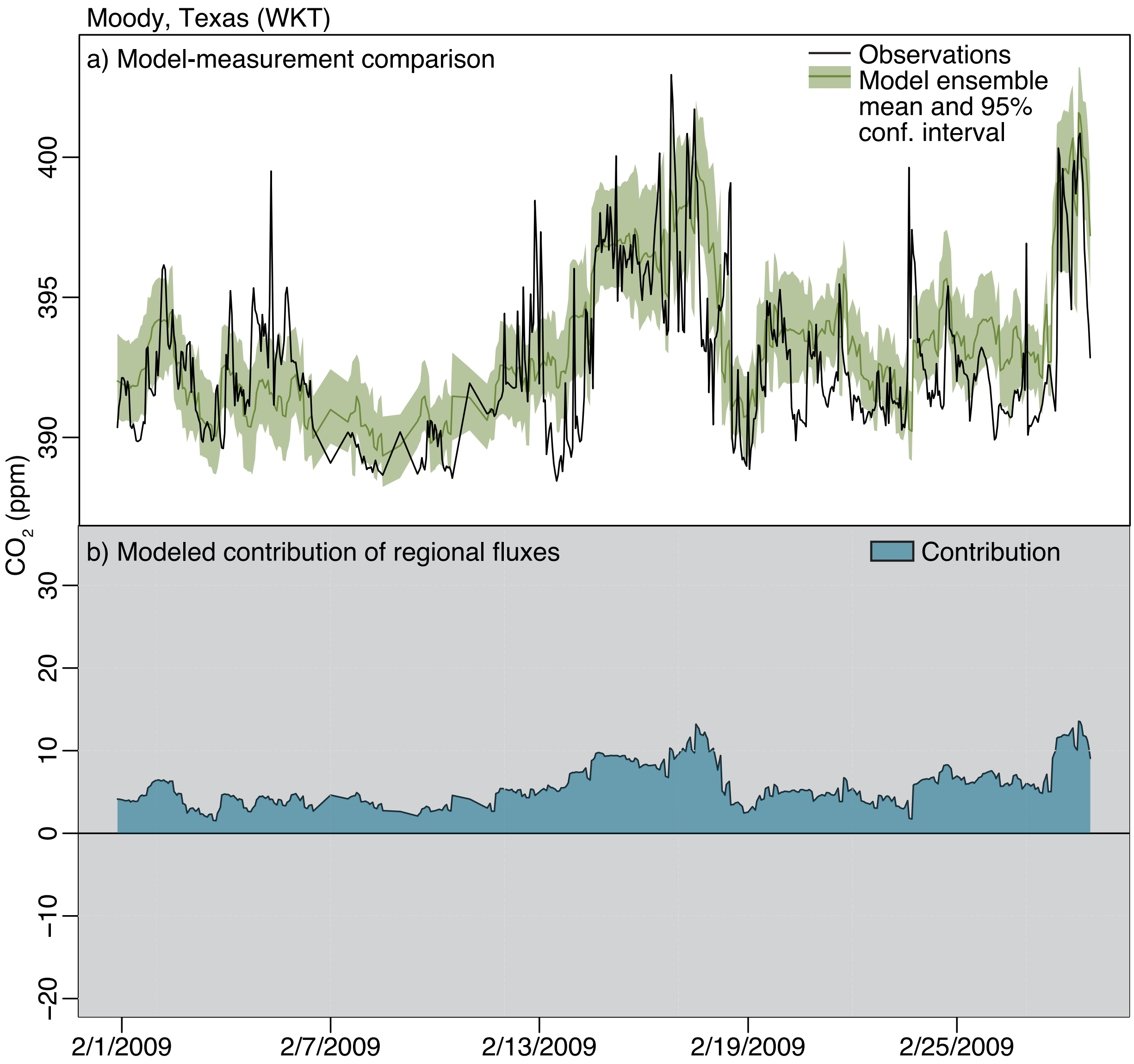


Figure E.14: This figure is analogous to Fig. E.9 but for Moody, Texas, in February 2009.

E.6 Plots of meteorological variables and uncertainties

This section describes, in greater detail, the monthly-averaged meteorological parameters considered in the synthetic tracer experiment (case study two, sections 7.2.5 and 7.3.4). Table S2 lists all of the meteorological parameters that we compare against the synthetic tracer CV. We compare the synthetic tracer against the monthly-averaged meteorological parameters, the standard deviation in the monthly mean parameters, and the CV of each meteorological parameter – 60 parameters in total.

Figures E.15-E.17 display a number of monthly-averaged meteorological parameters estimated by CAM-LETKF – both those listed in section 7.3.4 and several additional variables for reference. For example, these figures display monthly mean zonal and meridional winds, and the uncertainties (standard deviation) (Fig. E.17). These uncertainties exhibit a number of patterns consistent with well-known meso- and synoptic-scale circulation patterns. For example, the uncertainty in zonal winds is generally higher in many coastal regions including the west coast of North and South America. These patterns may reflect uncertainties in modeled sea breezes. Uncertainties in the zonal surface winds are also higher over many mountainous regions, including the US Rocky Mountains and Himalayas. These uncertainties may reflect the challenges of modeling winds over complex terrain. In addition, uncertainties in both zonal and meridional surface winds are higher along the Intertropical Convergence Zone (ITCZ).

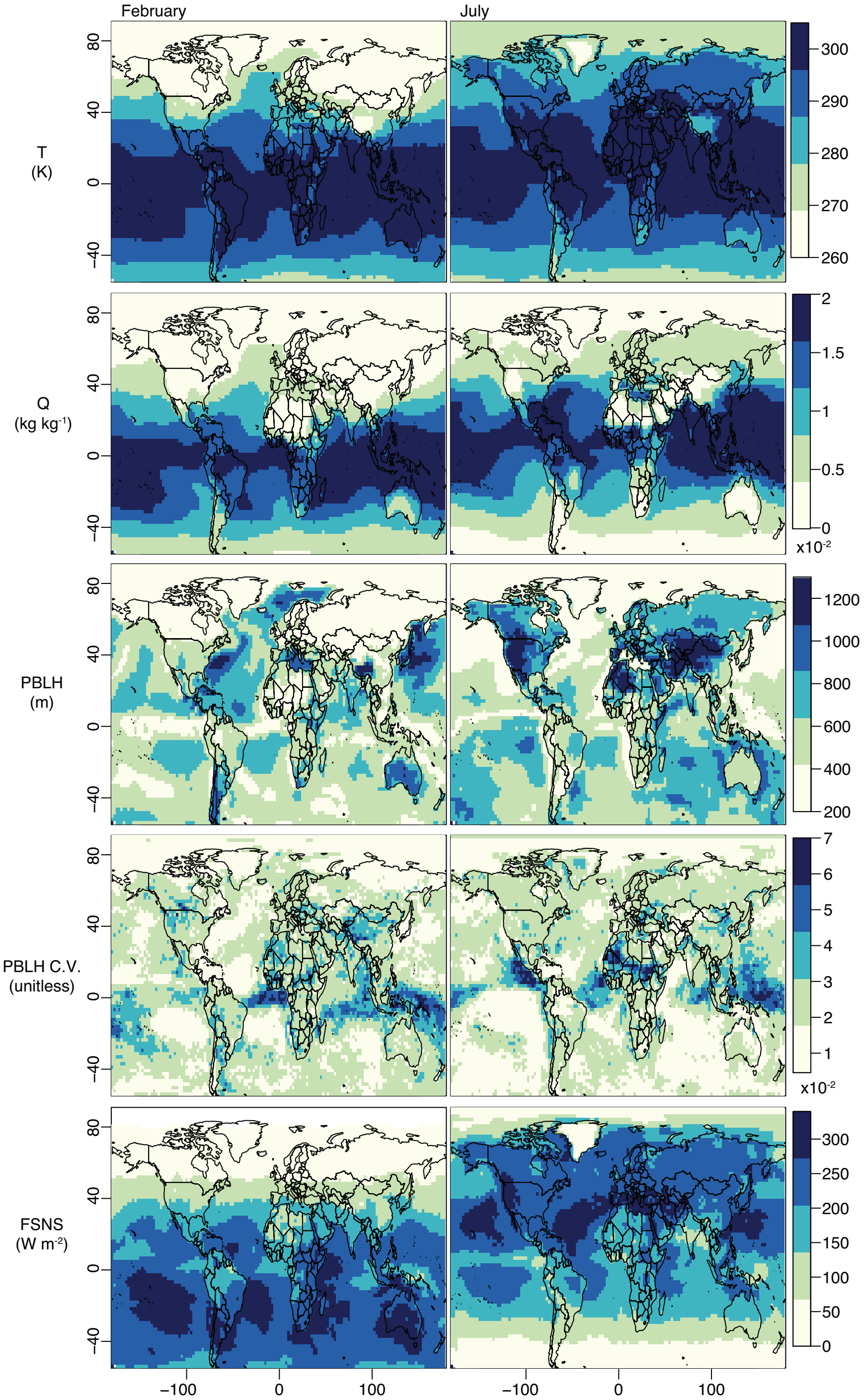


Figure E.15: Maps of monthly-averaged meteorological parameters as estimated by CAM-LETKF.

