**Atmospheric
Chemistry
and Physics**

Regional-scale geostatistical inverse modeling of North American CO$_2$ fluxes: a synthetic data study

S. M. Gourdji[1], **A. I. Hirsch**[2], **K. L. Mueller**[1], **V. Yadav**[1], **A. E. Andrews**[3], **and A. M. Michalak**[1,4]

[1]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, USA
[2]National Renewable Energy Laboratory, Golden, CO 80401, USA
[3]Global Monitoring Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO 80305, USA
[4]Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI 48109, USA

Abstract. A series of synthetic data experiments is performed to investigate the ability of a regional atmospheric inversion to estimate grid-scale CO$_2$ fluxes during the growing season over North America. The inversions are performed within a geostatistical framework without the use of any prior flux estimates or auxiliary variables, in order to focus on the atmospheric constraint provided by the nine towers collecting continuous, calibrated CO$_2$ measurements in 2004. Using synthetic measurements and their associated concentration footprints, flux and model-data mismatch covariance parameters are first optimized, and then fluxes and their uncertainties are estimated at three different temporal resolutions. These temporal resolutions, which include a four-day average, a four-day-average diurnal cycle with 3-hourly increments, and 3-hourly fluxes, are chosen to help assess the impact of temporal aggregation errors on the estimated fluxes and covariance parameters. Estimating fluxes at a temporal resolution that can adjust the diurnal variability is found to be critical both for recovering covariance parameters directly from the atmospheric data, and for inferring accurate ecoregion-scale fluxes. Accounting for both spatial and temporal a priori covariance in the flux distribution is also found to be necessary for recovering accurate a posteriori uncertainty bounds on the estimated fluxes. Overall, the results suggest that even a fairly sparse network of 9 towers collecting continuous CO$_2$ measurements across the continent, used with no auxiliary information or prior estimates of the flux distribution in time or space, can be used to infer relatively accurate monthly ecoregion scale CO$_2$ surface fluxes over North America within estimated uncertainty bounds. Simulated random transport error is shown to decrease the quality of flux estimates in under-constrained areas at the ecoregion scale, although the uncertainty bounds remain realistic. While these synthetic data inversions do not consider all potential issues associated with using actual measurement data, e.g. systematic transport errors or problems with the boundary conditions, they help to highlight the impact of inversion setup choices, and help to provide a baseline set of CO$_2$ fluxes for comparison with estimates from future real-data inversions.

1 Introduction

Improved estimates of regional-scale CO$_2$ land-atmosphere exchange are needed for the design and verification of carbon management policies, as well as for the validation of process-based models used to predict CO$_2$ fluxes. Continuous CO$_2$ fluxes cannot be directly measured at regional scales, and have instead been inferred from atmospheric concentration patterns using inverse modeling techniques. While earlier global inversion studies had used atmospheric CO$_2$ concentration measurements sampled in the free troposphere at remote or high-altitude locations to infer continental-scale CO$_2$ fluxes (e.g. Gurney et al., 2002a; Baker et al., 2006), the recent convergence of several factors has made it feasible to estimate sub-continental scale CO$_2$ fluxes in a regional inverse modeling framework (e.g. Peylin et al., 2005; Lauvaux

et al., 2008; Schuh et al., 2010). These factors include continuous ground-based measurements of atmospheric CO_2 taken at several North American and Eurasian sites (e.g. Bakwin et al., 1998; Haszpra, 1999) that provide data with high temporal (and, increasingly, high spatial) resolution to constrain carbon fluxes at finer scales. In addition, the continuous measurement locations tend to be sited in continental, low-altitude areas with strong biospheric activity, providing more information about flux variability at sub-continental scales relative to the measurements used in global inversions. Finally, recent advances in regional atmospheric transport modeling and the use of analyzed wind fields with high spatial resolution make it feasible to take advantage of continuous data from continental locations in regional inversions.

The use of continuous data in grid-scale CO_2 inversions is relatively new, and, therefore, many questions remain as to the optimal approach for taking advantage of these large and highly variable data streams. Synthetic data (a.k.a. "pseudo-data") experiments are useful in the design of inversions, because they include a set of specified baseline fluxes with which results can be compared, making it easier to diagnose any potential biases in inferred fluxes under a number of different scenarios (e.g. Law et al., 2002, 2003, 2004; Carouge et al., 2010a, b). Also, given a wide spread in inversion results, such as those from the North American Carbon Program Regional Interim Synthesis (Cook et al., 2009), synthetic data studies can help to isolate the impact of inversion setup choices, as opposed to other sources of bias in inversions, e.g. due to errors in the boundary conditions, or transport models.

The interpretation of flux estimates in a synthetic data inversion is simplified relative to a real data inversion in two important ways. First, synthetic measurements are only influenced by fluxes occurring within the domain of study, and therefore there is no need to specify boundary conditions at the edge of the domain. Errors in boundary conditions used in regional real-data inversions can bias flux estimates, particularly for smaller regions (e.g. Peylin et al., 2005; Göckede et al., 2010). Second, the effect of atmospheric transport model errors can be controlled by using the same transport model to create the synthetic measurements as is used to estimate fluxes in the inversion. The impact of transport model errors on atmospheric inversions has been a significant research focus for some time, principally at the global scale (e.g. Gurney et al., 2002b; Baker et al., 2006). However, transport model errors may be even more of a concern for regional inversions, where the models' ability to simulate small-scale variability in areas with high biospheric activity and/ or complex terrain is relatively unknown (Geels et al., 2004; Patra et al., 2008; Gerbig et al., 2009).

In addition to the simplifications associated with a synthetic data inversion, the geostatistical inversion approach (Michalak et al., 2004) further makes it possible to eliminate the impact of the choice of a priori flux estimates on inversions. Geostatistical inversions are Bayesian, but do not prescribe a prior estimate of the flux distribution from biospheric models and/or inventories. Therefore, this approach provides a unique opportunity to assess the information content of the available atmospheric measurement data (Mueller et al., 2008).

Geostatistical inversions, as well as some other recent inversions, also estimate fluxes directly on a grid at fine spatial resolutions, thereby minimizing spatial aggregation errors (e.g. Kaminski et al., 2001; Gourdji et al., 2008; Schuh et al., 2009) that can occur when fixed flux patterns are imposed for large regions (e.g. Law et al., 2002; Peters et al., 2007). These errors result because the inversion cannot adjust the flux patterns within specified regions, even though atmospheric observations are sensitive to sub-regional variability. Aggregation errors can be temporal as well as spatial, a topic explored in the current study. Temporal aggregation errors occur when estimating a single flux, or flux adjustment, over a time interval with significant intra-period variability, which is of particular concern for regional inversions using continuous data collected in areas with strong diurnal and synoptic flux variability. While estimating spatially and temporally finer scale fluxes helps to reduce aggregation errors, computational costs also grow as the number of estimated fluxes grows.

The current work uses a series of synthetic data inversion experiments to evaluate a regional geostatistical grid-scale ($1° \times 1°$) inversion for June 2004 over North America, using the nine continuous CO_2 observing towers operational in the United States and Canada at that time. The primary objectives of this study are threefold. First, we investigate the use of available atmospheric measurements to infer reasonable covariance parameters (both flux covariance and model-data mismatch), as well as the impact of these inferred parameters on the estimated fluxes. Second, we perform inversions estimating fluxes at three different temporal resolutions ranging from 3-hourly to 4-day averages, in order to evaluate the impact of temporal aggregation errors on inversion results. Third, we assess the information content of the limited atmospheric network for 2004, by comparing inversion results to the "true" fluxes at both the grid and aggregated eco-region scale. The effect of random transport model error is additionally explored throughout this study by adding random noise with realistic magnitude to the synthetic measurements, and then observing their impact on the inversion results.

2 Inversion method and components

This section summarizes the geostatistical inverse modeling method, and the setup of each component of the inversion for the analyses performed in this study.

2.1 Geostatistical inverse modeling (GIM)

The GIM approach to CO$_2$ flux estimation was developed to help limit the influence of a priori assumptions on inversion results (Michalak et al., 2004; Mueller et al., 2008; Gourdji et al., 2008). The approach is Bayesian, but, instead of prescribing a prior estimate of the flux distribution, the GIM approach uses: 1) a deterministic model of the trend that estimates the relationship of CO$_2$ flux to key covariates, and 2) a prior covariance matrix that describes the expected variability in flux departures from the trend, as a function of the separation distance in space and time between individual grid-scale fluxes. The model of the trend can be as simple as a single unknown mean flux across the domain, or can include more complex components, such as a linear combination of auxiliary variables related to CO$_2$ flux processes (Gourdji et al., 2008). The GIM equations are summarized below, and readers are referred to Michalak et al. (2004), Mueller et al. (2008), and Gourdji et al. (2008) for additional details.
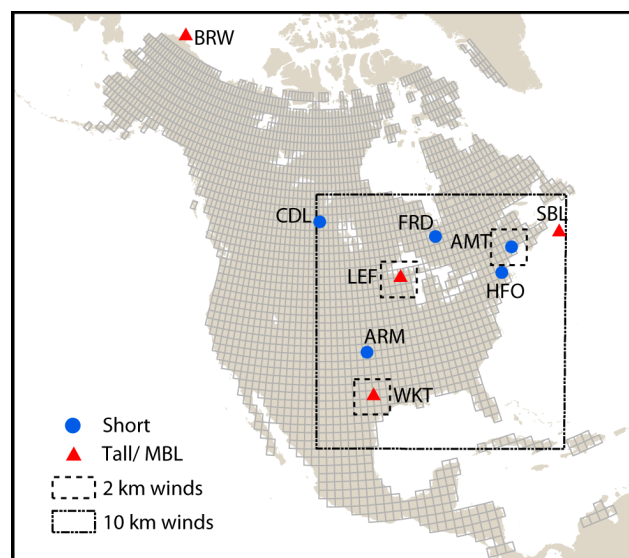
The GIM approach entails minimizing the following objective function:

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

(1)

where the vector z ($n \times 1$) represents the atmospheric CO$_2$ measurements (ppm), and s ($m \times 1$) is the vector of fluxes (μmol/(m^2 s)). $\mathbf{H}$ ($n \times m$) describes the sensitivity of CO$_2$ measurements to surface fluxes, as quantified from an atmospheric transport model, with units of ppm/(μmol/(m^2 s)), and $\mathbf{H}s$ therefore represents a vector of modeled CO$_2$ observations. $\mathbf{X}$ is a known ($m \times p$) matrix containing the flux covariates in the model of the trend, $\boldsymbol{\beta}$ are ($p \times 1$) unknown drift coefficients, and $\mathbf{X}\boldsymbol{\beta}$ is the model of the trend. The two covariance matrices in the objective function, $\mathbf{R}$ ($n \times n$) and $\mathbf{Q}$ ($m \times m$), balance the relative weight of the atmospheric data and the model of the trend in estimating the fluxes. $\mathbf{R}$ is the model-data mismatch covariance matrix, describing the expected magnitude of discrepancies between observed (z) and modeled ($\mathbf{H}s$) CO$_2$ concentrations (due to measurement, transport model, representation, and aggregation errors). $\mathbf{Q}$ ($m \times m$) is the a priori flux covariance matrix, characterizing how flux deviations from the model of the trend (i.e. $s - \mathbf{X}\boldsymbol{\beta}$) are correlated in time and space. The setup of each component of the GIM objective function is further discussed in Sects. 2.2 to 2.7.

