SAND2014-16474J

A statistical approach for isolating fossil fuel emissions in atmospheric inverse problems

Vineet Yadav[1#], **Anna M. Michalak**[1], **Jaideep Ray**[2] **and Yoichi P. Shiga**[3]

[1] Department of Global Ecology, Carnegie Institution for Science, Stanford, California, USA 94305

[2] Sandia National Laboratories, Livermore, California

[3] Department of Civil and Environmental Engineering, Stanford University, Stanford, California

Correspondence to: Vineet Yadav (vineety@stanford.edu)

Abstract

Independent identification and quantification of fossil fuel (FF) emissions constitutes a considerable scientific challenge. By coupling atmospheric observations of CO_2 with models of atmospheric transport, inverse models have the ability to overcome this challenge. However, disaggregating the biospheric and FF flux components from CO_2 concentration measurements has proven to be difficult and only been accomplished through isotopic tracers of fossil fuel emissions. In this study, we propose an alternative statistical inverse modeling scheme for disaggregating fluxes on the basis of novel error covariance models and covariates that are representative of the underlying processes affecting FF and biospheric fluxes. The application of the method is demonstrated with a synthetic and two real data inversions by using in-situ CO_2 measurements over North America. Inversions are performed only for the month of January as the given observation network allows reasonable, regional estimation and disaggregation of fluxes in this month. The quality of disaggregation is assessed primarily through examination of a posteriori covariance between disaggregated FF and biospheric fluxes at regional scales. Findings indicate that the proposed method is able to robustly disaggregate FF from biospheric fluxes at the monthly-regional scale. Error covariance models and covariates based on temporally varying FF inventory data provide a more robust disaggregation over static proxies (e.g. nightlight intensity, population density). However, synthetic data case studies show that disaggregation is still possible without the availability of temporally varying FF inventory data.

1 Introduction

The rising concentration of carbon dioxide (CO_2) in the atmosphere is the main driver of anthropogenic climate change. Spatial and temporal variations in global CO_2 fluxes leading to this increase can be inferred using top-down inverse models from atmospheric observations that reflect the combined influence of fossil fuel (FF), biospheric, and oceanic fluxes. In inverse models, CO_2 concentration measurements are combined with atmospheric transport models driven by observed meteorology to yield estimates of the net exchange of CO_2 at the land and ocean surface (e.g. Gurney et al., 2002; Rayner et al., 1999; Tans et al., 1990; Michalak et al., 2004).

Recently, atmospheric inverse models have been proposed as a potential tool for independent verification of inventory-based estimates of FF fluxes or emissions. Such applications currently do not exist at regional to continental scales, however, due to limitations associated with observational coverage (Pacala et al., 2010). Improvements in terms of increasing in-situ (e.g. Sloop and Novakovskaia, 2012) and satellite measurements (e.g. Duren and Miller, 2012) of CO_2 concentrations and radiocarbon isotope 14C (Miller et al., 2012) have been suggested as options towards reducing this uncertainty. Targeted efforts are ongoing with the focus on constraining the FF fluxes in urban areas (see for e.g. Megacities Carbon Project, Duren and Miller, 2012; Hestia Project, Gurney et al., 2012; Kort et al., 2012; Indianapolis Flux Experiment (INFLUX), http://influx.psu.edu/). These efforts do not scale to national or continental scales, however, especially in developing countries, and can typically only be used to validate FF fluxes obtained from inverse models in small regions, i.e., urban areas. In the next decade, remote sensing of CO_2 will provide a large number of observations (for discussion on CO_2 observations from space see; Crisp et. al., 2004; Olsen et al., 2008) that can reduce the uncertainty of the FF fluxes estimated within an inverse modeling framework.

Beyond an increase in observations, however, methodological improvements are also required in both transport and inverse models to realize the full potential of current and future CO_2 observations. In the case of inverse models, these methodological improvements include designing inverse modeling approaches that leverage the distinct statistical signatures of FF and biospheric fluxes in order to pinpoint their contributions to the total surface flux.