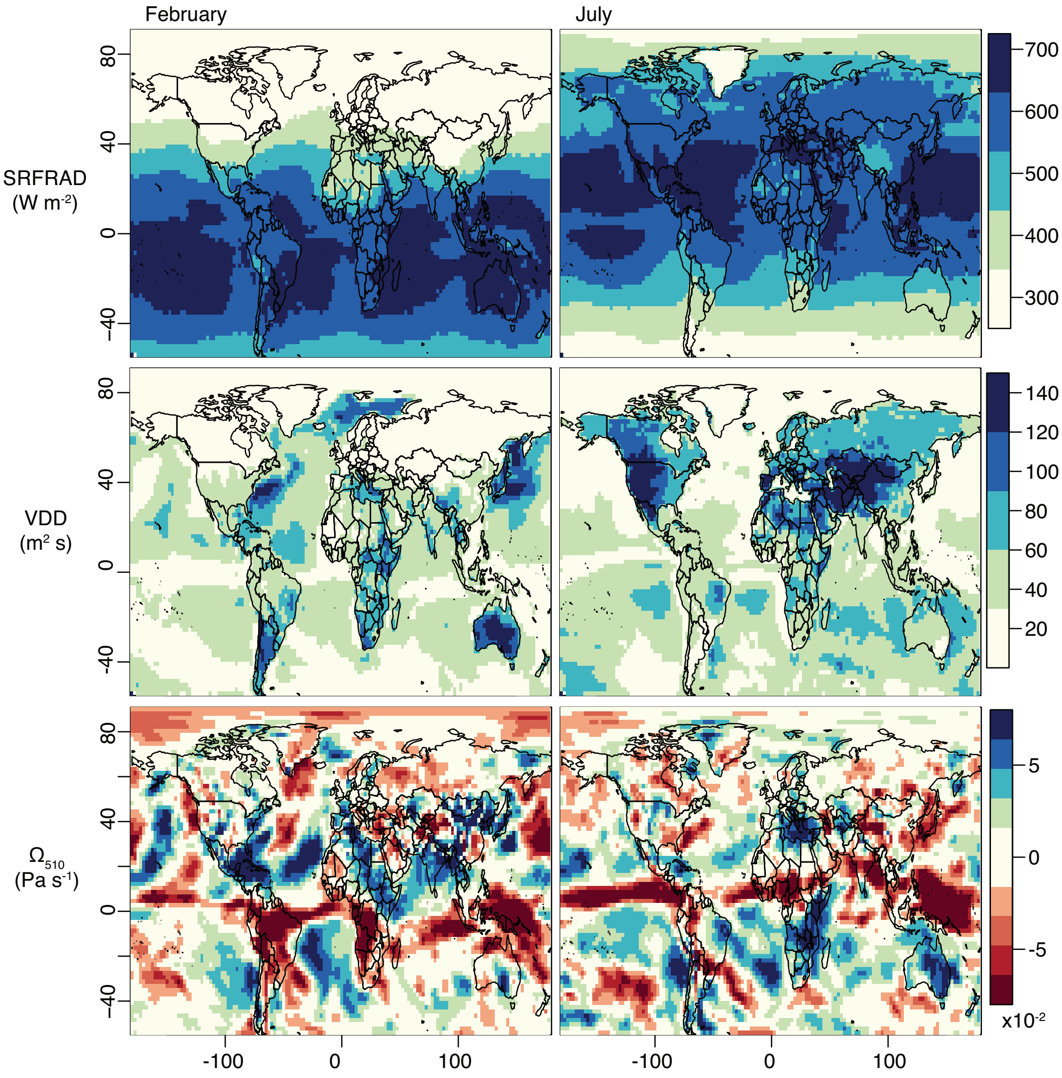


Figure E.16: Maps of monthly-averaged meteorological parameters as estimated by CAM-LETKF.

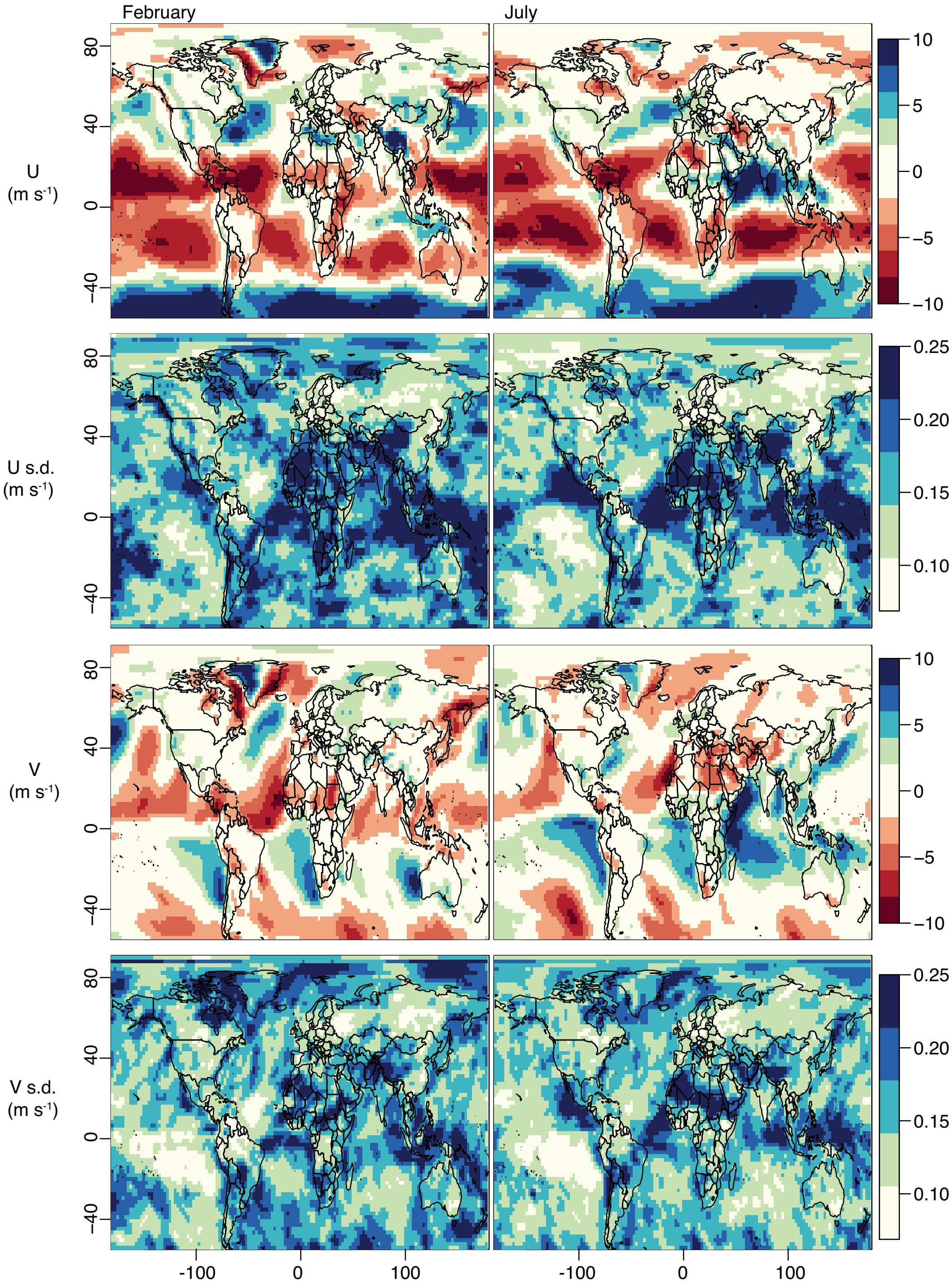


Figure E.17: Maps of monthly-averaged meteorological parameters and uncertainties (standard deviations) as estimated by CAM-LETKF.

Table S2: Candidate meteorological variables

Meteorological variable	Abbreviation	Units
Vertical velocity	Ω	Pa s^{-1}
Vertical velocity at 510hPa	Ω_{510}	Pa s^{-1}
Net longwave flux at the surface	FLNS	W m^{-2}
Downwelling solar flux at surface	FSDS	W m^{-2}
Net solar flux at surface	FSNS	W m^{-2}
Solar flux reflected from surface	FSRS	W m^{-2}
Liquid cloud water	LCWAT	kg kg^{-1}
Surface latent heat flux	LHFLX	W m^{-2}
Planetary boundary layer height	PBLH	m
Large-scale, stable precipitation rate	PRECL	m s^{-1}
Convective precipitation rate	PRECC	m s^{-1}
Specific humidity	Q	kg kg^{-1}
Relative humidity	RELHUM	%
Surface sensible heat flux	SHFLX	W m^{-2}
Net radiative flux at surface	SRFRAD	W m^{-2}
Temperature	T	K
Zonal wind	U	m s^{-1}
Meridional wind	V	m s^{-1}
Vertical diffusion diffusivity	VDD	m^2 s
Total wind velocity	wind	m s^{-1}

References

Advisory group to the Midwestern Greenhouse Gas Reduction Accord (MGGRA) (2009). Final draft recommendations.

Anderson, J. L. (2009). Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A*, 61(1), 72–83.

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., De Wekker, S. F. J., Wolfe, D. E., Munger, J. W., & Tans, P. P. (2014). Co_2, co, and ch_4 measurements from tall towers in the noaa earth system research laboratory's global greenhouse gas reference network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts. *Atmospheric Measurement Techniques*, 7(2), 647–687.

Andrieu, C., de Freitas, N., Doucet, A., & Jordan, M. (2003). An introduction to MCMC for machine learning. *Machine Learning*, 50, 5–43.

Antoniou, A. & Lu, W. (2007). *Practical Optimization: Algorithms and Engineering Applications*. New York, NY: Springer.

Avis, C. A., Weaver, A. J., & Meissner, K. J. (2011). Reduction in areal extent of high-latitude wetlands in response to permafrost thaw. *Nature Geoscience*, 4(7), 444–448.

Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., & Zhu, Z. (2006). Transcom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988–2003. *Global Biogeochemical Cycles*, 20(1), GB1002.

Barnes, R. & You, K. (1992). Adding bounds to kriging. *Mathematical Geosciences*, 24, 171–176.

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Dentener, F., Wagner, T., Platt, U., Kaplan, J. O., Koerner, S., Heimann, M., Dlugokencky, E. J., & Goede, A. (2007). Satellite chartography of atmospheric methane from SCIAMACHY on board ENVISAT: 2. evaluation based on inverse model simulations. *Journal of Geophysical Research – Atmospheres*, 112(D2).

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Villani, M. G., Houweling, S., Dentener, F., Dlugokencky, E. J., Miller, J. B., Gatti, L. V., Engel, A., & Levin, I. (2009). Inverse modeling of global and regional CH_4 emissions using SCIAMACHY satellite retrievals. *Journal of Geophysical Research – Atmospheres*, 114.

Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R. A., Dlugokencky, E., Wofsy, S. C., Kort, E. A., Sweeney, C., Schuck, T., Brenninkmeijer, C., Chen, H., Beck, V., & Gerbig, C. (2013). Atmospheric CH_4 in the first decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements. *Journal of Geophysical Research – Atmospheres*, 118(13), 7350–7369.

Bergamaschi, P., Krol, M., Meirink, J. F., Dentener, F., Segers, A., van Aardenne, J., Monni, S., Vermeulen, A. T., Schmidt, M., Ramonet, M., Yver, C., Meinhardt, F., Nisbet, E. G., Fisher, R. E., O'Doherty, S., & Dlugokencky, E. J. (2010). Inverse modeling of European CH_4 emissions 2001-2006. *Journal of Geophysical Research – Atmospheres*, 115.

Biraud, S. C., Torn, M. S., Smith, J. R., Sweeney, C., Riley, W. J., & Tans, P. P. (2013). A multi-year record of airborne CO_2 observations in the us southern great plains. *Atmospheric Measurement Technology*, 6, 751–763.

Bohn, T. J., Melton, J. R., Ito, A., Kleinen, T., Spahni, R., Stocker, B. D., Zhang, B., Zhu, X., Schroeder, R., Glagolev, M. V., Maksyutov, S., Brovkin, V., Chen, G., Denisov, S. N., Eliseev, A. V., Gallego-Sala, A., McDonald, K. C., Rawlins, M. A., Riley, W. J., Subin, Z. M., Tian, H., Zhuang, Q., & Kaplan, J. O. (2015). WETCHIMP-WSL: intercomparison of wetland methane emissions models over West Siberia. *Biogeosciences Discussions*, 12(2), 1907–1973.

Bolstad, W. (2012). *Understanding Computational Bayesian Statistics*. Wiley Series in Computational Statistics. Hoboken, New Jersey: John Wiley & Sons.

Bousquet, P., Ringeval, B., Pison, I., Dlugokencky, E. J., Brunke, E. G., Carouge, C., Chevallier, F., Fortems-Cheiney, A., Frankenberg, C., Hauglustaine, D. A., Krummel, P. B., Langenfelds, R. L., Ramonet, M., Schmidt, M., Steele, L. P., Szopa, S., Yver, C., Viovy, N., & Ciais, P. (2011). Source attribution of the changes in atmospheric methane for 2006-2008. *Atmospheric Chemistry and Physics*, 11(8), 3689–3700.

Bouwman, A. (1996). Direct emission of nitrous oxide from agricultural soils. *Nutrient Cycling in Agroecosystems*, 46(1), 53–70.

Bouwman, A. F., Vanderhoek, K. W., & Olivier, J. G. J. (1995). Uncertainties in the global source distribution of nitrous-oxide. *Journal of Geophysical Research – Atmospheres*, 100(D2), 2785–2800.

Bridgham, S. D., Cadillo-Quiroz, H., Keller, J. K., & Zhuang, Q. (2013). Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales. *Global Change Biology*, 19(5), 1325–1346.

Bubier, J., Costello, A., Moore, T. R., Roulet, N. T., & Savage, K. (1993). Microtopography and methane flux in boreal peatlands, northern Ontario, Canada. *Canadian Journal of Botany*, 71(8), 1056–1063.

Burrows, S. M., Rayner, P. J., Butler, T., & Lawrence, M. G. (2013). Estimating bacteria emissions from inversion of atmospheric transport: sensitivity to modelled particle characteristics. *Atmospheric Chemistry and Physics*, 13(11), 5473–5488.

Butler, J. (2014). The NOAA annual greenhouse gas index (AGGI).

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

Casella, G. & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3), pp. 167–174.

Chanton, J. P., Glaser, P. H., Chasar, L. S., Burdige, D. J., Hines, M. E., Siegel, D. I., Tremblay, L. B., & Cooper, W. T. (2008). Radiocarbon evidence for the importance of surface vegetation on fermentation and methanogenesis in contrasting types of boreal peatlands. *Global Biogeochemical Cycles*, 22(4).

Cheang, W.-K. & Reinsel, G. C. (2000). Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood. *Journal of the American Statistical Association*, 95(452), 1173–1184.

Chen, H., Zhou, T., Neale, R. B., Wu, X., & Zhang, G. J. (2010). Performance of the new NCAR CAM3.5 in East Asian summer monsoon simulations: Sensitivity to modifications of the convection scheme. *Journal of Climate*, 23(13), 3657–3675.

Chen, Y.-H. & Prinn, R. G. (2006). Estimation of atmospheric methane emissions between 1996 and 2001 using a three-dimensional global chemical transport model. *Journal of Geophysical Research – Atmospheres*, 111(D10).

Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4), pp. 327–335.

Ciais, P., Rayner, P., Chevallier, F., Bousquet, P., Logan, M., Peylin, P., & Ramonet, M. (2011). Atmospheric inversions for estimating CO_2 fluxes: methods and perspectives. In M. Jonas, Z. Nahorski, S. Nilsson, & T. Whiter (Eds.), *Greenhouse Gas Inventories* (pp. 69–92). Springer Netherlands.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., & Canadell, J. (2013). *Carbon and Other Biogeochemical Cycles - Final Draft Underlying Scientific Technical Assessment*, chapter 6. IPCC Secretariat: Geneva.

Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., Chang, P., Doney, S. C., Hack, J. J., Henderson, T. B., Kiehl, J. T., Large, W. G., McKenna, D. S., Santer, B. D., & Smith, R. D. (2006). The community climate system model version 3 (CCSM3). *Journal of Climate*, 19(11), 2122 – 2143.

Comas, X., Slater, L., & Reeve, A. (2005). Geophysical and hydrological evaluation of two bog complexes in a northern peatland: Implications for the distribution of biogenic gases at the basin scale. *Global Biogeochemical Cycles*, 19(4).

Commisssion, C. E. (2006). Inventory of California greenhouse gas emissions and sinks: 1990 to 2004.

Conn, A., Gould, N., & Toint, P. (2000). *Trust-Region Methods*. Mps-Siam Series on Optimization. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Corazza, M., Bergamaschi, P., Vermeulen, A., Aalto, T., Haszpra, L., Meinhardt, F., Doherty, S., & Thompson, R. (2011). Inverse modelling of European N_2O emissions: assimilationg observations from different networks. *Atmospheric Chemistry and Physics*, 11, 2381–2398.

Corbeil, R. R. & Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics*, 18(1), pp. 31–38.

Cotton, W. R., Pielke, R. A., Walko, R. L., Liston, G. E., Tremback, C. J., Jiang, H., McAnelly, R. L., Harrington, J. Y., Nicholls, M. E., Carrio, G. G., & McFadden, J. P. (2003). RAMS 2001: Current status and future directions. *Meteorology and Atmospheric Physics*, 82(1-4), 5–29.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., & Vitart, F. (2011). The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597.

Denman, K., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P., Dickinson, R., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., Dias, P. d. S., Wofsy, S., & Zhang, X. (2007). *Couplings Between Changes in the Climate System and Biogeochemistry*, chapter 7. Cambridge University Press: Cambridge, UK.

Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3385–3396.

Devore, J. (2012). *Probability and Statistics for Engineering and the Sciences*. Boston: Brooks/Cole, Cengage Learning.

Dimitrov, L. I. (2002). Mud volcanoes – the most important pathway for degassing deeply buried sediments. *Earth-Science Reviews*, 59(1 –4), 49 – 76.

Dlugokencky, E. J., Bruhwiler, L., White, J. W. C., Emmons, L. K., Novelli, P. C., Montzka, S. A., Masarie, K. A., Lang, P. M., Crotwell, A. M., Miller, J. B., & Gatti, L. V. (2009). Observational constraints on recent increases in the atmospheric CH_4 burden. *Geophysical Research Letters*, 36.

Dlugokencky, E. J., Nisbet, E. G., Fisher, R., & Lowry, D. (2011). Global atmospheric methane: budget, changes and dangers. *Philosophical Transactions of the Royal Society A*, 369(1943), 2058–2072.

Draxler, R. R. & Hess, G. (1998). An overview of the hysplit_4 modelling system for trajectories. *Australian Meteorological Magazine*, 47(4), 295–308.

Dutaur, L. & Verchot, L. V. (2007). A global inventory of the soil CH_4 sink. *Global Biogeochemical Cycles*, 21(4).

Eckhardt, S., Prata, A. J., Seibert, P., Stebel, K., & Stohl, A. (2008). Estimation of the vertical profile of sulfur dioxide injection into the atmosphere by a volcanic eruption using satellite column measurements and inverse transport modeling. *Atmospheric Chemistry and Physics*, 8(14), 3881–3897.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., & Tarpley, J. D. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research – Atmospheres*, 108(D22).

Enting, I. (2002). *Inverse Problems in Atmospheric Constituent Transport*. Cambridge Atmospheric and Space Science Series. Cambridge University Press.

Environment Canada (2013). National inventory report 1990-2011: Greenhouse gas sources and sinks in Canada - executive summary.

Etiope, G., Lassey, K. R., Klusman, R. W., & Boschi, E. (2008). Reappraisal of the fossil methane budget and related emission from geologic sources. *Geophysical Research Letters*, 35(9).

European Comission, Joint Research Center (JRC)/ Netherlands Environmental Assessment Agency (PBL) (2009). Emission database for global atmospheric research (EDGAR).

European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL) (2010). Emission database for global atmospheric research (EDGAR), release version 4.2.

European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL) (2013). Emission database for global atmospheric research (edgar), release edgarv4.2 ft2010.

Fang, Y., Michalak, A. M., Shiga, Y. P., & Yadav, V. (2014). Using atmospheric observations to evaluate the spatiotemporal variability of CO_2 fluxes simulated by terrestrial biospheric models. *Biogeosciences*, 11(23), 6985–6997.