The GIM problem involves estimating the fluxes, s, as well as the drift coefficients, $\boldsymbol{\beta}$ (e.g. Michalak et al., 2004). Minimizing Eq. (1) with respect to these variables yields a system of linear equations:

$$\begin{bmatrix} \mathbf{HQH}^{\mathrm{T}} + \mathbf{R} & \mathbf{HX} \\ (\mathbf{HX})^{\mathrm{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^{\mathrm{T}} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{HQ} \\ \mathbf{X}^{\mathrm{T}} \end{bmatrix}$$

(2)

Fig. 1. Location of nine measurement towers used in the study, as well as the domains for the two levels of high-resolution nesting with the WRF winds. Outside of the 2 and 10 km resolution nests, 40 km resolution winds were used for the remainder of the domain. The background grid represents the flux estimation resolution of $1° \times 1°$.

that is solved for the matrices $\boldsymbol{\Lambda}$ and $\mathbf{M}$, which are then used to define the a posteriori best estimates of s and $\boldsymbol{\beta}$ as:

$$\widehat{s} = \boldsymbol{\Lambda} z$$

(3)

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Q}^{-1}\boldsymbol{\Lambda} z$$

(4)

The best estimates of flux can alternately be expressed as the sum of a deterministic model of the trend ($\mathbf{X}\widehat{\boldsymbol{\beta}}$) and a spatiotemporally correlated stochastic component:

$$\widehat{s} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{QH}^{\mathrm{T}}\left(\mathbf{HQH}^{\mathrm{T}} + \mathbf{R}\right)^{-1}\left(z - \mathbf{HX}\widehat{\boldsymbol{\beta}}\right)$$
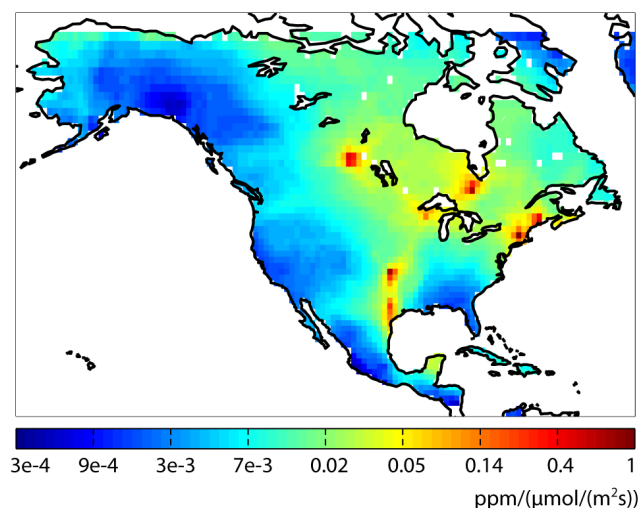
(5)

The a posteriori uncertainties of the grid-scale fluxes are given by:

$$\mathbf{V}_{\widehat{s}} = -\mathbf{XM} + \mathbf{Q} - \mathbf{QH}^{\mathrm{T}}\boldsymbol{\Lambda}^{\mathrm{T}}$$

(6)

where this full covariance matrix represents a composite of the uncertainty associated with the estimation of unknown drift coefficients ($\boldsymbol{\beta}$) in the model of the trend, the spatiotemporal variability of fluxes as represented in $\mathbf{Q}$, and the overall constraint on fluxes as determined by the concentration footprints, the model of the trend and the prior covariance matrices.

2.2 Flux estimation resolution (s)

For all inversions, fluxes are estimated at a $1° \times 1°$ grid-scale spatial resolution, with the domain including all land cells

Fig. 2. Average monthly sensitivity of June 2004 measurements at nine towers to all fluxes.

within the range of $10° - 70°$ N and $50° - 170°$ W, yielding 2641 estimation regions (Fig. 1). Fluxes are estimated from 1 June to 2 July 2004, in universal time (UTC) using three different temporal resolutions: 3-hourly (henceforth referred to as F3hr), a 4-day average diurnal cycle with 3-hourly time increments (F4d-diurnal), and a flat 4-day average without any diurnal variability (F4d). These three temporal resolutions make it possible to investigate the benefit of directly estimating the diurnal cycle of fluxes, and, conversely, the risk of temporal aggregation error associated with estimating fluxes averaged over multiple days. Despite the potential benefits associated with estimating finer-scale fluxes, the number of estimated fluxes and associated computational costs grow as the temporal resolution becomes finer, as shown in Table 1. Additional details associated with the setup for each temporal flux resolution are described in Sects. 2.3 to 2.7.

2.3 Atmospheric transport (H)

Atmospheric transport models are necessary for CO$_2$ inversions in order to quantify the sensitivity of measured concentrations to surface fluxes, or the concentration footprints that populate the atmospheric transport matrix **H**. The Stochastic Time-Inverted Lagrangian Transport Model (STILT) model (Lin et al., 2003) is used for the current study. STILT, which has already been applied in several pilot studies aimed at constraining CO$_2$ sources and sinks in the United States (Gerbig et al., 2003, 2006; Lin et al., 2004; Matross et al., 2006), represents air arriving at observation locations as an ensemble of particles that are transported backward in time. The particle velocities in STILT are in turn derived from meteorological fields generated by gridded numerical weather prediction models, in this case from the Weather Research & Forecast-

ing (WRF) model (Skamarock et al., 2005), version 2.2. For this study, WRF v2.2 was configured to use three levels of high resolution nesting: a 2-km resolution grid around the three tallest measurement towers (LEF, AMT and WKT, see Table 2 and Fig. 1), embedded in a 10-km resolution grid over the northern Midwest, Gulf Region, and New England extending to approximately 105° W, and then an outermost 40-km resolution grid covering the remainder of the overall domain of the inversions.

At each measurement location, 10-day back-trajectories of 500 particles were generated using STILT every hour from 1 June to 8 July 2004. Concentration footprints, or sensitivities, were then calculated at 3-hourly intervals back in time, by integrating these particle trajectories over the North American $1° \times 1°$ grid as described in Lin et al. (2003). Finally, these high-resolution **H** matrices were aggregated to the temporal resolution of the concentration data, described in Sect. 2.4, and the three flux temporal resolutions.

A map of the average sensitivity of measurements to fluxes for June 2004, derived from the concentration footprints, is shown in Fig. 2. As seen here, many parts of North America are not well-constrained by the 9-tower measurement network in 2004. These areas include northwest Canada and Alaska, the southwestern and southeastern United States, and parts of Central America. In contrast, the eastern temperate forests and Midwestern agricultural areas have a stronger atmospheric data constraint.

2.4 Synthetic concentration time series (z)

One goal of this study is to assess the projected accuracy of North American estimates of CO$_2$ flux using a contemporary observation network. Therefore, synthetic data were generated at the highest sampling elevation of the nine towers that were collecting continuous high-precision calibrated CO$_2$ measurements in North America in June of 2004 (Fig. 1, Table 2). A full set of synthetic measurements without data gaps from 1 June to 8 July were generated by multiplying 3-hourly CO$_2$ surface flux estimates (**s**) from a biospheric model by the atmospheric transport matrices (**H**).

The biospheric fluxes used in this study are taken from the Carnegie Ames Stanford Approach terrestrial carbon cycle model, as configured for the Global Fire Emissions Database v2 project (henceforth referred to as CASA-GFEDv2; Randerson et al., 1997; Van der Werf et al., 2006). CASA-GFEDv2 was chosen because it is a well-accepted model that has been used for specifying prior flux estimates in several synthesis Bayesian inversion studies (e.g. Baker et al., 2006; Peters et al., 2007), although the choice of biospheric model is flexible here, given that the aim of a synthetic data inversion is to assess the accuracy of the setup relative to a given set of prescribed fluxes. The monthly-average CASA-GFEDv2 Net Ecosystem Exchange (NEE) values were temporally downscaled to a 3-hourly resolution in order to test the ability of the inversion setup to accurately

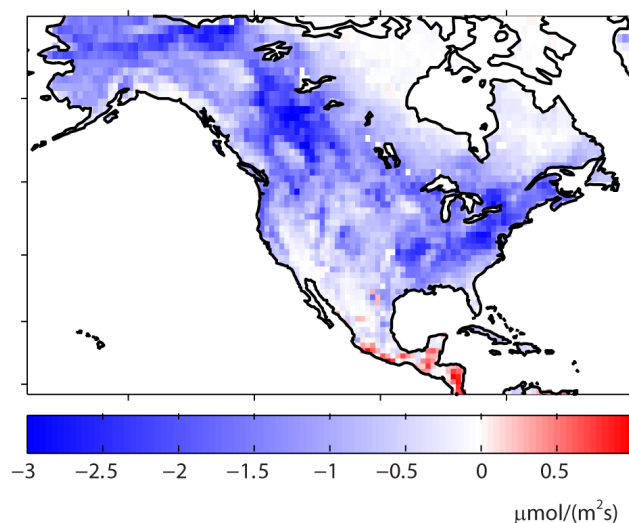**Table 1.** Inversion characteristics for three flux temporal resolutions.