Up to this point, inversion efforts aimed at separating FF emissions from biospheric fluxes have relied on the use of co-emitted or isotopic tracers of FF CO_2 emissions (e.g., Brioude et al., 2012) that can be used to identify the FF contribution to the total CO_2 signal (for details on tracers of FF CO_2 see Miller et al., 2012). However, the spatial coverage of these FF tracer measurements is limited and a large fraction of the variance in FF fluxes remains unexplained by them (Miller et al., 2012).

Just as we utilize a unique chemical signature in a FF tracer to isolate FF emissions from biospheric fluxes, so too can we utilize the unique spatiotemporal features of FF emissions to separate FF emissions from biospheric fluxes. Here we present a geostatistical inverse modeling methodology that does not rely on FF tracers to separate FF and biospheric fluxes. In order to achieve this, we postulate that: (1) the covariates and the error covariance models for FF and biospheric fluxes are distinct, and (2) the covariates and error covariance models can represent the magnitude and spatio-temporal variations in the terrestrial sources (FF and biospheric) and sinks (biospheric) of CO_2 fluxes. The applicability of the proposed method is demonstrated within the context of one synthetic and two real data inversions at 1° by 1° resolution for North America for the month of January 2008. In these inversions, the covariates and the error covariance model for biospheric fluxes are prescribed, whereas they are chosen from a set of candidate covariates and error covariance models for FF fluxes. The quality of the separation of the fluxes is assessed by examining a posteriori cross-covariances between FF and biospheric fluxes, and by comparing the inferred spatial distribution and magnitude of the estimated fluxes with true fluxes in the case of the synthetic data test case.

The study focuses on the month of January, because previous work has shown that this is a month during which the impact of FF emissions is observed even given the limited observational network currently available (Shiga et al., 2014). This is due to the seasonality of the sensitivity of observations to surface fluxes, the smaller interference by the biospheric signal, and the ability to adequately model the atmospheric transport over the domain of interest (Shiga et al., 2014). For more extensive applications, a method such as the one proposed here would need to be coupled with the more extensive observational coverage described above.

2 Method for flux disaggregation

A geostatistical formulation of the atmospheric inverse problem has been used to explore a variety questions related to CO_2 fluxes (e.g. Michalak et al. 2004, Gourdji et al. 2012, Shiga et al. 2014, Fang et al. 2014). Unlike other Bayesian methods, this approach does not obtain its prior mean from a process-based model; instead, it models the prior as a linear combination of a set of covariates with weights that are treated as hyperparameters and estimated as part of the inverse problem. The objective function for the standard geostatistical inverse modeling (GIM) approach can be written as:

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

where $\mathbf{z}$ are measurements of CO_2 concentrations, $\mathbf{H}$ is a Jacobian matrix representing the sensitivity of measurements to underlying flux, $\mathbf{s}$ are the CO_2 fluxes, $\mathbf{R}$ is the model-data mismatch error covariance matrix, $\mathbf{X}$ is a matrix of covariates of $\mathbf{s}$, $\boldsymbol{\beta}$ are the coefficients or weights of individual covariates and $\mathbf{Q}$ is the error covariance matrix (aka prior covariance) describing the deviations of $\mathbf{s}$ from $\mathbf{X}\boldsymbol{\beta}$.

In this study, we modify this objective function to separately account for biospheric and FF fluxes. This modified objective function becomes:

$$\begin{aligned} L_{\mathbf{s}_{bio},\mathbf{s}_{ff},\boldsymbol{\beta}_{bio},\boldsymbol{\beta}_{ff}} = {} & \frac{1}{2}\left(\mathbf{z}-\left[\mathbf{H}_{bio}\mathbf{s}_{bio}+\mathbf{H}_{ff}\mathbf{s}_{ff}\right]\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{z}-\left[\mathbf{H}_{bio}\mathbf{s}_{bio}+\mathbf{H}_{ff}\mathbf{s}_{ff}\right]\right) \\ & + \frac{1}{2}(\mathbf{s}_{bio}-\mathbf{X}_{bio}\boldsymbol{\beta}_{bio})^{\mathrm{T}}\mathbf{Q}_{bio}^{-1}(\mathbf{s}_{bio}-\mathbf{X}_{bio}\boldsymbol{\beta}_{bio}) \\ & + \frac{1}{2}\left(\mathbf{s}_{ff}-\mathbf{X}_{ff}\boldsymbol{\beta}_{ff}\right)^{\mathrm{T}}\mathbf{Q}_{ff}^{-1}\left(\mathbf{s}_{ff}-\mathbf{X}_{ff}\boldsymbol{\beta}_{ff}\right) \end{aligned} \quad (2)$$