Fast, J. (2005). Evaluation of the boundary layer characteristics and particulates in Mexico City predicted by WRF. In *WRF/MM5 Users' Workshop*.

Fienen, M., Kitanidis, P., Watson, D., & Jardine, P. (2004). An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory. *Mathematical Geosciences*, 36(1), 101–126.

Fienen, M. N., Luo, J., & Kitanidis, P. K. (2006). A Bayesian geostatistical transfer function approach to tracer test analysis. *Water Resources Research*, 42(7).

Fiore, A. M., Jacob, D. J., Field, B. D., Streets, D. G., Fernandes, S. D., & Jang, C. (2002). Linking ozone pollution and climate change: The case for controlling methane. *Geophysical Research Letters*, 29(19), 1919.

Food and Agriculture Organization of the United Nations, Plant Production Protection Division (2012). Fertistat: Fertilizer use statistics.

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D., Haywood, J., Lean, J., Lowe, D., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., & Dorland, R. V. (2007). *Changes in atmospheric constituents and in radiative forcing*. Cambridge University Press: Cambridge, UK.

Fraser, A., Palmer, P. I., Feng, L., Boesch, H., Cogan, A., Parker, R., Dlugokencky, E. J., Fraser, P. J., Krummel, P. B., Langenfelds, R. L., O'Doherty, S., Prinn, R. G., Steele, L. P., van der Schoot, M., & Weiss, R. F. (2013). Estimating regional methane surface fluxes: the relative importance of surface and GOSAT mole fraction measurements. *Atmospheric Chemistry and Physics*, 13(11), 5697–5713.

Gay, D. (1984). A trust-region approach to linearly constrained optimization. In D. Griffiths (Ed.), *Numerical Analysis*, volume 1066 of *Lecture Notes in Mathematics* (pp. 72–105). Springer Berlin Heidelberg.

Gedney, N., Cox, P., & Huntingford, C. (2004). Climate feedback from wetland methane emissions. *Geophysical Research Letters*, 31(20).

Gelfand, A. E., Smith, A. F. M., & Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, 87(418), pp. 523–532.

Gelman, A. (2004). *Bayesian Data Analysis.* Texts in Statistical Science. Boca Raton, Florida: Chapman & Hall/CRC, 2 edition.

Gerbig, C., Körner, S., & Lin, J. C. (2008). Vertical mixing in atmospheric tracer transport models: error characterization and propagation. *Atmospheric Chemistry and Physics*, 8(3), 591–602.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., & Grainger, C. A. (2003). Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. analysis of cobra data using a receptor-oriented framework. *Journal of Geophysical Research – Atmospheres*, 108(D24).

Geyer, C. (2011). *Handbook of Markov Chain Monte Carlo*, chapter Introduction to Markov Chain Monte Carlo, (pp. 3–48). Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis: Boca Raton; London.

Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization*. London, UK: Academic Press.

Göckede, M., Turner, D. P., Michalak, A. M., Vickers, D., & Law, B. E. (2010). Sensitivity of a subregional scale atmospheric inverse CO_2 modeling framework to boundary conditions. *Journal of Geophysical Research – Atmospheres*, 115.

Gourdji, S. M., Hirsch, A. I., Mueller, K. L., Yadav, V., Andrews, A. E., & Michalak, A. M. (2010). Regional-scale geostatistical inverse modeling of north american CO_2 fluxes: a synthetic data study. *Atmospheric Chemistry and Physics*, 10(13), 6151–6167.

Gourdji, S. M., Mueller, K. L., Schaefer, K., & Michalak, A. M. (2008). Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data. *Journal of Geophysical Research – Atmospheres*, 113(D21).

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., & Michalak, A. M. (2012). North American CO_2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion. *Biogeosciences*, 9(1), 457–475.

Grell, G. A. & Devenyi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, 29(14).

Gurney, K., Law, R., Denning, A., Rayner, P., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fan, S., Fung, I., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., & Yuen, C. (2002). Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models. *Nature*, 415(6872), 626–630.

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

Hendriks, D. M. D., van Huissteden, J., & Dolman, A. J. (2010). Multi-technique assessment of spatial and temporal variability of methane fluxes in a peat meadow. *Agriculture and Forest Meteorology*, 150(6, SI), 757–774.

Henze, D. K., Hakami, A., & Seinfeld, J. H. (2007). Development of the adjoint of GEOS-Chem. *Atmospheric Chemistry and Physics*, 7(9), 2413–2433.

Hirsch, A. I., Michalak, A. M., Bruhwiler, L. M., Peters, W., Dlugokencky, E. J., & Tans, P. P. (2006). Inverse modeling estimates of the global nitrous oxide surface flux from 1998-2001. *Global Biogeochemical Cycles*, 20(1).

Hodson, E. L., Poulter, B., Zimmermann, N. E., Prigent, C., & Kaplan, J. O. (2011). The El Niño–Southern Oscillation and wetland methane interannual variability. *Geophysical Research Letters*, 38(8).

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., & Pfister, L. (1995). Stratosphere-troposphere exchange. *Reviews of Geophysics*, 33(4), 403–439.

Howarth, R. W., Santoro, R., & Ingraffea, A. (2011). Methane and the greenhouse-gas footprint of natural gas from shale formations. *Clim Change*, 106(4), 679–690.

Hristov, A. N., Johnson, K. A., & Kebreab, E. (2014). Livestock methane emissions in the United States. *Proceedings of the National Academy of Sciences*, 111(14), E1320.

Huang, J., Golombek, A., Prinn, R., Weiss, R., Fraser, P., Simmonds, P., Dlugokencky, E. J., Hall, B., Elkins, J., Steele, P., Langenfelds, R., Krummel, P., Dutton, G., & Porter, L. (2008). Estimation of regional emissions of nitrous oxide from 1997 to 2005 using multinetwork measurements, a chemical transport model, and an inverse method. *Journal of Geophysical Research – Atmospheres*, 113(D17).

Hudman, R. C., Murray, L. T., Jacob, D. J., Millet, D. B., Turquety, S., Wu, S., Blake, D. R., Goldstein, A. H., Holloway, J., & Sachse, G. W. (2008). Biogenic versus anthropogenic sources of CO in the United States. *Geophysical Research Letters*, 35(4).

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., & Swanson, D. K. (2013). The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth System Science Data*, 5(1), 3–13.

Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., Sauer, T., Szunyogh, I., Yorke, J. A., & Zimin, A. V. (2004). Four-dimensional ensemble kalman filtering. *Tellus A*, 56(4), 273–277.

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, 230(1-2), 112–126.

Huntzinger, D., Post, W., Wei, Y., Michalak, A., West, T., Jacobson, A., Baker, I., Chen, J., Davis, K., Hayes, D., Hoffman, F., Jain, A., Liu, S., McGuire, A., Neilson, R., Potter, C., Poulter, B., Price, D., Raczka, B., Tian, H., Thornton, P., Tomelleri, E., Viovy, N., Xiao, J., Yuan, W., Zeng, N., Zhao, M., & Cook, R. (2012). North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison. *Ecological Modelling*, 232(0), 144 – 157.

Huntzinger, D. N., Gourdji, S. M., Mueller, K. L., & Michalak, A. M. (2011). The utility of continuous atmospheric measurements for identifying biospheric CO_2 flux variability. *Journal of Geophysical Research*, 116(D6).

Hurst, D. F., Bakwin, P. S., & Elkins, J. W. (1998). Recent trends in the variability of halogenated trace gases over the United States. *Journal of Geophysical Research – Atmospheres*, 103(D19), 25299–25306.

Hurst, D. F., Bakwin, P. S., Myers, R. C., & Elkins, J. W. (1997). Behavior of trace gas mixing ratios on a very tall tower in North Carolina. *Journal of Geophysical Research – Atmospheres*, 102(D7), 8825–8835.

Intergovernmental Panel On Climate Change (2006). 2006 IPCC guidelines for national greenhouse gas inventories.

Jacob, D. (1999). *Introduction to Atmospheric Chemistry*. Princeton, NJ: Princeton University Press.

Jeong, S., Hsu, Y.-K., Andrews, A. E., Bianco, L., Vaca, P., Wilczak, J. M., & Fischer, M. L. (2013). A multitower measurement network estimate of California's methane emissions. *Journal of Geophysical Research – Atmospheres*.

Jeong, S., Zhao, C., Andrews, A., Bianco, L., Wilczak, J., & Fischer, M. (2012). Seasonal variation of CH_4 emissions from central California. *Journal of Geophysical Research*, 117(D11), D11306.

Judd, A. G., Hovland, M., Dimitrov, L. I., García Gil, S., & Jukes, V. (2002). The geological methane budget at continental margins and its influence on climate change. *Geofluids*, 2(2), 109–126.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., & Potter, G. L. (2002). NCEP–DOE AMIP-II Reanalysis (R-2). *Bulletin of the American Meteorological Society*, 83(11), 1631–1643.

Kang, J.-S., Kalnay, E., Miyoshi, T., Liu, J., & Fung, I. (2012). Estimation of surface carbon fluxes with an advanced data assimilation methodology. *Journal of Geophysical Research – Atmospheres*, 117(D24), D24101.

Kaplan, J. (2002). Wetlands at the Last Glacial Maximum: Distribution and methane emissions. *Geophysical Research Letters*, 29(6).

Karion, A., Sweeney, C., Petron, G., Frost, G., Michael Hardesty, R., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. A., Schnell, R., Tans, P., Trainer, M., Zamora, R., & Conley, S. (2013). Methane emissions estimate from airborne measurements over a western United States natural gas field. *Geophysical Research Letters*, 40(16), 4393–4397.

Kass, R. & Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773–795.

Katzenstein, A., Doezema, L., Simpson, I., Balke, D., & Rowland, F. (2003). Extensive regional atmospheric hydrocarbon pollution in the southwestern United States. *Proceedings of the National Academy of Sciences of the United States of America*, 100(21), 11975–11979.