Case	Flux resolution	# of estimated fluxes	Structure of trend (**X**)	Structure of flux covariance (**Q**)
F4d	4-day average	$(2641 \times 8) = 21\,128$	One spatiotemporal mean	Full spatiotemporal flux covariance
F4d-diurnal	4-day average diurnal cycle (with 3-hourly bins)	$(2641 \times 64) = 169\,024$	Eight spatial means by 3-hourly bins	Full spatiotemporal flux covariance across 4-day periods, but not within diurnal cycle
F3hr	3-hourly	$(2641 \times 256) = 676\,096$	Eight spatial means by 3-hourly bins	Full spatiotemporal flux covariance across days, but not within diurnal cycle

Table 2. Measurement locations used in the inversions.

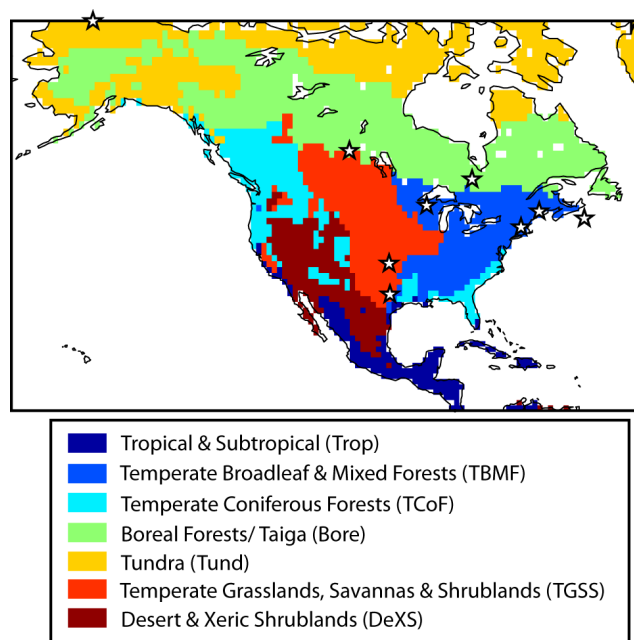| Tower	Location	Coordinates	Height	Maintained by	Type
LEF	Park Falls, Wisconsin	45.93 N, 90.27 W	396 m	NOAA/GMD	Tall
WKT	Moody, Texas	31.32 N, 97.33 W	457 m	NOAA/GMD	Tall
BRW	Barrow, Alaska	71.32 N, 156.60 W	10 m	NOAA/GMD	MBL
SBL	Sable Island, Nova Scotia	43.93 N, 60.02 W	25 m	Met Service Canada	MBL
AMT	Argyle, Maine	45.03 N, 68.68 W	107 m	NOAA/GMD	Short
ARM	Norman, Oklahoma	36.62 N, 97.50 W	60 m	US Dept. of Energy	Short
CDL	Candle Lake, Saskatchewan	53.99 N, 105.12 W	30 m	Met Service Canada	Short
FRD	Fraserdale, Ontario	49.84 N, 81.52 W	40 m	Met Service Canada	Short
HFO	Petersham, Massachusetts	42.54 N, 72.17 W	30 m	Harvard University	Short

recover diurnally-varying fluxes. This was accomplished using the method of Olsen and Randerson (2004), which is based on net shortwave radiation and near-surface temperature data from the NASA Global Land Data Assimilation System (GLDAS; Rodell et al., 2004). These downscaled 3-hourly CASA-GFEDv2 fluxes, shown in Fig. 3 at the aggregated monthly scale, represent the "truth" to which inversion results are compared.

The vector of modeled observations (i.e. **H**s), obtained by multiplying the "true" 3-hourly fluxes by the concentration footprints, was first generated at the hourly resolution corresponding to the STILT particle releases for the nine tower locations. Then, the synthetic observation vectors were averaged to a 3-hourly timescale. Sensitivity tests were additionally performed for inversions using daily and 8-day average concentration vectors. These tests revealed that using higher temporal resolution observations yielded superior flux estimates, consistent with Law et al. (2002), who found that using 4-hourly measurements relative to more coarsely-averaged observations helped to reduce biases over the Australasian subcontinent when fluxes were estimated at a sufficiently fine spatial resolution. As a result, only cases considering 3-hourly averaged observations are presented here.

Fig. 3. "True" CASA-GFEDv2 fluxes, aggregated to the monthly scale.

Finally, for inversions that simulate the effect of transport model error, three sets of inversions using different realizations of uncorrelated errors added to the synthetic

Fig. 4. Seven ecoregions (modified from Olson, 2001) used for analyzing inversion results at spatially aggregated scales. Stars represent nine measurement locations.

measurements were conducted. The variance of these errors remains the same across realizations, and corresponds to the expected magnitude of model-data mismatch seen in real measurements for each tower, as discussed in Sect. 3.1.

2.5 Use of night-time measurements

Due to stable conditions, night-time measurements taken from shorter towers within the nocturnal boundary layer provide little information about fluxes at large scales (Haszpra, 1999). In addition, meteorological fields used in transport models have difficulty in reliably simulating the height of the night-time planetary boundary layer (PBL), or the sharp gradient across it, which can lead to biased flux estimates from inversions using night-time measurements from within the PBL (Geels et al., 2007). For example, Gerbig et al. (2008) found biases of up to 50% in night-time PBL height in a study using high-resolution winds from the European Centre for Medium-Range Weather Forecasts (i.e. ECMWF, available from http://data.ecmwf.int/data/); comparison of the high resolution WRF wind fields (T. Nehrkorn, personal communication) used in this study with wind profiler PBL-height measurements yielded a similar conclusion.

Given their local footprints, the use of night-time near-surface measurement data in regional inversions could lead to higher aggregation errors near the towers, relative to those for afternoon measurements sampling well-mixed air. In addition, biases in night-time PBL height would affect future real-data inversions. For these reasons, only afternoon mea-

surements were included here for the five "Short" towers (see Table 2) that are consistently within the nocturnal boundary layer. For these towers, which are all in the Eastern or Central Standard Time zones, "afternoon" was considered to be 18:00–24:00 UTC. In contrast, all 24 h of atmospheric data were included in the inversions for the four tall (≥400 m) or marine boundary layer towers ("Tall" or "MBL" in Table 2). At these towers, observations sample relatively well-mixed air throughout the diurnal cycle, and therefore night-time measurements are assumed to be better-represented by the WRF/ STILT model relative to those at Short towers. This was qualitatively confirmed by comparing the diurnal cycle of actual observations at the tallest sampling levels of these towers to those from transported CASA-GFEDv2 fluxes.

Two sensitivity tests were performed to evaluate the choice of including night-time data for Tall and MBL towers and excluding them for Short towers. First, night-time data were included for the Short towers, such that 24 h of measurements were used for all sampling locations. Second, night-time data were excluded for the Tall/MBL towers, such that only afternoon measurements were used for all nine towers. (For the MBL towers, afternoon values were shifted to reflect local time zones.) Overall, such experiments help to assess biases associated with the use of night-time measurements, relative to the potential additional constraint on fluxes they can provide.

2.6 Model of the trend (Xβ)

A very simple model of the trend is applied in the current study for all inversions, analogous to those used in Michalak et al. (2004) and Mueller et al. (2008), where no additional auxiliary environmental variables are included in the model of the trend. As discussed in Sect. 2.1, the flux estimates ($\hat{s}$) are a composite of the inferred trend ($\mathbf{X}\hat{\beta}$) and a spatiotemporally-correlated stochastic component, such that any flux variability seen through the atmospheric observations but not captured by the trend can be still recovered through the stochastic component of the best estimate. For the F4d inversions, $\mathbf{X}$ is represented as a vector of ones, such that the corresponding drift coefficient (β) represents the mean value in space and time of fluxes across all grid-cells. For the two temporal resolutions resolving the diurnal cycle (F4d-diurnal and F3hr), the $\mathbf{X}$ matrix is instead structured to allow for eight spatial means defined for each 3-hourly bin of the diurnal cycle. Longitudinal gradients that could capture the changing day/ night gradient across the continent for different UTC time intervals were also considered for these inversions, but ultimately not included because they did not improve flux estimates.

2.7 Covariance matrices (Q and R)

Model-data mismatch errors are assumed uncorrelated in space and time, yielding a diagonal matrix $\mathbf{R}$, as is typical

in most inversion studies. A different variance was used for each measurement tower, based on results from initial tests showing significantly reduced errors in inferred fluxes as compared to an inversion using only two separate variances for Tall/MBL and Short towers.

In contrast to the diagonal structure of **R**, the covariance matrix **Q** contains off-diagonal entries describing the spatial and/or temporal correlation of the flux deviations from the model of the trend **Xβ**. The estimated fluxes for this study are sorted first in space, and then in time. Therefore, if only spatial covariance were considered, **Q** would be a block diagonal matrix, with each block describing the correlation between grid-scale fluxes for each time period of the inversion. (Because of the simple models of the trend used in this study (Sect. 2.6), the flux deviations in the diagonal blocks of **Q** represent residuals from a constant mean, and therefore these blocks describe the covariance of the fluxes themselves.) When temporal covariance is additionally considered, the off-diagonal blocks in **Q** contain diagonal entries describing the correlation among grid-cells with themselves over time. Finally, if cross spatial-temporal covariance is included, the off-diagonal blocks in **Q** become full, and they describe the spatial covariance between fluxes across different time periods.

In the current study, preliminary tests showed that including full spatiotemporal covariance between grid-scale fluxes helped to recover accurate uncertainty bounds for recovered fluxes, especially at spatially and temporally aggregated scales. (A subset of this analysis is presented in Table 5.) Therefore, cross spatial-temporal covariance was included in **Q** for all flux temporal resolutions. However, for the two flux resolutions resolving the diurnal cycle, spatial-temporal covariance is only assumed for the same 3-hourly interval across days or periods, but not within the diurnal cycle. For example, grid-scale fluxes from 00:00–03:00 UTC are correlated with fluxes from 00:00–03:00 UTC in neighboring days or periods, but not with fluxes from 03:00–06:00 UTC.