where the subscripts *bio* and *ff* represent the biospheric and FF component of the terms defined in equation 1. This modified objective function embodies the assumptions that suitable covariates (in **X)** and error covariance models (**Q**) can be defined to statistically isolate FF and biospheric fluxes. Thus, the covariates ($\mathbf{X}_{bio}$, $\mathbf{X}_{ff}$) and error covariance models ($\mathbf{Q}_{bio}$, $\mathbf{Q}_{ff}$) in equation 2 play a vital role, as they capture our understanding of the processes affecting FF and biospheric fluxes. Overall, a small set of covariates identified using a statistical model selection procedure are used to represent a portion of the variation in estimated fluxes (see, Yadav et al. 2010). The weights ($\boldsymbol{\beta}_{bio}$, $\boldsymbol{\beta}_{ff}$) of these covariates are optimal when the underlying error covariance models are correctly specified (in the context of geostatistics see Stein 1988).

The covariates and error covariance models in section 2.1 and 2.2 are discussed specifically in the context of the three inversion case studies presented in this work. Other covariates and error covariance models could be implemented within equation 2 as needed for other applications.

2.1 Covariates and error covariance model for biospheric fluxes

For the three inversion case studies presented here, the only covariates used for biospheric fluxes in $\mathbf{X}_{bio}$ are indices binning fluxes into eight 3-hourly blocks, where three hours is the temporal resolution at which the biospheric fluxes are estimated (see, section 3 for details on the resolution of inversions). This

choice models the mean diurnal variations in the biospheric fluxes, but any spatiotemporal deviations therefrom must be captured by the error covariance matrix $\mathbf{Q}_{bio}$.

Biospheric fluxes vary relatively smoothly, exhibit spatial autocorrelation, and are largely independent of FF fluxes in North America. Thus, it is assumed for the inversion case studies that the error covariance for biospheric fluxes can be modeled through a stationary (for definition of stationarity see Cressie 1991) spatio-temporal exponential covariance model (see, Gourdji et al 2012). This error covariance model can be written as (for details see e.g., Gourdji et al., 2010; Yadav and Michalak, 2013):

$$\mathbf{Q}_{bio} = \sigma^2 \left[\exp\left(\frac{-\mathbf{h}_{temporal_{bio}}}{l_{temporal_{bio}}} \right) \otimes \exp\left(\frac{-\mathbf{h}_{spatial_{bio}}}{l_{spatial_{bio}}} \right) \right] \quad (3)$$

where σ^2 is the variance in space and time, $\mathbf{h}_{spatial_{bio}}$ and $\mathbf{h}_{temporal_{bio}}$, are the separation distances between estimation locations of biospheric fluxes in space and time, and $l_{temporal_{bio}}$ and $l_{spatial_{bio}}$ are the spatial and temporal correlation range parameters and $\otimes$ denotes the Kronecker product.

2.2 Covariates and error covariance model for fossil fuel fluxes

To aid in the disaggregation of FF fluxes from the biospheric fluxes, we include covariates that are correlated with FF fluxes in $\mathbf{X}_{ff}$. As there are many covariates that are correlated with FF fluxes, a set of covariates that best describe FF fluxes are chosen from a superset of candidate covariates on the basis of the Bayesian Information Criterion (BIC; Schwarz, 1978) as described in section 2.3. For the inversions presented here, the superset of candidate covariates of FF fluxes includes (1) the radiance intensity of night lights for 2008 (Elvidge et al., 1997), (2) population density per sq. km for 2008 (SEDAC, 2004), (3) % built up area (Miteva, 2002), (4) % urban area for 2009 (Schneider et al., 2009), and (5) a mixed, scaled estimate of FF fluxes of North America for 2008 from Vulcan and ODIAC (see Section 3.1).