Khvorostyanov, D. V., Ciais, P., Krinner, G., & Zimov, S. A. (2008). Vulnerability of east Siberia's frozen carbon stores to future warming. *Geophysical Research Letters*, 35(10).

Kim, H.-S., Maksyutov, S., Glagolev, M. V., Machida, T., Patra, P. K., Sudo, K., & Inoue, G. (2011). Evaluation of methane emissions from West Siberian wetlands based on inverse modeling. *Environmental Research Letters*, 6(3).

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., Cameron-Smith, P., Castaldi, S., Chevallier, F., Feng, L., Fraser, A., Heimann, M., Hodson, E. L., Houweling, S., Josse, B., Fraser, P. J., Krummel, P. B., Lamarque, J.-F., Langenfelds, R. L., Le Quere, C., Naik, V., O'Doherty, S., Palmer, P. I., Pison, I., Plummer, D., Poulter, B., Prinn, R. G., Rigby, M., Ringeval, B., Santini, M., Schmidt, M., Shindell, D. T., Simpson, I. J., Spahni, R., Steele, L. P., Strode, S. A., Sudo, K., Szopa, S., van der Werf, G. R., Voulgarakis, A., van Weele, M., Weiss, R. F., Williams, J. E., & Zeng, G. (2013). Three decades of global methane sources and sinks. *Nature Geoscience*, 6(10), 813–823.

Kitanidis, P. (1987). Parametric estimation of covariances of regionalized variables. *Journal of the American Water Resources Association*, 23(4), 557–567.

Kitanidis, P. (1995). Quasi-linear geostatistical theory for inversing. *Water Resources Research*, 31(10), 2411–2419.

Kitanidis, P. K. (1983b). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research*, 19(4), 909–921.

Kitanidis, P. K. (1986). Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Research*, 22(4), 499–507.

Kitanidis, P. K. & Lane, R. W. (1985). Maximum likelihood parameter estimation of hydrologic spatial processes by the gauss-newton method. *Journal of Hydrology*, 79(1–2), 53 – 71.

Kitanidis, P. K. & Vomvoris, E. G. (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resources Research*, 19(3), 677–690.

Konishi, S. & Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer Series in Statistics. Springer.

Kopacz, M., Jacob, D. J., Henze, D. K., Heald, C. L., Streets, D. G., & Zhang, Q. (2009). Comparison of adjoint and analytical Bayesian inversion methods for constraining Asian sources of carbon monoxide using satellite (MOPITT) measurements of CO columns. *Journal of Geophysical Research – Atmospheres*, 114(D4).

Koppmann, R. (2008). *Volatile Organic Compounds in the Atmosphere*. Wiley.

Kort, E. A., Daube, B. C., Wofsy, S. C., Andrews, A., Hirsch, A., C., S., Dlugokencky, E. J., Miller, J. B., Tans, P. P., Eluszkiewicz, J., Nehrkorn, T., Michalak, A. M., Stephens, B. B., Gerbig, C., Kaplan, J. O., & Houweling, S. (2010). Atmospheric constraints on 2004 emissions of methane and nitrous oxide in North America from atmospheric measurements and receptor-oriented modeling framework. *Journal of Integrative Environmental Sciences*, 7(2), 125–133.

Kort, E. A., Eluszkiewicz, J., Stephens, B. B., Miller, J. B., Gerbig, C., Nehrkorn, T., Daube, B. C., Kaplan, J. O., Houweling, S., & Wofsy, S. C. (2008). Emissions of CH_4 and N_2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations. *Geophysical Research Letters*, 35(18).

Kort, E. A., Patra, P. K., Ishijima, K., Daube, B. C., Jiménez, R., Elkins, J., Hurst, D., Moore, F. L., Sweeney, C., & Wofsy, S. C. (2011). Tropospheric distribution and variability of N_2O: Evidence for strong tropical emissions. *Geophysical Research Letters*, 38(15).

Koven, C. D., Ringeval, B., Friedlingstein, P., Ciais, P., Cadule, P., Khvorostyanov, D., Krinner, G., & Tarnocai, C. (2011). Permafrost carbon-climate feedbacks accelerate global warming. *Proceedings of the National Academy of Sciences of the United States of America*, 108(36), 14769–14774.

Kretschmer, R., Gerbig, C., Karstens, U., Biavati, G., Vermeulen, A., Vogel, F., Hammer, S., & Totsche, K. U. (2014). Impact of optimized mixing heights on simulated regional atmospheric transport of CO_2. *Atmospheric Chemistry and Physics*, 14(14), 7149–7172.

Kretschmer, R., Gerbig, C., Karstens, U., & Koch, F.-T. (2012). Error characterization of CO_2 vertical mixing in the atmospheric transport model wrf-vprm. *Atmospheric Chemistry and Physics*, 12(5), 2441–2458.

Kvenvolden, K. & Rogers, B. (2005). Gaia's breath – global methane exhalations. *Marine and Petroleum Geology*, 22(4), 579–590. Symposium on Near-Surface Hydrocarbon Migration, Vancouver, CANADA, APR 07-10, 2002.

Lark, R. M. & Cullis, B. R. (2004). Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science*, 55(4), 799–813.

Lark, R. M., Cullis, B. R., & Welham, S. J. (2006). On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science*, 57(6), 787–799.

Lauvaux, T., Pannekoucke, O., Sarrat, C., Chevallier, F., Ciais, P., Noilhan, J., & Rayner, P. J. (2009). Structure of the transport uncertainty in mesoscale inversions of CO_2 sources and sinks using ensemble model simulations. *Biogeosciences*, 6(6), 1089–1102.

Law, R. M., Rayner, P. J., Steele, L. P., & Enting, I. G. (2002). Using high temporal frequency data for CO_2 inversions. *Global Biogeochemical Cycles*, 16(4).

Li, H., Kalnay, E., & Miyoshi, T. (2009). Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(639), 523–533.

Lin, C. & More, J. (1999). Newton's method for large bound-constrained optimization problems. *SIAM Journal on Optimization*, 9(4), 1100–1127.

Lin, J., Gerbig, C., Wofsy, S., Andrews, A., Daube, B., Davis, K., & Grainger, C. (2003). A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model. *Journal of Geophysical Research – Atmospheres*, 108(D16).

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

Lin, J. C., Pejam, M. R., Chan, E., Wofsy, S. C., Gottlieb, E. W., Margolis, H. A., & McCaughey, J. H. (2011). Attributing uncertainties in simulated biospheric carbon fluxes to different error sources. *Global Biogeochemical Cycles*, 25(2), GB2018.

Liu, J., Fung, I., Kalnay, E., & Kang, J.-S. (2011). CO_2 transport uncertainties from the uncertainties in meteorological fields. *Geophysical Research Letters*, 38(12), L12808.

Liu, J., Fung, I., Kalnay, E., Kang, J.-S., Olsen, E. T., & Chen, L. (2012). Simultaneous assimilation of AIRS XCO_2 and meteorological observations in a carbon climate model with an ensemble Kalman filter. *Journal of Geophysical Research – Atmospheres*, 117(D5), D05309.

Liu, J. S., Liang, F., & Wong, W. H. (2000). The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association*, 95(449), 121–134.

Livingston, G. & Hutchinson, G. (2009). *Enclosure-based measurement of trace gas exchange: applications and sources of error*, chapter 2. Methods in ecology. Wiley: London.

Lupascu, M., Wadham, J., Hornibrook, E. R., & Pancost, R. (2012). Temperature sensitivity of methane production in the permafrost active layer at Stordalen, Sweden: a comparison with non-permafrost northern wetlands. *Arctic, Antarctic, and Alpine Research*, 44(4), 469–482.

Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G., & Derwent, R. G. (2011). Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach. *Journal of Geophysical Research – Atmospheres*, 116.

Manning, A. J., Ryall, D. B., Derwent, R. G., Simmonds, P. G., & O'Doherty, S. (2003). Estimating European emissions of ozone-depleting and greenhouse gases using observations and a modeling back-attribution technique. *Journal of Geophysical Research – Atmospheres*, 108(D14).

Masarie, K. A., Pétron, G., Andrews, A., Bruhwiler, L., Conway, T. J., Jacobson, A. R., Miller, J. B., Tans, P. P., Worthy, D. E., & Peters, W. (2011). Impact of CO_2 measurement bias on CarbonTracker surface flux estimates. *Journal of Geophysical Research – Atmospheres*, 116(D17), D17305.

Matross, D. M., Andrews, A., Pathmathevan, M., Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Gottlieb, E. W., Chow, V. Y., Lee, J. T., Zhao, C. L., Bakwin, P. S., Munger, J. W., & Hollinger, D. Y. (2006). Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data. *Tellus Series B-Chemical and Physical Meteorology*, 58(5), 344–358.

McGilchrist, C. (1989). Bias of ML and REML estimators in regression models with ARMA errors. *Journal of Statistical Computation and Simulation*, 32(3), 127–136.

McKain, K., Wofsy, S. C., Nehrkorn, T., Eluszkiewicz, J., Ehleringer, J. R., & Stephens, B. B. (2012). Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22), 8423–8428.

Medvigy, D., Moorcroft, P. R., Avissar, R., & Walko, R. L. (2005). Mass conservation and atmospheric dynamics in the regional atmospheric modeling system (RAMS). *Environmental Fluid Mechanics*, 5(1-2), 109–134.

Meirink, J. F., Bergamaschi, P., Frankenberg, C., d'Amelio, M. T. S., Dlugokencky, E. J., Gatti, L. V., Houweling, S., Miller, J. B., Roeckmann, T., Villani, M. G., & Krol, M. C. (2008). Four-dimensional variational data assimilation for inverse modeling of atmospheric methane emissions: Analysis of SCIAMACHY observations. *Journal of Geophysical Research – Atmospheres*, 113(D17).

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., Eliseev, A. V., Denisov, S. N., Hopcroft, P. O., Lettenmaier, D. P., Riley, W. J., Singarayer, J. S., Subin, Z. M., Tian, H., Zurcher, S., Brovkin, V., van Bodegom, P. M., Kleinen, T., Yu, Z. C., & Kaplan, J. O. (2013a). Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences*, 10(2), 753–788.