The correlation structure in **Q** is modeled using a covariance function that varies in space and time as a function of separation distance. Here, as in Michalak et al. (2004), we use an isotropic exponential decay model:

$$\mathbf{Q}\left(h_x, h_t \mid \sigma_\mathbf{Q}^2, l, \tau\right) = \sigma_\mathbf{Q}^2 \exp\left(-\frac{h_x}{l}\right)\exp\left(-\frac{h_t}{\tau}\right) \qquad (7)$$

where h_x and h_t are the separation distances between grid cells in space and time, respectively, l is the spatial correlation range parameter, τ is the temporal correlation range parameter, and $\sigma_\mathbf{Q}^2$ is the asymptotic variance of fluxes at large separation distances. The correlation lengths for an exponential model are approximately $3l$ and 3τ.

Multiple variance parameters ($\sigma_\mathbf{Q}^2$) were initially considered for different times of the day for those inversions resolving the diurnal cycle (i.e. F4d-diurnal and F3hr), because the underlying "true" CASA-GFED fluxes are significantly more

variable during the day-time compared to the night. However, the use of multiple variance parameters resulted in only small changes to the inferred fluxes, and in some areas biased the results. Therefore, for simplicity, only one flux variance parameter was used for each inversion.

Covariance parameters were estimated from the atmospheric data using the Restricted Maximum Likelihood (RML) approach, described in detail for atmospheric applications in Michalak et al. (2004) and Mueller et al. (2008). More specifics on the covariance parameter optimization for both **R** and **Q** are included below in Sect. 3.1.

3 Covariance parameter optimization

This section describes the approach taken to estimate covariance parameters in this study, as well as an analysis of the inferred parameters.

3.1 Setup for testing RML optimization with atmospheric data

To infer unbiased fluxes with accurate uncertainty estimates, it is important to correctly specify the flux covariance parameters (Gerbig et al., 2006), as well as the model-data mismatch variances. The RML approach provides a way to statistically optimize these parameters using the atmospheric data in an inverse setup (henceforth referred to as RML-Inv). If the recovered covariance parameters can be shown to yield accurate flux estimates, then this approach eliminates the need to use proxy methods for estimating covariance parameters. While this approach was previously demonstrated in Michalak et al. (2004) to perform well with synthetic data experiments for the global scale, it is investigated here for regional inversions using continuous data.

To estimate covariance parameters with the atmospheric data, the RML-Inv approach minimizes the negative log-likelihood of the available atmospheric observations with respect to the covariance parameters ($\boldsymbol{\theta}$), which include the three flux covariance parameters in **Q** ($\sigma_\mathbf{Q}^2$, l and τ) and the nine model-data mismatch variances in **R** (one for each tower). The corresponding objective function is (Kitanidis, 1995):

$$\begin{aligned}
L = \ln\left|\mathbf{HQH}^\mathrm{T}+\mathbf{R}\right| + \ln\left|(\mathbf{HX})^\mathrm{T}\left(\mathbf{HQH}^\mathrm{T}+\mathbf{R}\right)^{-1}\mathbf{HX}\right| \\
+ \left[z^\mathrm{T}\left(\left(\mathbf{HQH}^\mathrm{T}\right)^{-1}-\left(\mathbf{HQH}^\mathrm{T}\right)^{-1}\mathbf{HX}\right.\right. \\
\cdot \left.\left.\left((\mathbf{HX})^\mathrm{T}\left(\mathbf{HQH}^\mathrm{T}\right)^{-1}\mathbf{HX}\right)^{-1}(\mathbf{HX})^\mathrm{T}\left(\mathbf{HQH}^\mathrm{T}\right)^{-1}\right)z\right]
\end{aligned}$$
$$(8)$$

To test performance, parameters inferred using RML-Inv for each flux temporal resolution are compared to reference values, derived using the underlying true fluxes, as described

Table 3. Inferred parameters using RML-Krig and RML-Inv for the a priori flux covariance matrix (**Q**). RML-Inv parameters were estimated using observations with and without simulated transport error. The cell shading indicates the factor by which RML-Inv parameters differ from the RML-Krig, a.k.a. reference, values:

(no fill)   = 2×   2× to 4×   3× to 4×   >4×

Q	F4d			F4d-diurnal			F3hr		
		RML-Inv			RML-Inv			RML-Inv	
	RML-Krig	Perfect trans.	Trans. error	RML-Krig	Perfect trans.	Trans. error	RML-Krig	Perfect trans.	Trans. error
σ_Q^2 (μmol/(m^2 s))^2	1.0	97.1	25.8	11.1	35.5	27.3	13.4	24.3	20.7
l (km)	610	0	57	809	389	363	601	661	528
τ (days)	6.8	2.6	6.9	80.7	9.2	9.6	8.6	2.7	3.5

Table 4. "True" and RML-Inv inferred variances by tower for the model-data mismatch matrix (**R**). Both "True" and RML-Inv results are shown as calculated using observations with and without simulated transport error. The cell shading indicates the factor by which RML-Inv parameters differ from the "true", or reference, values:

(no fill)   = 2×   2× to 3×   3× to 4×   >4×

R (ppm^2)	F4d				F4d-diurnal				F3hr			
	"True"		RML-Inv		"True"		RML-Inv		"True"		RML-Inv	
	Perfect trans.	Trans. error	Perfect trans.	Trans. error	Perfect trans.	Trans. error	Perfect trans.	Trans. error	Perfect trans.	Trans. error	Perfect trans.	Trans. error
LEF	10.7	13.8	2.7	7.9	2.2	4.7	0.1	2.7	0.0	1.8	0.0	1.9
WKT	5.0	18.0	2.3	15.7	0.7	10.5	0.1	7.9	0.0	10.1	0.0	8.8
SBL	6.2	10.6	0.1	5.6	2.7	7.3	0.1	5.0	0.0	4.6	0.0	4.3
BRW	0.2	1.3	0.0	1.1	0.1	1.2	0.0	1.1	0.0	1.1	0.0	1.2
ARM	3.5	14.9	0.1	13.1	1.2	12.4	0.4	10.3	0.0	11.1	0.0	11.2
HF	15.6	34.4	5.2	25.7	8.8	24.4	6.8	17.7	0.0	23.5	2.2	23.2
AMT	9.5	23.4	3.9	17.6	4.8	21.5	1.3	14.9	0.0	11.6	0.3	13.4
FRD	3.4	11.7	0.0	7.4	2.4	11.1	0.0	7.1	0.0	8.8	0.0	8.6
CDL	2.9	3.6	0.0	1.0	1.2	1.4	0.0	0.0	0.0	0.0	0.0	0.0

below. In addition, the impact of using RML-Inv parameters on estimated fluxes is investigated in Sect. 4.

Because this is a pseudo-data setup, the covariance parameters for **Q** can also be estimated by implementing RML directly on the "true" underlying CASA-GFEDv2 fluxes (*s*). This approach will henceforth be referred to as RML-Krig, where "Krig" refers to the kriging setup of RML; see Mueller et al. (2008) and Gourdji et al. (2008) for details. Here, the RML objective function becomes:

$$L_Q = \frac{1}{2}\ln|\mathbf{Q}| + \frac{1}{2}\ln\left|\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right|$$
$$+ \frac{1}{2}\left[s^T\left(\mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{X}\left(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Q}^{-1}\right)s\right] \quad (9)$$

Because RML-Krig does not use the atmospheric measurements, the RML-Krig parameters are not affected by the simulation of transport error.

In order to derive "true" model-data mismatch variances in **R**, an approach other than RML-Krig must be applied using the true fluxes transported forward to the measurement locations. In the current study, fluxes are estimated at the native spatial resolution of the true fluxes (i.e. 1° × 1°). Therefore, there is technically no spatial aggregation error, and for cases that do not consider simulated random transport error, the model-data mismatch variances in **R** are exclusively determined by temporal aggregation error. These errors can be directly calculated as the variance of the difference between two synthetic data vectors: the observations used in the inversion (generated using 3-hourly fluxes), and a second set of observations generated using fluxes pre-averaged to coarser timescales (i.e. the 4-day or 4-day diurnal cycle) and then multiplied by aggregated transport matrices (**H**). Temporal aggregation error is technically zero in this study when

Table 5. Percent of true fluxes falling within two standard deviations of the a posteriori grid-scale monthly flux estimates. Results are shown without added transport errors for the three flux temporal resolutions, two sets of covariance parameters, and with and without assumed temporal flux covariance in **Q**.

	Reference parameters		RML-Inv parameters	
	Spatial covariance only	Spatiotemporal covariance	Spatial covariance only	Spatiotemporal covariance
F4d	66%	88%	100%	100%
F4d-diurnal	80%	96%	98%	99%
F3hr	49%	93%	63%	95%

estimating 3-hourly fluxes, although a floor of 0.01 ppm^2 is set for the model-data mismatch variance in all inversions using this flux resolution.

For inversions that consider transport model errors, the variance of the noise added to the measurements is added to the temporal aggregation error variance to arrive at the "true" total model-data mismatch variance. The variances of the added noise were determined per tower in the following manner. RML-Inv was first used to estimate **Q** and **R** parameters with actual atmospheric measurements for June 2004. Then, the difference between these "real data" model-data mismatch variances and the RML-Inv inferred temporal aggregation errors (with synthetic data) were taken as a measure of the magnitude of residual model-data mismatch for each tower, most of which is likely attributable to transport model error. This procedure was repeated for each of the examined flux resolutions, and then these inferred differences by tower were used as the variance of the random noise added to the synthetic measurements for the transport error analyses.

3.2 Comparison of reference covariance parameters by flux temporal resolution

The RML-Krig **Q** parameters and the "true" **R** covariance parameters are shown in Tables 3 and 4 for each of the examined flux resolutions. Because these parameters were inferred using the true underlying fluxes, they are henceforth referred to as reference values, for later comparison with the RML-Inv parameters inferred with the synthetic measurements. However, first we compare the reference parameters themselves across flux temporal resolutions.

The RML-Krig **Q** parameters show that the spatial correlation ranges (l) are not substantially different across temporal resolutions, but that the overall variance ($\sigma_{\mathbf{Q}}^2$) of the fluxes decreases as the estimated temporal resolution becomes coarser. This is expected, as more of the short-term variability in the spatial flux distribution is averaged out. The temporal correlation range (τ) is much longer for the 4-day diurnal cycle relative to the other timescales, although this is most likely an unreliable value given that the calculated
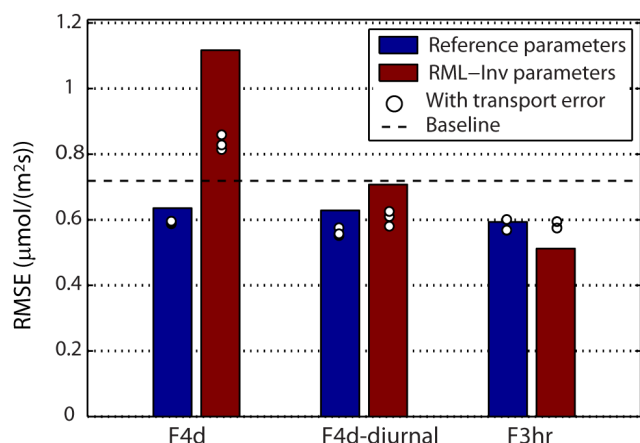
correlation length is much longer than the one-month time period of analysis.

The "true" model-data mismatch variances in **R** for inversions without transport error represent temporal aggregation errors, and they become higher, as expected, as the flux temporal resolution becomes coarser. This is the essence of aggregation error, where averaging out the "true" temporal variability in the fluxes and then transporting them forward to the sampling locations cannot properly reproduce the measured concentrations. For a given temporal resolution, these aggregation errors also tend to be higher for towers in highly active biospheric regions (e.g. LEF, AMT and HF), where temporal variability in nearby fluxes has a strong influence on measured concentrations. The "true" model-data mismatch with simulated transport error is increased by the magnitude of the random noise added to the measurements, which varies from a standard deviation of about 0.5 ppm for CDL to about 5 ppm for AMT.

3.3 Results of RML-Inv optimization

Estimated RML-Inv parameters for **Q** and **R** are also shown in Tables 3 and 4. The RML-Inv values for inversions with transport error represent the average parameters inferred using three realizations of random noise. Although RML-Inv parameters are compared here with the RML-Krig values (for **Q**) and "true" model-data mismatch variances (for **R**) to assess the relative ability of the atmospheric data to recover covariance parameters, the ultimate concern is the impact of these parameters on the inversions, results of which are presented in Sect. 4.

The correspondence between the inferred RML-Inv parameters and the reference values, as seen in Tables 3 and 4, is found to be strongly dependent on the flux temporal resolution. For the F3hr case, RML-Inv is able to recover covariance parameters within a factor of two for the flux variance ($\sigma_{\mathbf{Q}}^2$), spatial correlation range (l) and seven of the nine model-data mismatch variances with perfect transport. In contrast, for the F4d case, most of the recovered RML-Inv covariance parameters differ by more than a factor of four, particularly the flux variance ($\sigma_{\mathbf{Q}}^2$). The F4d-diurnal

Fig. 5. Root Mean Square Error (RMSE) between estimated and "true" grid-scale fluxes, aggregated to monthly averages, for inversions with two sets of covariance parameters and three flux temporal resolutions. Bars show the RMSE for inversions without transport error, and the white dots show the RMSE for inversions conducted with three different realizations of simulated transport error. The baseline RMSE, as described in Sect. 4.2, represents the value associated with inferring a perfect flat mean monthly flux across the continent.

case exhibits intermediate performance between the other flux resolutions.

The flux variance parameter ($\sigma_{\mathbf{Q}}^2$) is higher than the reference value for all temporal resolutions. This is most likely due to the fact that the majority of towers are sited in biospherically active regions, which have above-average flux variability as compared to the continent as a whole. Surprisingly, the addition of simulated random transport error helps to bring the flux variance closer to the reference value, but this may be due to the transport errors obscuring some of the "true" flux variability that would otherwise be seen through the measurement data.

Overall, estimating covariance parameters with the atmospheric data and the coarse flux resolution (F4d) appears to yield consistently unreliable parameter estimates, whereas the RML-Inv approach with the other two temporal resolutions yields results that are more consistent with the reference values. The impact of these inferred parameter estimates on inversion results in explored in the next section.

4 Inversion experiments

This section describes the setup and results of inversions used to test the impact of covariance parameter optimization methods, flux temporal resolutions, and other inversion assumptions and data choices.

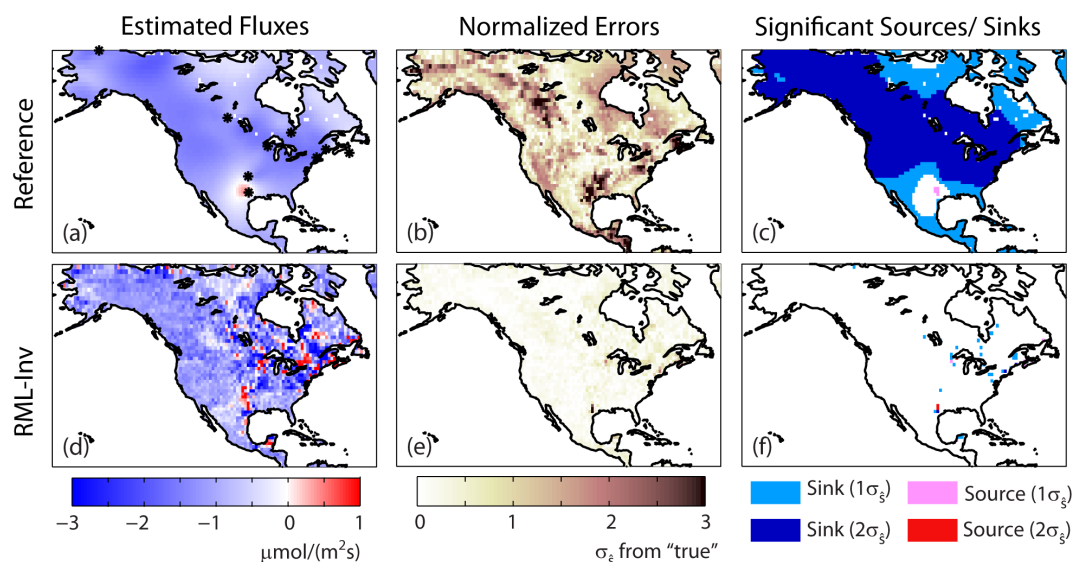### 4.1 Inversion setups and diagnostics

After the covariance parameter optimization analysis, a series of inversions was run to test the impact on inferred fluxes and uncertainties of a) varying the temporal flux resolution (shown in Table 1), b) using inferred RML-Inv vs. the reference covariance parameters, and c) including simulated random transport error. All combinations of a), b) and c) yielded 24 inversions, i.e. 2 sets of covariance parameters for each of 3 flux temporal resolutions, once with perfect transport and three times with different realizations of random transport error. In addition, sensitivity tests were run for the inversions with no transport model error to test the impact of excluding cross spatial-temporal covariance in **Q**, and the inclusion or exclusion of night-time data.

The a posteriori flux estimates from all inversions, as well as the true CASA-GFEDv2 fluxes, were averaged to a monthly timescale in order to compare results at a scale relevant for carbon-cycle science. The inferred grid-scale fluxes were compared using two quantitative metrics. First, the root mean square error (RMSE) (e.g. Law et al., 2002) between the true and estimated fluxes was calculated at the native 1° × 1° spatial resolution for all land grid-cells across the continent (Fig. 5). Second, the accuracy of the estimated a posteriori uncertainties (from Eq. 6) was evaluated by calculating the percent of 1° × 1° true fluxes that fall within two standard deviations of the estimated fluxes (Table 5). Ideally, 95% of fluxes should fall within this interval. Values significantly below 95% indicate an underestimation of the true a posteriori uncertainties. The results of this second metric are compared for inversions with and without cross spatial-temporal covariance in **Q**, in order to examine the impact of accounting for temporal covariance on the recovered flux uncertainties.
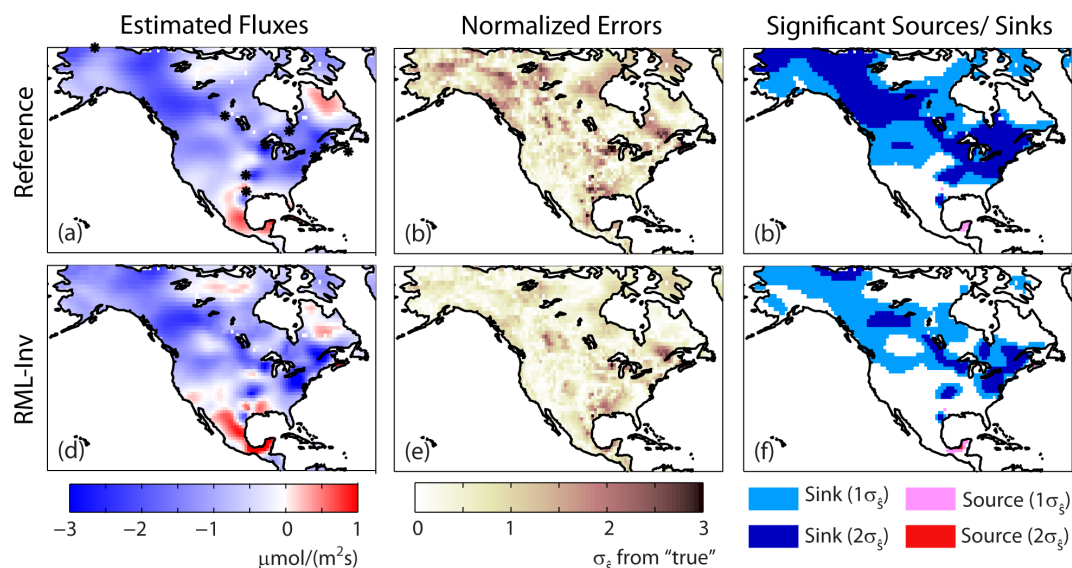
Inversion results were also compared qualitatively by examining the spatial patterns of inferred fluxes to those of the true fluxes (Figs. 6, 7, and 8). Finally, monthly fluxes and uncertainties were aggregated to seven ecoregions (Fig. 4) as well as to the North American continent (Fig. 9). These ecoregions are loosely defined based on the work of Olson (2001), and represent large, mostly contiguous, regions with similar climate, land cover and land use. An area-weighted RMSE at the ecoregion scale was also calculated.