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., Eliseev, A. V., Denisov, S. N., Hopcroft, P. O., Lettenmaier, D. P., Riley, W. J., Singarayer, J. S., Subin, Z. M., Tian, H., Zürcher, S., Brovkin, V., van Bodegom, P. M., Kleinen, T., Yu, Z. C., & Kaplan, J. O. (2013b). Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences*, 10(2), 753–788.

Mesinger, F., Dimego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jovi, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., & Shi, W. (2006). North American Regional Reanalysis. *Bull. Am. Meteorol. Soc.*, 87, 343–360.

MeteoSwiss Federal Office of Meteorology and Climatology (2014). GAWSIS station information system.

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

Michalak, A. & Kitanidis, P. (2003). A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification. *Water Resources Research*, 39(2).

Michalak, A. M. (2008). A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling. *Water Resources Research*, 44(9).

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

Michalak, A. M. & Kitanidis, P. K. (2004). Application of geostatistical inverse modeling to contaminant source identification at Dover AFB, Delaware. *Journal of Hydraulic Research*, 42(sup1), 9–18.

Miller, J. B., Lehman, S. J., Montzka, S. A., Sweeney, C., Miller, B. R., Karion, A., Wolak, C., Dlugokencky, E. J., Southon, J., Turnbull, J. C., & Tans, P. P. (2012a). Linking emissions of fossil fuel co2 and other anthropogenic trace gases using atmospheric 14co2. *Journal of Geophysical Research – Atmospheres*, 117(D8).

Miller, R. & Kahn, J. (1962). *Statistical analysis in the geological sciences*. Wiley.

Miller, S. (2007). STILT-CO: A new Lagrangian air pollution model for carbon monoxide. Undergraduate thesis, Harvard University.

Miller, S. M., Hayek, M. N., Andrews, A. E., Fung, I., & Liu, J. (2015). Biases in atmospheric CO_2 estimates from correlated meteorology modeling errors. *Atmospheric Chemistry and Physics*, 15(5), 2903–2914.

Miller, S. M., Kort, E. A., Hirsch, A. I., Dlugokencky, E. J., Andrews, A. E., Xu, X., Tian, H., Nehrkorn, T., Eluszkiewicz, J., Michalak, A. M., & Wofsy, S. C. (2012b). Regional sources of nitrous oxide over the United States: Seasonal variation and spatial distribution. *Journal of Geophysical Research – Atmospheres*, 117(D6), D06310.

Miller, S. M., Matross, D. M., Andrews, A. E., Millet, D. B., Longo, M., Gottlieb, E. W., Hirsch, A. I., Gerbig, C., Lin, J. C., Daube, B. C., Hudman, R. C., Dias, P. L. S., Chow, V. Y., & Wofsy, S. C. (2008). Sources of carbon monoxide and formaldehyde in North America determined from high-resolution atmospheric data. *Atmospheric Chemistry and Physics*, 8(24), 7673–7696.

Miller, S. M., Michalak, A. M., & Levi, P. J. (2014a). Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions. *Geoscientific Model Development*, 7(1), 303–315.

Miller, S. M., Michalak, A. M., & Wofsy, S. C. (2014b). Reply to hristov et al.: Linking methane emissions inventories with atmospheric observations. *Proceedings of the National Academy of Sciences*, 111(14), E1321.

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., Dlugokencky, E. J., Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., Miller, B. R., Miller, J. B., Montzka, S. A., Nehrkorn, T., & Sweeney, C. (2013). Anthropogenic emissions of methane in the United States. *Proceedings of the National Academy of Sciences*, 110(50), 20018–20022.

Miller, S. M., Worthy, D. E. J., Michalak, A. M., Wofsy, S. C., Kort, E. A., Havice, T. C., Andrews, A. E., Dlugokencky, E. J., Kaplan, J. O., Levi, P. J., Tian, H., & Zhang, B. (2014c). Observational constraints on the distribution, seasonality, and environmental predictors of North American boreal methane emissions. *Global Biogeochemical Cycles*, 28(2), 146–160.

Mitchell, L. E., Brook, E. J., Sowers, T., McConnell, J. R., & Taylor, K. (2011). Multidecadal variability of atmospheric methane, 1000-1800 CE. *Journal of Geophysical Research – Biogeosciences*, 116.

Miyoshi, T. (2011). The Gaussian approach to adaptive covariance inflation and its implementation with the Local Ensemble Transform Kalman Filter. *Monthly Weather Review*, 139(5), 1519–1535.

Miyoshi, T. & Kunii, M. (2012). The Local Ensemble Transform Kalman Filter with the Weather Research and Forecasting Model: Experiments with real observations. *Pure and Applied Geophysics*, 169(3), 321–333.

Monteil, G., Houweling, S., Dlugockenky, E. J., Maenhout, G., Vaughn, B. H., White, J. W. C., & Rockmann, T. (2011). Interpreting methane variations in the past two decades using measurements of CH_4 mixing ratio and isotopic composition. *Atmospheric Chemistry and Physics*, 11(17), 9141–9153.

Montzka, S. (2003). *Controlled substances and other source gases*, (pp. 1.1–1.83). World Meteorological Organization: Geneva.

More, J. (1988). Trust regions and projected gradients. *Lecture Notes in Control and Information Sciences*, 113, 1–13.

Mueller, K. L., Gourdji, S. M., & Michalak, A. M. (2008). Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 1. results using atmospheric measurements. *Journal of Geophysical Research – Atmospheres*, 113(D21).

Mueller, K. L., Yadav, V., Curtis, P. S., Vogel, C., & Michalak, A. M. (2010). Attributing the variability of eddy-covariance CO_2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest. *Global Biogeochemical Cycles*, 24(3).

Muller, J. & Stavrakou, T. (2005). Inversion of CO and NO(x) emissions using the adjoint of the IMAGES model. *Atmospheric Chemistry and Physics*, 5, 1157–1186.

National Research Council (2010). *Verifying Greenhouse Gas Emissions:Methods to Support International Climate Agreements*. The National Academies Press.

NEG/ECP, T. (2001). Climate action plan 2001.

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

Nevison, C. & Holland, E. (1997). A reexamination of the impact of anthropogenically fixed nitrogen on atmospheric N_2O and the stratospheric O_3 layer. *Journal of Geophysical Research – Atmospheres*, 102(D21), 25519–25536.

Nevison, C. D., Lueker, T. J., & Weiss, R. F. (2004). Quantifying the nitrous oxide source from coastal upwelling. *Global Biogeochemical Cycles*, 18(1).

Newman, S., Jeong, S., Fischer, M. L., Xu, X., Haman, C. L., Lefer, B., Alvarez, S., Rappenglueck, B., Kort, E. A., Andrews, A. E., Peischl, J., Gurney, K. R., Miller, C. E., & Yung, Y. L. (2013). Diurnal tracking of anthropogenic CO_2 emissions in the Los Angeles basin megacity during spring 2010. *Atmospheric Chemistry and Physics*, 13(8), 4359–4372.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research – Atmospheres*, 116(D12).

NOAA ESRL (2013). Carbon cycle greenhouse gas group aircraft program.

O'Connor, F. M., Boucher, O., Gedney, N., Jones, C. D., Folberth, G. A., Coppell, R., Friedlingstein, P., Collins, W. J., Chappellaz, J., Ridley, J., & Johnson, C. E. (2010). Possible role of wetlands, permafrost, and methane hydrates in the methane cycle under future climate change: A review. *Reviews of Geophysics*, 48.

Olefeldt, D., Turetsky, M. R., Crill, P. M., & McGuire, A. D. (2013). Environmental and physical controls on northern terrestrial methane emissions across permafrost zones. *Global Change Biology*, 19(2), 589–603.

Oleson, K. W., Niu, G.-Y., Yang, Z.-L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., Stöckli, R., Dickinson, R. E., Bonan, G. B., Levis, S., Dai, A., & Qian, T. (2008). Improvements to the community land model and their impact on the hydrological cycle. *Journal of Geophysical Research – Biogeosciences*, 113(G1), G01021.

Olivier, J. & Janssens-Maenhout, G. (2012). *CO_2 emissions from fuel combustion, 2012 edition*, chapter III: Greenhouse-gas emissions. International Energy Agency: Paris.

Olivier, J. G. J. & Peters, J. (2005). CO_2 from non-energy use of fuels: A global, regional and national perspective based on the IPCC tier 1 approach. *Resources Conservation and Recycling*, 45(3), 210–225.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 51(11), 933–938.

Oppenheim, A. & Schafer, R. (1989). *Discrete-time Signal Processing*. Prentice-Hall signal processing series. Prentice Hall.

Pan, L. L., Bowman, K. P., Atlas, E. L., Wofsy, S. C., Zhang, F., Bresch, J. F., Ridley, B. A., Pittman, J. V., Homeyer, C. R., Romashkin, P., & Cooper, W. A. (2010). The Stratosphere-Troposphere Analyses of Regional Transport 2008 Experiment. *Bulletin of the American Meteorological Society*, 91, 327–342.

Papa, F., Prigent, C., Aires, F., Jimenez, C., Rossow, W. B., & Matthews, E. (2010). Interannual variability of surface water extent at the global scale, 1993–2004. *Journal of Geophysical Research – Atmospheres*, 115(D12).

Parazoo, N. C., Denning, A. S., Kawa, S. R., Pawson, S., & Lokupitiya, R. (2012). CO_2 flux estimation errors associated with moist atmospheric processes. *Atmospheric Chemistry and Physics*, 12(14), 6405–6416.