4.2 Results of grid-scale diagnostics

Figure 5 shows the grid-scale RMSE's for the six inversions with no transport error (using two sets of covariance parameters and grouped by the three examined flux resolutions). With the reference covariance parameters, there is little difference in continental grid-scale RMSE among the three flux resolutions, although the 3-hourly resolution shows a slight advantage. Using the RML-Inv parameters, inversion performance degrades for coarser estimation timescales, consistent with the fact that the recovered RML-Inv covariance
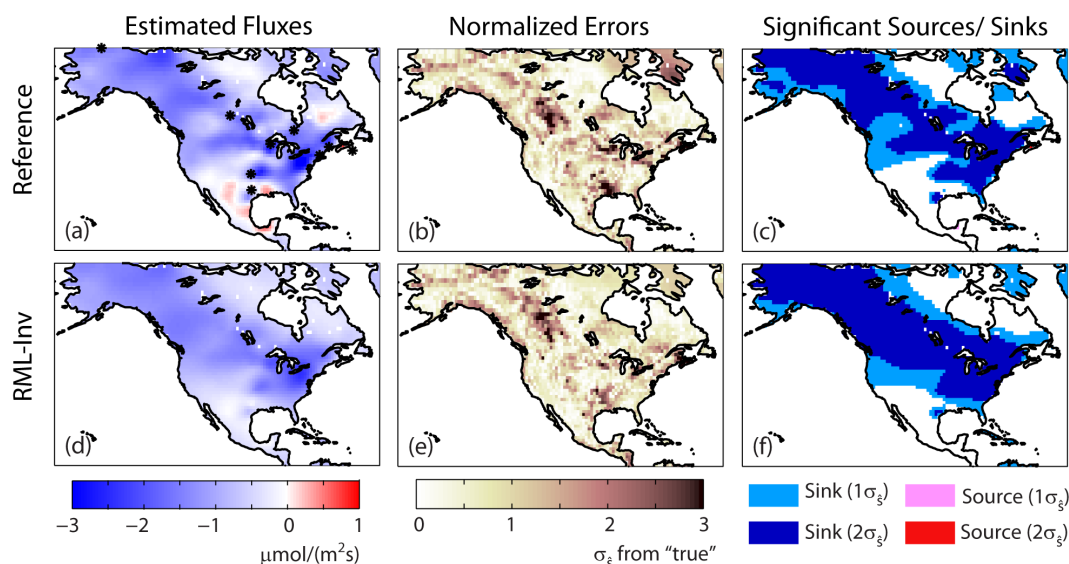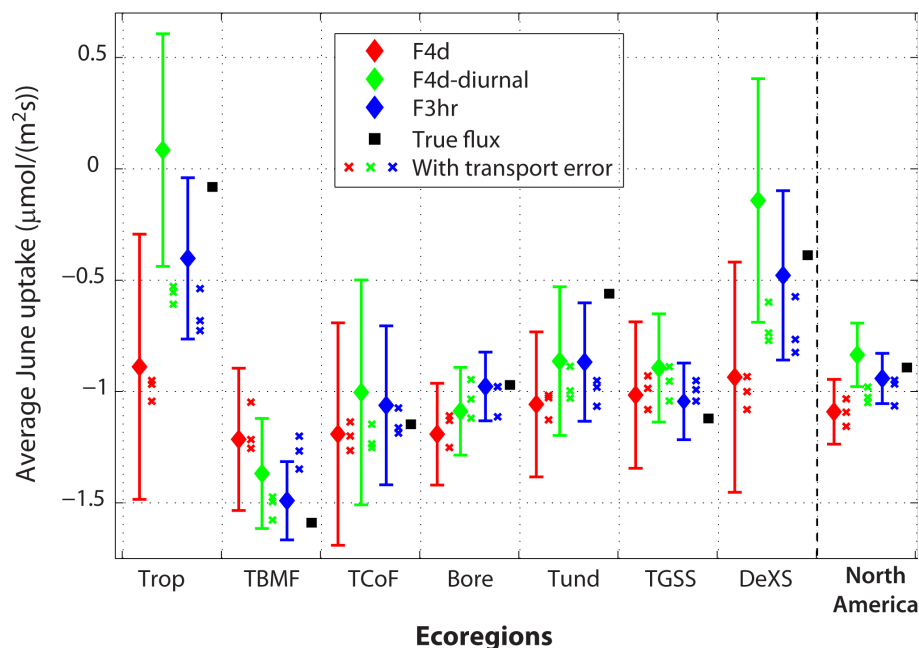
Fig. 6. Monthly grid-scale fluxes estimated from F4d inversion. The first row presents results obtained using the reference covariance parameters, and the second row shows results obtained using RML-Inv parameters. The maps show a posteriori fluxes aggregated to monthly averages (**a, d**), errors relative to the true fluxes (Fig. 3) normalized by the a posteriori standard deviations (**b, e**), and locations of significant sources and sinks at $1\sigma_{\hat{s}}$ and $2\sigma_{\hat{s}}$ (**c, f**).

Fig. 7. Monthly grid-scale fluxes estimated from F4d-diurnal inversion. The first row presents results obtained using the reference covariance parameters, and the second row shows results obtained using RML-Inv parameters. The maps show a posteriori fluxes aggregated to monthly averages (**a, d**), errors relative to the true fluxes (Fig. 3) normalized by the a posteriori standard deviations (**b, e**), and locations of significant sources and sinks at $1\sigma_{\hat{s}}$ and $2\sigma_{\hat{s}}$ (**c, f**).

parameters became farther from the reference values as fluxes were temporally aggregated (Sect. 3.3). If the inversion were to infer the exact mean monthly flux across the continent with no spatiotemporal variability, the RMSE would be 0.72 μmol/(m^2 s). Therefore, all inversions, except for the F4d case with RML-Inv parameters, perform better than this baseline value.

The RMSE corresponding to the three realizations of simulated transport error for each inversion setup are also shown in Fig. 5. The grid-scale fluxes (results not shown) from these inversions show that transport error has the effect of damping down the variability in the inferred flux signal. For the two timescales with temporal aggregation error (i.e. F4d and F4d-diurnal), this has a positive impact on the RMSE, with

Fig. 8. Monthly grid-scale fluxes estimated from F3hr inversion. The first row presents results obtained using the reference covariance parameters, and the second row shows results obtained using RML-Inv parameters. The maps show a posteriori fluxes aggregated to monthly averages **(a, d)**, errors relative to the true fluxes (Fig. 3) normalized by the a posteriori standard deviations **(b, e)**, and locations of significant sources and sinks at $1\sigma_{\hat{s}}$ and $2\sigma_{\hat{s}}$ **(c, f)**.

Fig. 9. Estimated fluxes from inversions with no transport error using RML-Inv covariance parameters, aggregated to monthly average ecoregion (Fig. 4) and continental scales. Error bars represent 95% uncertainty bounds. Estimated fluxes from inversions using three different realizations of transport error are also shown for each temporal resolution. Uncertainty bounds are not shown for these inversions for simplicity, but were similar in magnitude to their equivalents with no transport error.

the difference being more pronounced with the RML-Inv parameters due to the improvement in quality of the covariance optimization for these cases. In contrast, for the F3hr case, simulated transport error has a minimal impact on the grid-scale RMSE. While these results are promising, it is not clear whether this result would hold true with more realistic systematic, non-random transport errors.

The fraction of true fluxes lying within two standard deviations (Eq. 6) of the estimated fluxes is presented at the monthly timescale for the inversions with no transport error in Table 5, for both covariance parameter optimization methods, and with and without temporal covariance included in **Q**. The inclusion of temporal covariance is found to be important for obtaining accurate a posteriori uncertainties when using the reference covariance parameters for all flux resolutions. The same result holds using the RML-Inv parameters with the 3-hourly flux resolution, whereas for the F4d-diurnal and F4d cases, accurate grid-scale uncertainties can be obtained even with spatial-only covariance if RML-Inv parameters are used, due to the large estimated flux variance parameter in **Q**. Results with simulated transport error are not shown in Table 5, although the same conclusions hold. Overall, these results highlight that accounting for both spatial and temporal flux covariance yields accurate a posteriori uncertainty bounds at the grid-scale much more reliably than accounting for only spatial correlations.

4.3 Inferred grid-scale spatial patterns

Grid-scale maps of monthly-averaged flux estimates are shown in Figs. 6, 7 and 8 for inversions performed with the three flux resolutions and two sets of covariance parameters, without any simulated random transport error. For comparison, the true monthly-averaged fluxes are shown in Fig. 3. Overall, the inversions detect correct large-scale patterns of sources and sinks, with significant sinks recovered from the eastern United States to northwest Canada and Alaska. As expected, the inferred fluxes show significantly less overall variability relative to the true fluxes, due to the sparse atmospheric network and the absence of auxiliary environmental variables within the trend (e.g. Gourdji et al., 2008).

The F4d inversions (Fig. 6) perform least well in capturing the true grid-scale spatial patterns. Fluxes estimated using the reference covariance parameters remain close to their mean value for the continent with little spatial variability. In addition, high normalized errors exist near the WKT, ARM and SBL measurement towers, most likely due to temporal aggregation errors near the sampling locations. High errors are also seen in northwest Canada, where the strong sinks in this region fall outside of the areas well-constrained by the atmospheric measurements (Fig. 2). With the RML-Inv parameters, the opposite problem occurs such that there is unrealistic spatial variability in the recovered fluxes associated with the artificially high flux variance parameter in **Q**. These results are consistent with those presented in Sect. 3.3, confirming that estimating fluxes directly at highly aggregated temporal scales is not an optimal setup for regional inversions.

The F4d-diurnal inversions (Fig. 7) show more realistic spatial variability in the fluxes. Sources are now properly recovered in Central America, while normalized errors are reduced in all areas. Also, fluxes recovered using both sets of

covariance parameters yield consistent results with reasonable grid-scale spatial patterns. This indicates that, although the covariance parameters for the F4d-diurnal case inferred using the atmospheric data (RML-Inv) differed from the reference covariance parameters in some cases (Tables 3 and 4), the RML-Inv parameters can still be used to recover fluxes of comparable quality to those obtained using idealized covariance parameters. The a posteriori uncertainties are more affected by the use of RML-Inv parameters than the fluxes, as reflected in the lower normalized errors and fewer significant sources and sinks relative to the inversion using the reference parameters.

The F3hr inversions (Fig. 8) also yield realistic spatial variability, especially when using the reference parameters. With the RML-Inv parameters, the spatial variability is slightly reduced, although as with the F4d-diurnal case, inferred fluxes are similar using the two sets of parameters. Also, given the more realistic recovered uncertainties with this resolution as compared to the other cases (Table 5), the largest number of significant sources and sinks are recovered at the grid-scale with the RML-Inv parameters for this case.

Overall, this comparison of the grid-scale flux maps demonstrates that an inversion that directly estimates the diurnal cycle of the fluxes (i.e. F4d-diurnal and F3hr) can recover reasonably accurate grid-scale spatial patterns across the continent using only a 9-tower measurement network. Also, as previously shown in Fig. 5, the quality of inferred fluxes is preserved even when covariance parameters are estimated from the available atmospheric data, and not assumed known a priori. Inversions that estimate fluxes at coarser timescales (i.e. F4d) that average out the diurnal cycle do not perform nearly as well, consistent with the covariance parameter conclusions presented in Sect. 3.3.

4.4 Results at ecoregion scale

Figure 9 presents estimated fluxes and their uncertainties from the inversions using covariance parameters inferred with the atmospheric observations (RML-Inv), aggregated a posteriori to the monthly-average ecoregion (Fig. 4) and continental scales. The RML-Inv parameters were used, because this is most consistent with what would be possible in future real-data inversions. Also, results using the reference covariance parameters are very similar at this aggregated scale to those presented in Fig. 9.

At the ecoregion scale, the inversions resolving the diurnal cycle (F4d-diurnal and F3hr) are seen to yield more accurate fluxes, which is also confirmed by the RMSE's at this aggregated scale (0.03, 0.04 and 0.16 µmol/(m^2 s) for the F3hr, F4d-diurnal, and F4d cases, respectively). The relative performance of inversions using the three temporal resolutions is particularly evident in the better-constrained ecoregions, such as the Temperate Broad & Mixed-leaf Forest (TBMF) and Boreal Forest (Bore), where temporal aggregation error has the larger impact on fluxes due to their proximity to

the towers and strong flux variability. In addition, the F3hr and F4d-diurnal inversions also infer more realistic ecoregion fluxes in the far-field, e.g. in the Tropics (Trop) and the Desert and Xeric Shrubland (DeXS). In contrast, the F4d case yields aggregated fluxes that remain close to the mean continental flux across all ecoregions. This last result is interesting in that high grid-scale spatial variability for this case (seen in Fig. 6) did not translate into large differences in spatially aggregated ecoregion scale fluxes.

For the F3hr and F4d-diurnal cases, the 95% uncertainty bounds capture the true flux for all or most ecoregions. Also, an analysis of inversions for these cases that were performed with and without accounting for temporal covariance between fluxes (results not shown) confirms that accounting for this correlation a priori is necessary for recovering accurate uncertainty bounds at the monthly ecoregion scale. In contrast, the quality of inferred fluxes and uncertainties is lower for the F4d case regardless of spatial-temporal covariance assumptions, with only three of seven aggregated fluxes falling within two standard deviations of the true ecoregion flux in Fig. 9. Overall, these results show that realistic a posteriori uncertainties can be recovered by the inversion at aggregated ecoregion scales, in addition to grid-scales (Table 5), as long as the diurnal cycle is estimated in the inferred fluxes, and both spatial and temporal correlation are considered in the a priori flux covariance matrix.

As expected, the addition of transport error degrades the quality of inferred ecoregion scale fluxes for all temporal resolutions, with the inferred values being closer to the mean continental flux, as seen in Fig. 9. The impact is greater for inversions that resolve the diurnal cycle (F3hr and F4d-diurnal), perhaps because these cases were originally better able to resolve ecoregion-scale variability in a setup without transport model error. In addition, for under-constrained areas (e.g. Tropics (Trop) and Desert and Xeric Shrubland (DeXS)), random noise may obscure the diffuse signal from these areas more than in the near-field of the tower locations. Despite the fact that random transport errors are shown here to degrade inversion quality, uncertainty bounds for the cases resolving the diurnal cycle are still realistic for most ecoregions (results not shown).

At the North American continental scale, inferred fluxes from all temporal resolutions, with and without transport error, are reasonably close to the true net flux (a difference of $\leq$0.25 μmol/(m^2 s)). Consistent with previous results, the F4d case performs the least well, with the continental sink being statistically significantly different from the true value. However, for all cases, as demonstrated by the narrower uncertainty bounds at the continental scale, results confirm that fluxes can be inferred more precisely at the continental scale than at smaller spatial scales.

Overall, the ecoregion-scale results confirm the importance of estimating the diurnal cycle of fluxes directly. This is particularly true in the near-field of the tower locations due to temporal aggregation errors, but it also appears to help con-

strain fluxes in the far-field. Random transport errors degrade the ecoregion-scale flux signal towards the continental mean, although their impact may potentially decrease as more areas become better-constrained by a growing measurement network.

4.5 Sensitivity tests with night-time data

In order to investigate the potential value of including night-time data for both Short and Tall/MBL towers, additional sensitivity tests were performed using the F4d-diurnal case. This resolution was chosen because of its comparable quality to the F3hr case with lower computational cost, which makes it most promising for annual or multi-year inversions. Inversions using reference covariance parameters and no transport error were used for these tests in order to isolate the impact of temporal aggregation errors associated with night-time measurements.

Results show that the continental grid-scale RMSE changes only marginally when night-time data are eliminated for the Tall/MBL towers (0.62 vs. 0.63 μmol/(m^2 s)). However, including night-time data for both Short and Tall/MBL towers substantially increased the RMSE to 0.71 μmol/(m^2 s), most likely due to the temporal aggregation errors associated with the smaller footprints for night-time measurements at the Short towers. An analysis of results at the aggregated ecoregion scale showed that the setup used for most inversions in the current study, using night-time data only for the Tall/MBL towers, minimizes ecoregion scale RMSE's. This setup eliminates the temporal aggregation errors associated with including night-time data at the Short towers, while also allowing for a stronger constraint on far-field fluxes through night-time data from the Tall/ MBL towers. Again, in a real-data environment, the value of using night-time measurements from the Tall/MBL towers may be reduced for nights when the towers are within the PBL, and the PBL height is consistently over or under-estimated.

5 Summary and conclusions

This study evaluated the constraint on CO$_2$ fluxes provided by atmospheric data from nine continuous measurement locations across the North American continent, within the context of a regional geostatistical inversion without additional auxiliary variables. Estimating fluxes at a temporal resolution that can directly estimate the diurnal variability (F4d-diurnal and F3hr cases) was found to be crucial both for recovering covariance parameters directly from the atmospheric data, and for inferring ecoregion-scale fluxes that were statistically consistent with the "true" fluxes. Accounting a priori for both spatial and temporal covariance in the flux distribution was also found to be necessary for recovering accurate a posteriori uncertainty bounds on the estimated fluxes.

The poor performance of inversions that did not estimate the diurnal cycle (i.e. F4d) were due to the high temporal aggregation errors associated with not being able to adjust the strong diurnal and synoptic variability of the fluxes, particularly near the measurement locations. For time periods outside of the growing season, the impact of temporal aggregation error may be lower, because fluxes are expected to be less variable, although the ability to infer accurate fluxes during the growing season is necessary for inferring accurate annual or multi-year carbon budgets. Also, while temporal aggregation errors may be of particular concern for geostatistical inversions because they do not assume the shape of the diurnal cycle a priori, any errors in the diurnal cycle in prior flux estimates in synthesis Bayesian inversions would also yield temporal aggregation errors. This is likely to be of at least some concern, given the differences observed in the diurnal cycles predicted by different biospheric models. This impact is the subject of ongoing work.

In terms of the two flux resolutions resolving the diurnal cycle (i.e. F3hr and F4d-diurnal), both cases yielded flux estimates of comparable quality, despite the fact that covariance parameters estimated with the atmospheric data were more consistent with reference values for the F3hr relative to the F4d-diurnal case. In fact, for both of these flux resolutions (with no transport model error) at the ecoregion scale, the two standard deviation uncertainty bounds captured the true flux for all or almost all eco-regions. This is an encouraging result, showing that data from only nine measurement towers sparsely located across the continent can be used to constrain ecoregion-scale fluxes without additional auxiliary information from remote-sensing datasets or biospheric models, as long as the diurnal cycle is resolved in the fluxes. Between the two cases, the F4d-diurnal case has the additional advantage of estimating four times fewer fluxes than the F3hr case, yielding substantial computational savings that will become important for longer-term inversions. Therefore, given the comparable quality of inversion results for these two cases, performing regional inversions in a way that resolves a multi-day average diurnal cycle appears to be the most promising avenue for regional inversions using real atmospheric data.

In this study, simulated random transport error was shown to decrease the quality of flux estimates in under-constrained areas at the ecoregion scale. It is important to note that non-random transport errors due to phasing or systematic biases, that may be more typical of the real inaccuracies in existing atmospheric transport models, may have more of an impact on flux estimation in the near-field where small errors in the transport matrices (**H**) could translate into large differences in inferred fluxes. The impact of transport errors on regional inversions has been investigated more thoroughly in other studies, e.g. Law et al. (2003), and should also be explored further in future studies focusing on North America.