Peischl, J., Ryerson, T. B., Holloway, J. S., Trainer, M., Andrews, A. E., Atlas, E. L., Blake, D. R., Daube, B. C., Dlugokencky, E. J., Fischer, M. L., Goldstein, A. H., Guha, A., Karl, T., Kofler, J., Kosciuch, E., Misztal, P. K., Perring, A. E., Pollack, I. B., Santoni, G. W., Schwarz, J. P., Spackman, J. R., Wofsy, S. C., & Parrish, D. D. (2012). Airborne observations of methane emissions from rice cultivation in the Sacramento Valley of California. *Journal of Geophysical Research – Atmospheres*, 117(D24), n/a–n/a.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., & Tans, P. P. (2007). An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proceedings of the National Academy of Sciences of the United States of America*, 104(48), 18925–18930.

Petrescu, A. M. R., van Beek, L. P. H., van Huissteden, J., Prigent, C., Sachs, T., Corradi, C. A. R., Parmentier, F. J. W., & Dolman, A. J. (2010). Modeling regional to global CH_4 emissions of boreal and arctic wetlands. *Global Biogeochemical Cycles*, 24(4).

Petron, G., Frost, G., Miller, B. R., Hirsch, A. I., Montzka, S. A., Karion, A., Trainer, M., Sweeney, C., Andrews, A. E., Miller, L., Kofler, J., Bar-Ilan, A., Dlugokencky, E. J., Patrick, L., Moore, Jr., C. T., Ryerson, T. B., Siso, C., Kolodzey, W., Lang, P. M., Conway, T., Novelli, P., Masarie, K., Hall, B., Guenther, D., Kitzis, D., Miller, J., Welsh, D., Wolfe, D., Neff, W., & Tans, P. (2012). Hydrocarbon emissions characterization in the Colorado Front Range: A pilot study. *Journal of Geophysical Research – Atmospheres*, 117.

Pickett-Heaps, C. A., Jacob, D. J., Wecht, K. J., Kort, E. A., Wofsy, S. C., Diskin, G. S., Worthy, D. E. J., Kaplan, J. O., Bey, I., & Drevet, J. (2011). Magnitude and seasonality of wetland methane emissions from the Hudson Bay Lowlands (Canada). *Atmospheric Chemistry and Physics*, 11(8), 3773–3779.

Pielke, R. A., Cotton, W. R., Walko, R. L., Tremback, C. J., Lyons, W. A., Grasso, L. D., Nicholls, M. E., Moran, M. D., Wesley, D. A., Lee, T. J., & Copeland, J. H. (1992). A comprehensive meteorological modeling system - RAMS. *Meteorology and Atmospheric Physics*, 49(1-4), 69–91.

Pillai, D., Gerbig, C., Ahmadov, R., Rödenbeck, C., Kretschmer, R., Koch, T., Thompson, R., Neininger, B., & Lavrié, J. V. (2011). High-resolution simulations of atmospheric CO_2 over complex terrain representing the Ochsenkopf mountain tall tower. *Atmospheric Chemistry and Physics*, 11(15), 7445–7464.

Pillai, D., Gerbig, C., Kretschmer, R., Beck, V., Karstens, U., Neininger, B., & Heimann, M. (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.

Pino, D., Vilà-Guerau de Arellano, J., Peters, W., Schröter, J., van Heerwaarden, C. C., & Krol, M. C. (2012). A conceptual framework to quantify the influence of convective boundary layer development on carbon dioxide mixing ratios. *Atmospheric Chemistry and Physics*, 12(6), 2969–2985.

Potter, P., Ramankutty, N., Bennett, E. M., & Donner, S. D. (2010). Characterizing the spatial patterns of global fertilizer application and manure production. *Earth Interactions*, 14.

Prater, J. L., Chanton, J. P., & Whiting, G. J. (2007). Variation in methane production pathways associated with permafrost decomposition in collapse scar bogs of Alberta, Canada. *Global Biogeochemical Cycles*, 21(4).

Prather, M. J., Holmes, C. D., & Hsu, J. (2012). Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry. *Geophysical Research Letters*, 39(9).

Prigent, C., Papa, F., Aires, F., Rossow, W. B., & Matthews, E. (2007). Global inundation dynamics inferred from multiple satellite observations, 1993–2000. *Journal of Geophysical Research – Atmospheres*, 112(D12).

Prinn, R., Cunnold, D., Rasmussen, R., Simmonds, P., Alyea, F., Crawford, A., Fraser, P., & Rosen, R. (1990). Atmospheric emissions and trends of nitrous-oxide deduced from 10 years of ale-gauge data. *Journal of Geophysical Research – Atmospheres*, 95(D11), 18369–18385.

Raczka, B. M., Davis, K. J., Huntzinger, D., Neilson, R. P., Poulter, B., Richardson, A. D., Xiao, J., Baker, I., Ciais, P., Keenan, T. F., Law, B., Post, W. M., Ricciuto, D., Schaefer, K., Tian, H., Tomelleri, E., Verbeeck, H., & Viovy, N. (2013). Evaluation of continental carbon cycle simulations with North American flux tower observations. *Ecological Monographs*, 83(4), 531–556.

Rahn, T. & Wahlen, M. (2000). A reassessment of the global isotopic budget of atmospheric nitrous oxide. *Global Biogeochemical Cycles*, 14(2), 537–543.

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2008). Farming the planet: 1. geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, 22(1).

Ramsey, F. & Schafer, D. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis*. Boston: Cengage Learning, 3 edition.

Ravishankara, A. R., Daniel, J. S., & Portmann, R. W. (2009). Nitrous oxide (N_2O): The dominant ozone-depleting substance emitted in the 21st century. *Science*, 326(5949), 123–125.

Rigby, M., Manning, A. J., & Prinn, R. G. (2011). Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models. *Atmospheric Chemistry and Physics*, 11(18), 9887–9898.

Riley, W. J., Subin, Z. M., Lawrence, D. M., Swenson, S. C., Torn, M. S., Meng, L., Mahowald, N. M., & Hess, P. (2011). Barriers to predicting changes in global terrestrial methane fluxes: analyses using CLM4Me, a methane biogeochemistry model integrated in CESM. *Biogeosciences*, 8(7), 1925–1953.

Ringeval, B., de Noblet-Ducoudré, N., Ciais, P., Bousquet, P., Prigent, C., Papa, F., & Rossow, W. B. (2010). An attempt to quantify the impact of changes in wetland extent on methane emissions on the seasonal and interannual time scales. *Global Biogeochemical Cycles*, 24(2).

Robinson, D. L. (1987). Estimation and use of variance components. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 36(1), pp. 3–14.

Rockmann, T., Kaiser, J., & Brenninkmeijer, C. A. M. (2003). The isotopic fingerprint of the preindustrial and the anthropogenic N_2O source. *Atmospheric Chemistry and Physics*, 3(2), 315–323.

Rodgers, C. (2000). *Inverse methods for atmospheric sounding: Theory and practice*. Series on Atmospheric, Ocean, and Planetary Physics. Singapore: World Scientific.

Roulet, N. T., Ash, R., & Moore, T. R. (1992). Low boreal wetlands as a source of atmospheric methane. *Journal of Geophysical Research*, 97(D4), 3739–3749.

Sanchez-Ccoyllo, O. R., Silva Dias, P. L., Andrade, M. D., & Freitas, S. R. (2006). Determination of O_3-, CO- and PM10-transport in the metropolitan area of Sao Paulo, Brazil through synoptic-scale analysis of back trajectories. *Meteorology and Atmospheric Physics*, 92(1-2), 83–93.

Savitzky, A. & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639.

Schuh, A. E., Denning, A. S., Corbin, K. D., Baker, I. T., Uliasz, M., Parazoo, N., Andrews, A. E., & Worthy, D. E. J. (2010). A regional high-resolution carbon flux inversion of North America for 2004. *Biogeosciences*, 7(5), 1625–1644.

Schuur, E., Abbott, B., Bowden, W., Brovkin, V., Camill, P., Canadell, J., Chanton, J., Chapin, F.S., I., Christensen, T., Ciais, P., Crosby, B., Czimczik, C., Grosse, G., Harden, J., Hayes, D., Hugelius, G., Jastrow, J., Jones, J., Kleinen, T., Koven, C., Krinner, G., Kuhry, P., Lawrence, D., McGuire, A., Natali, S., O'Donnell, J., Ping, C., Riley, W., Rinke, A., Romanovsky, V., Sannel, A., Schädel, C., Schaefer, K., Sky, J., Subin, Z., Tarnocai, C., Turetsky, M., Waldrop, M., Walter Anthony, K., Wickland, K., Wilson, C., & Zimov, S. (2013). Expert assessment of vulnerability of permafrost carbon to climate change. *Climatic Change*, 119(2), 359–374.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.

Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L., & Yadav, V. (2014). Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO_2 measurements. *Geophysical Research Letters*, 41(12), 4381–4388.

Shiga, Y. P., Michalak, A. M., Randolph Kawa, S., & Engelen, R. J. (2013). In-situ CO_2 monitoring network evaluation and design: A criterion based on atmospheric CO_2 variability. *Journal of Geophysical Research – Atmospheres*, 118(4), 2007–2018.

Singarayer, J. S., Valdes, P. J., Friedlingstein, P., Nelson, S., & Beerling, D. J. (2011). Late holocene methane rise caused by orbitally controlled increase in tropical sources. *Nature*, 470(7332), 82–85.

Sitch, S., Smith, B., Prentice, I., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M., Thonicke, K., & Venevsky, S. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, 9(2), 161–185.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Wang, W., & Powers, J. (2005). A description of the advanced research WRF version 2.

Snodgrass, M. & Kitanidis, P. (1997). A geostatistical approach to contaminant source identification. *Water Resources Research*, 33(4), 537–546.

Snyman, J. (2005). *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. Applied Optimization. Boston, MA: Springer.

Sorensen, D. C. (1982). Newton's method with a model trust region modification. *SIAM Journal on Numerical Analysis*, 19(2), pp. 409–426.

Spahni, R., Wania, R., Neef, L., van Weele, M., Pison, I., Bousquet, P., Frankenberg, C., Foster, P. N., Joos, F., Prentice, I. C., & van Velthoven, P. (2011). Constraining global methane emissions and uptake by ecosystems. *Biogeosciences*, 8(6), 1643–1665.