Finally, it is important to note that real-data inversions are subject to additional complications in flux interpretation relative to synthetic data studies, due to potential biases introduced by the boundary conditions, non-random transport errors, and aggregation errors from flux variability at scales finer than the scale of the estimation grid. Overall, however, synthetic data experiments provide a baseline for the best achievable performance of real-data inversions, and help to highlight the impact of setup choices that may be obscured by the additional complexity associated with using real data.

In summary, synthetic data experiments were shown in this work to help illuminate the constraint on fluxes achieved by various regional inversion setup choices. The results suggest that even a fairly sparse network of continuous CO$_2$ measurements, used with no auxiliary information or prior estimates of flux variability in time or space, can be used to infer accurate monthly ecoregion scale CO$_2$ surface fluxes over North America, as long as the diurnal cycle is resolved in the estimated fluxes and both a priori spatial and temporal flux covariances are considered. Statistically significant sinks can also be recovered at the grid-scale, although uncertainties remain high at this fine spatial scale. The incorporation of additional atmospheric data and auxiliary variables in future real data geostatistical inversions can only help to further improve the recovery of fluxes at finer spatial resolutions.

References

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., and Zhu, Z.: TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO$_2$ fluxes, 1988–2003, Global Biogeochem. Cycles, 20, GB1002, doi:10.1029/2004GB002439, 2006.

Bakwin, P. S., Tans, P. P., Hurst, D., and Zhao, C.: Measurements of carbon dioxide on very tall towers: Results of the NOAA/CMDL program, Tellus, 50B, 401–415, 1998.

Carouge, C., Bousquet, P., Peylin, P., Rayner, P. J., and Ciais, P.: What can we learn from European continuous atmospheric CO2 measurements to quantify regional fluxes – Part 1: Potential of the 2001 network, Atmos. Chem. Phys., 10, 3107–3117, doi:10.5194/acp-10-3107-2010, 2010a.

Carouge, C., Rayner, P. J., Peylin, P., Bousquet, P., Chevallier, F., and Ciais, P.: What can we learn from European continuous atmospheric CO2 measurements to quantify regional fluxes – Part 2: Sensitivity of flux accuracy to inverse setup, Atmos. Chem. Phys., 10, 3119–3129, doi:10.5194/acp-10-3119-2010, 2010b.

Cook, R., Huntzinger, D. N., Post, M., Jacobson, A., Wei, Y., and NACP Interim Synthesis Participants: Preliminary results of the NACP Regional Interim Synthesis, presentation at AmeriFlux Annual Meeting, Washington, D.C., Sep. 22, 2009, available at http://nacp.ornl.gov/int_synth_contreg.shtml, 2009.

Geels, C., Doney, S. C., Dargaville, R., Brandt, J., and Christensen, J. H.: Investigating the sources of synoptic variability in atmospheric CO$_2$ measurements over the Northern Hemisphere continents: a regional model study, Tellus, 56B, 35–50, 2004.

Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P., Vermeulen, A. T., Dargaville, R., Aalto, T., Brandt, J., Christensen, J. H., Frohn, L. M., Haszpra, L., Karstens, U., Rödenbeck, C., Ramonet, M., Carboni, G., and Santaguida, R.: Comparing atmospheric transport models for future regional inversions over Europe – Part 1: mapping the atmospheric CO2 signals, Atmos. Chem. Phys., 7, 3461–3479, doi:10.5194/acp-7-3461-2007, 2007.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO$_2$ with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, J. Geophys. Res., 108 (D24), 4757, doi:10.1029/2003JD003770, 2003.

Gerbig, C., Lin, J. C., Munger, J. W., and Wofsy, S. C.: What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes?, Atmos. Chem. Phys., 6, 539–554, doi:10.5194/acp-6-539-2006, 2006.

Gerbig, C., Körner, S., and Lin, J. C.: Vertical mixing in atmospheric tracer transport models: error characterization and propagation, Atmos. Chem. Phys., 8, 591–602, doi:10.5194/acp-8-591-2008, 2008.

Gerbig, C., Dolman, A. J., and Heimann, M.: On observational and modelling strategies targeted at regional carbon exchange over continents, Biogeosciences, 6, 1949–1959, doi:10.5194/bg-6-1949-2009, 2009.

Göckede, M., Turner, D. P., Michalak, A. M., Vickers, D., and Law, B. E.: Sensitivity of a sub-regional scale atmospheric inverse

CO$_2$ modeling framework to boundary conditions, J. Geophys. Res.-Biogeosciences, in review, 2010.

Gourdji, S. M., Mueller, K. L., Schaefer, K., and Michalak, A. M.: Global monthly-averaged CO$_2$ fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, J. Geophys. Res., 113, D21115, doi:10.1029/2007JD009733, 2008.

Gurney, K., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: Towards robust regional estimates of CO$_2$ sources and sinks using atmospheric transport models, Nature, 415(6872), 626–630, 2002a.

Gurney, K., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Kowalczyk, E., Maki, T., Maksyutov, S., Peylin, P., Prather, M., Pak, B. C., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: TransCom 3 CO$_2$ inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information, Tellus, 55B, 555–579, 2002b.

Haszpra, L.: On the representativeness of carbon dioxide measurements, J. Geophys. Res., 104(D21), 26953–26960, 1999.

Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, J. Geophys. Res., 106, 4703–4715, 2001.

Kitanidis, P.: Quasi-linear geostatistical theory for inversing, Water Resour. Res., 31(10), 2411–2419, 1995.

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K. J., Ciais, P., Denning, A. S., and Rayner, P. J.: Mesoscale inversion: first results from the CERES campaign with synthetic data, Atmos. Chem. Phys., 8, 3459–3471, doi:10.5194/acp-8-3459-2008, 2008.

Law, R. M., Rayner, P. J., Steele, L. P., and Enting, I. G.: Using high temporal frequency data for CO$_2$ inversions, Global Biogeochem. Cycles, 4, 1053, doi:10.1029/2001GB001593, 2002.

Law, R. M., Rayner, P. J., Steele, L. P., and Enting, I. G.: Data and modelling requirements for CO$_2$ inversions using high frequency data, Tellus, 55B, 512–521, doi:10.1034/j.1600-0560.2003.0029.x, 2003.

Law, R. M., Rayner, P. J., and Wang, Y. P.: Inversion of diurnally-varying synthetic CO$_2$: network optimisation for an Australian test case, Global Biogeochem. Cycles, 18, GB1044, doi:10.1029/2003GB002136, 2004.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: the Stochastic Time-Inverted Lagrangian Transport (STILT) model, J. Geophys. Res., 108(D16), 4493, doi:10.1029/2002JD003161, 2003.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Grainger, C. A., Stephens, B. B., Bakwin, P. S., and Hollinger, D. Y.: Measuring fluxes of trace gases at regional scales by Lagrangian observations: application to the CO$_2$ Budget and Rectification Airborne (COBRA) study, J. Geophys. Res., 109, D15304, doi:10.1029/2004JD004754, 2004.

Matross, D. M., Andrews, A., Pathmathevan, M., Gerbig, C., Lin, J.

C., Wofsy, S. C., Daube, B. C., Gottlieb, E. W., Lee, J. T., Zhao, C., Bakwin, P. S., Munger, J. W., and Hollinger, D. Y.: Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based, and satellite data, Tellus, 58(5), 344–358, 2006.

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res., 109(D14), D14109, doi:10.1029/2003JD004422, 2004.

Mueller, K. L., Gourdji, S. M., and Michalak, A. M.: Global monthly-averaged CO$_2$ fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements, J. Geophys. Res., 113, D21114, doi:10.1029/2007JD009734, 2008.

Olsen, J. C. and Randerson, J. T.: Differences between surface and column atmospheric CO$_2$ and implications for carbon cycle research, J. Geophys. Res.-Atmos., 109, D02301, doi:10.1029/2003JD003968, 2004.

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., and Kassem, K. R.: Terrestrial ecoregions of the world: a new map of life on earth, Bioscience, 51(11), 933–938, 2001.

Patra, P. K., Law, R. M., Peters, W., Rodenbeck, C., Takigawa, M., Aulagnier, C., Baker, I., Bergmann, D. J., Bousquet, P., Brandt, J., Bruhwiler, L., Cameron-Smith, P. J., Christensen, J. H., Delage, F., Denning, A. S., Fan, S., Geels, C., Houweling, S., Imasu, R., Karstens, U., Kawa, S. R., Kleist, J., Krol, M. C., Lin, S.-J., Lokupitiya, R., Maki, T., Maksyutov, S., Niwa, Y., Onishi, R., Parazoo, N., Pieterse, G., Rivier, L., Satoh, M., Serrar, S., Taguchi, S., Vautard, R., Vermeulen, A. T., and Zhu, Z.: TransCom model simulationsof hourly atmospheric CO$_2$: Analysis of synoptic-scale variations for the period 2002–2003, Global Biogeochem. Cycles, 22, GB4013, doi:10.1029/2007GB003081, 2008.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, PNAS, 104(48), 18925–18930, 2007.

Peylin, P., Rayner, P. J., Bousquet, P., Carouge, C., Hourdin, F., Heinrich, P., Ciais, P., and AEROCARB contributors: Daily CO$_2$ flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology, Atmos. Chem. Phys., 5, 3173–3186, doi:10.5194/acp-5-3173-2005, 2005.

Randerson, J. T., Thompson, M. V., Conway, T. J., Fung, I. Y., and Field, C. B.: The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide, Global Biogeochem. Cycles, 11(4), 535–560, 1997.

Rodell, M., Houser, P. R., Jambor, U., et al.: The Global Land Data Assimilation System, B. Am. Meteorol. Soc., 85, 381–394, 2004.

Schuh, A. E., Denning, A. S., Uliasz, M., and Corbin, K. D.: Seeing the forest through the trees: Recovering large-scale carbon flux biases in the midst of small-scale variability, J. Geophys. Phys.-Biogeo., 114, G03007, doi:10.1029/2008JG000842, 2009.

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

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 2. Technical Note 468+STR, MMM Division, NCAR, Boulder, CO, 88 pp., available from http://www.mmm.ucar.edu/wrf/users/docs/arw.pdf, 2005.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos. Chem. Phys., 6, 3423–3441, doi:10.5194/acp-6-3423-2006, 2006.