Sprenger, M. & Wernli, H. (2003). A northern hemispheric climatology of cross-tropopause exchange for the ERA15 time period (1979-1993). *Journal of Geophysical Research – Atmospheres*, 108(D12).

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., Aoki, S., Machida, T., Inoue, G., Vinnichenko, N., Lloyd, J., Jordan, A., Heimann, M., Shibistova, O., Langenfelds, R. L., Steele, L. P., Francey, R. J., & Denning, A. S. (2007). Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO_2. *Science*, 316(5832), 1732–1735.

Stohl, A., Seibert, P., Wotawa, G., Arnold, D., Burkhart, J. F., Eckhardt, S., Tapia, C., Vargas, A., & Yasunari, T. J. (2012). Xenon-133 and caesium-137 releases into the atmosphere from the Fukushima Dai-ichi nuclear power plant: determination of the source term, atmospheric dispersion, and deposition. *Atmospheric Chemistry and Physics*, 12(5), 2313–2343.

Sussmann, R., Forster, F., Rettinger, M., & Bousquet, P. (2012). Renewed methane increase for five years (2007-2011) observed by solar FTIR spectrometry. *Atmospheric Chemistry and Physics*, 12(11), 4885–4891.

Szunyogh, I., Kostelich, E., Gyarmati, G., Kalnay, E., Hunt, B., Ott, E., Satterfield, E., & Yorke, J. (2008). A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus A*, 60(1), 113–130.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.

Tarnocai, C. (2009). The impact of climate change on Canadian peatlands. *Canadian Water Resources Journal*, 34(4), 453–466.

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., & Zimov, S. (2009). Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochemical Cycles*, 23.

Theil, H. & Panne, C. V. D. (1960). Quadratic programming as an extension of classical quadratic maximization. *Management Science*, 7(1), pp. 1–20.

Thompson, R., Gerbig, C., & Roedenbeck, C. (2011). A Bayesian inversion estimate of N_2O emissions for western and central Europe and the assessment of aggregation errors. *Atmospheric Chemistry and Physics*, 11, 3443–3458.

Thoning, K., Tans, P., Komhyr, W., et al. (1989). Atmospheric carbon dioxide at Mauna Loa observatory 2. analysis of the NOAA GMCC data, 1974–1985. *Journal of Geophysical Research*, 94(06), 8549–8565.

Tian, H., Lu, C., Chen, G., Tao, B., Pan, S., Grosso, S. J. D., Xu, X., Bruhwiler, L., Wofsy, S. C., Kort, E. A., & Prior, S. A. (2012). Contemporary and projected biogenic fluxes of methane and nitrous oxide in North American terrestrial ecosystems. *Frontiers in Ecology and the Environment*, 10(10), 528–536.

Tian, H., Xu, X., Liu, M., Ren, W., Zhang, C., Chen, G., & Lu, C. (2010). Spatial and temporal patterns of CH_4 and N_2O fluxes in terrestrial ecosystems of North America during 1979-2008: application of a global biogeochemistry model. *Biogeosciences*, 7(9), 2673–2694.

Tian, H., Xu, X., Lu, C., Liu, M., Ren, W., Chen, G., Melillo, J., & Liu, J. (2011). Net exchanges of CO_2, CH_4, and N_2O between China's terrestrial ecosystems and the atmosphere and their contributions to global climate warming. *Journal of Geophysical Research – Biogeosciences*, 116.

US Energy Information Administration (2013). *Natural Gas Annual 2011*. Technical report.

US EPA (2013). *Inventory of U.S. greenhouse gas emissions and sinks: 1990 - 2011*. Technical Report EPA 430-R-13-001.

USDA National Agricultural Statistics Service (2012). Charts and maps, county maps, 2010.

van Hulzen, J., Segers, R., van Bodegom, P., & Leffelaar, P. (1999). Temperature effects on soil methane production: an explanation for observed variability. *Soil Biology and Biochemistry*, 31(14), 1919 – 1929.

Villani, M. G., Bergamaschi, P., Krol, M., Meirink, J. F., & Dentener, F. (2010). Inverse modeling of European CH_4 emissions: sensitivity to the observational network. *Atmospheric Chemistry and Physics*, 10(3), 1249–1267.

Vogelezang, D. H. P. & Holtslag, A. A. M. (1996). Evaluation and model impacts of alternative boundary-layer height formulations. *Boundary-Layer Meteorology*, 81(3-4), 245–269.

Waddington, J. & Roulet, N. (1996). Atmosphere-wetland carbon exchanges: Scale dependency of CO_2 and CH_4 exchange on the developmental topography of a peatland. *Global Biogeochemical Cycles*, 10(2), 233–245.

Walvoort, D. & de Gruijter, J. (2001). Compositional kriging: A spatial interpolation method for compositional data. *Mathematical Geosciences*, 33(8), 951–966.

Wang, J. & Zabaras, N. (2005). Hierarchical Bayesian models for inverse problems in heat conduction. *Inverse Problems*, 21(1), 183–206.

Wang, J. & Zabaras, N. (2006). A Markov random field model of contamination source identification in porous media flow. *International Journal of Heat and Mass Transfer*, 49(5-6), 939–950.

Wang, J. S., Logan, J. A., McElroy, M. B., Duncan, B. N., Megretskaia, I. A., & Yantosca, R. M. (2004). A 3-D model analysis of the slowdown and interannual variability in the methane growth rate from 1988 to 1997. *Global Biogeochemical Cycles*, 18(3).

Wania, R., Melton, J. R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Chen, G., Eliseev, A. V., Hopcroft, P. O., Riley, W. J., Subin, Z. M., Tian, H., van Bodegom, P. M., Kleinen, T., Yu, Z. C., Singarayer, J. S., Zurcher, S., Lettenmaier, D. P., Beerling, D. J., Denisov, S. N., Prigent, C., Papa, F., & Kaplan, J. O. (2013). Present state of global wetland extent and wetland methane modelling: methodology of a model inter-comparison project (WETCHIMP). *Geoscientific Model Development*, 6(3), 617–641.

Wania, R., Ross, I., & Prentice, I. C. (2010). Implementation and evaluation of a new methane model within a dynamic global vegetation model: LPJ-WHyMe v1.3.1. *Geoscientific Model Development*, 3(2), 565–584.

Wecht, K. J., Jacob, D. J., Frankenberg, C., Jiang, Z., & Blake, D. R. (2014). Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data. *Journal of Geophysical Research – Atmospheres*, 119(12), 7741–7756.

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

Western Climate Initiative (2010). Design for the WCI Regional Program.

Whalen, S. (2005). Biogeochemistry of methane exchange between natural wetlands and the atmosphere. *Environmental Engineering Science*, 22(1), 73–94.

Wilks, D. (2011). *Statistical Methods in the Atmospheric Sciences*. Oxford, UK: Academic Press, 3 edition.

Williams, I. N., Riley, W. J., Torn, M. S., Berry, J. A., & Biraud, S. C. (2011). Using boundary layer equilibrium to reduce uncertainties in transport models and CO_2 flux inversions. *Atmospheric Chemistry and Physics*, 11(18), 9631–9641.

Wilson, G. T. (1989). On the use of marginal likelihood in time series model estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(1), pp. 15–27.

Winderlich, J. (2012). *Setup of a CO_2 and CH_4 measurement system in Central Siberia and modeling of its results*. PhD thesis, University of Hamburg, Hamburg, Germany.

Worthy, D., Levin, I., Hopper, F., Ernst, M., & Trivett, N. (2000). Evidence for a link between climate and northern wetland methane emissions. *Journal of Geophysical Research – Atmospheres*, 105, 4031–4038.

Xiang, B., Miller, S. M., Kort, E. A., Santoni, G. W., Daube, B. C., Commane, R., Angevine, W. M., Ryerson, T. B., Trainer, M. K., Andrews, A. E., Nehrkorn, T., Tian, H., & Wofsy, S. C. (2013a). Nitrous oxide (N_2O) emissions from California based on 2010 CalNex airborne measurements. *Journal of Geophysical Research – Atmospheres*.

Xiang, B., Miller, S. M., Kort, E. A., Santoni, G. W., Daube, B. C., Commane, R., Angevine, W. M., Ryerson, T. B., Trainer, M. K., Andrews, A. E., Nehrkorn, T., Tian, H., & Wofsy, S. C. (2013b). Nitrous oxide (N_2O) emissions from California based on 2010 CalNex airborne measurements. *Journal of Geophysical Research – Atmospheres*, 118(7), 2809–2820.

Yadav, V., Mueller, K. L., & Michalak, A. M. (2013). A backward elimination discrete optimization algorithm for model selection in spatio-temporal regression models. *Environmental Modelling & Software*, 42(0), 88 – 98.

Young, P. (2012). *Annual Coal Report 2011*. Technical report, Office of Oil, Gas, and Coal Supply Statistics, US Energy Information Administration, Washington, DC 20585.

Yuan, Y.-x. (2000). *Proceedings of the Fourth International Congress on Industrial & Applied Mathematics, ICIAM 99, Edinburgh, [5 - 9 July 1999]*, chapter A review of trust region algorithms for optimization, (pp. 271–282). Oxford Univ. Press: Oxford, UK.

Zanini, A. & Kitanidis, P. K. (2009). Geostatistical inversing for large-contrast transmissivity fields. *Stochastic Environmental Research and Risk Assessment*, 23(5), 565–577.

Zhang, Y., Sachs, T., Li, C., & Boike, J. (2012). Upscaling methane fluxes from closed chambers to eddy covariance based on a permafrost biogeochemistry integrated model. *Global Change Biology*, 18(4), 1428–1440.

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

Zhu, X., Zhuang, Q., Chen, M., Sirin, A., Melillo, J., Kicklighter, D., Sokolov, A., & Song, L. (2011). Rising methane emissions in response to climate change in Northern Eurasia during the 21st century. *Environmental Research Letters*, 6(4), 045211.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, 44(1), 41 – 61.