Any spatiotemporal deviations from $\mathbf{X}_{ff}\boldsymbol{\beta}_{ff}$ are assumed to be independent, and can thus be represented through a diagonal error covariance matrix with a different variance for each spatial location (i.e., each grid-cell). This is consistent with the fact that fluxes tend to be spatially localized, at least at the spatial resolution (1° by 1°) at which FF fluxes are estimated in this study (see section 3 for details on spatial resolution of inversions).

The FF error covariance is thus defined here as:

$$\mathbf{Q}_{ff} = \left(a \begin{bmatrix} k_1 & 0 & 0 \\ 0 & . & 0 \\ 0 & 0 & k_r \end{bmatrix} + b \begin{bmatrix} 1 & 0 & 0 \\ 0 & . & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \quad (4)$$

where b is a constant variance component for all locations and times, and a and $k_1 \ldots . k_r$ (k's) define additional error variance that are spatially dependent.

We assume that the k's in equation 3 can be prescribed based on geospatial datasets related to FF fluxes, ten of which are considered here. The first nine are the mean, maximum and variance of night lights, population density, and % built up area for each 1^o x 1^o grid-cell in the inversion domain. These can be defined because all three of these datasets are available at higher resolution than the resolution of the inversions. The final considered dataset is a FF inventory (Vulcan combined with ODIAC; see section 4) at the resolution of the inversion, with this final dataset being temporally, as well as spatially, variable.

A single one of these datasets is selected to populate the k's, and thereby to define $\mathbf{Q}_{ff}$, using BIC (see Section 2.3). The primary objective is to obtain an optimal model that, in combination with covariates in $\mathbf{X}_{ff}$, can explain the spatiotemporal variability of FF fluxes.

2.3 Covariate and Covariance Selection from Bayesian Information Criterion

BIC evaluates the tradeoff between the explanatory power and the complexity of a statistical model. It is used for selecting an appropriate set of covariates from a superset of candidate covariates of the dependent variable. The set of covariates that forms the model with the lowest BIC value, i.e., the "best" model, optimally balances explanatory power with model complexity. Here, BIC is used to select covariates for both $\mathbf{X}_{ff}$ and $\mathbf{Q}_{ff}$. BIC is defined as:

$$BIC = \underbrace{RSS + \ln|\mathbf{\Psi}|}_{log\ likelihood} + \underbrace{p \ln(n)}_{penalty\ term} \tag{5}$$

where | | denotes the matrix determinant, p are the number of parameters or covariates in the model and n is the number of observations.

RSS in equation 5 is defined as ,

$$RSS = \left[\mathbf{z}^T \left(\mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1}\mathbf{\Omega}\left(\mathbf{\Omega}^T\mathbf{\Psi}^{-1}\mathbf{\Omega}\right)^{-1}\mathbf{\Omega}^T\mathbf{\Psi}^{-1}\right)\mathbf{z}\right] \tag{6}$$

where

$$\mathbf{\Psi} = [\mathbf{H}_{bio} \quad \mathbf{H}_{ff}]\begin{bmatrix}\mathbf{Q}_{bio} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{ff}\end{bmatrix}[\mathbf{H}_{bio} \quad \mathbf{H}_{ff}]^T + \mathbf{R} \tag{7}$$

and

$$\mathbf{\Omega} = [\mathbf{H}_{\boldsymbol{bio}} \quad \mathbf{H}_{\boldsymbol{ff}}]\begin{bmatrix}\mathbf{X}_{\boldsymbol{bio}} \\ \mathbf{X}_{\boldsymbol{ff}}\end{bmatrix} \tag{8}$$

Note that BIC (eq. 5) depends on the covariance parameters in $\mathbf{Q}_{ff}$, $\mathbf{Q}_{bio}$ and $\mathbf{R}$, which themselves depend on the covariates used to define $\mathbf{X}_{ff}$ and $\mathbf{Q}_{ff}$. The covariates and covariance parameters must therefore be adjusted in tandem to identify the overall best statistical model.. We proceed as follows:

(1) Adopt one of the ten covariates considered for populating the FF error covariance model ($\mathbf{Q}_{ff}$, eq. 4).
(2) Use the discrete optimization branch and bound algorithm (see; Yadav et al., 2013) and Restricted Maximum Likelihood (RML; for details see; Kitanidis 1995) to simultaneously select covariates for inclusion in $\mathbf{X}_{ff}$ that minimize BIC (eq. 5), and optimize the corresponding covariance parameters for $\mathbf{Q}_{bio}$, $\mathbf{R}$ and $\mathbf{Q}_{ff}$.
(3) Repeat steps 1 and 2 for each of the ten FF error covariance model covariates.
(4) Compare BIC obtained in step 2 for all ten FF error covariance covariates and select $\mathbf{X}_{ff}$, $\mathbf{Q}_{bio}$, $\mathbf{R}$ and $\mathbf{Q}_{ff}$ that results in the minimum overall BIC.

2.4 Flux and a posteriori covariance estimation

The FF and biospheric fluxes are estimated by solving a linear system of equations (e.g. Michalak et al. (2004)):

$$\begin{bmatrix}\mathbf{\Psi} & \mathbf{\Omega} \\ \mathbf{\Omega}^T & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{\Lambda}^T \\ \mathbf{M}\end{bmatrix} = \begin{bmatrix}[\mathbf{H}_{bio} \quad \mathbf{H}_{ff}]\begin{bmatrix}\mathbf{Q}_{bio} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{ff}\end{bmatrix} \\ \begin{bmatrix}\mathbf{X}_{bio} \\ \mathbf{X}_{ff}\end{bmatrix}\end{bmatrix} \tag{9}$$

$$\begin{bmatrix}\hat{\mathbf{s}}_{bio} \\ \hat{\mathbf{s}}_{ff}\end{bmatrix} = [\mathbf{\Lambda}_{bio} \quad \mathbf{\Lambda}_{ff}]\mathbf{z} \tag{10}$$

$$\mathbf{V} = \begin{bmatrix} \boldsymbol{V}_{bio} & \boldsymbol{V}_{ff,bio}^{T} \\ \boldsymbol{V}_{ff,bio} & \boldsymbol{V}_{ff} \end{bmatrix} = -\begin{bmatrix} \mathbf{X}_{bio} \\ \mathbf{X}_{ff} \end{bmatrix} \mathbf{M} + \begin{bmatrix} \mathbf{Q}_{bio} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{ff} \end{bmatrix} - \begin{bmatrix} \mathbf{Q}_{bio} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{ff} \end{bmatrix} [\mathbf{H}_{bio} \quad \mathbf{H}_{ff}]^{T} \boldsymbol{\Lambda}^{T} \qquad (11)$$

where $\mathbf{V}$ is the a posteriori covariance of the estimated fluxes $\hat{\mathbf{s}}_{bio}$ and $\hat{\mathbf{s}}_{ff}$, $\boldsymbol{\Lambda}$ is the matrix of weights, $\mathbf{M}$ are lagrange multipliers, and the remaining terms are as defined earlier. $\boldsymbol{\Lambda}$ and $\mathbf{M}$ are obtained by solving the system of equations in eq. 9.

The joint inversion scheme can result in negative FF fluxes, thus non-negativity constraints are imposed on $\hat{\mathbf{s}}_{ff}$. There are several methods for imposing non-negativity constraints on $\hat{\mathbf{s}}_{ff}$ (e.g. Miller et al. 2014), and a deterministic optimization approach based on Lagrange multipliers (for details, see; Gockede 2010; Miller et. al. 2014) is used here.

Note that non-negativity constraints are imposed on FF fluxes obtained from an unconstrained inversion that utilizes the covariance model and covariates selected as described in section 2.3. Since $\hat{\mathbf{s}}_{bio}$and $\hat{\mathbf{s}}_{ff}$ are updated when imposing the non-negativity constraints, the uncertainty associated with the FF and biospheric fluxes given in section 4 are those obtained from the unconstrained inversion before non-negativity constraints are imposed on $\hat{\mathbf{s}}_{ff}$. Due to this reason and assumptions of normality, the uncertainty bounds on the FF fluxes can have negative values.

3 Inversion Case Studies

Three inversion case studies are used to evaluate the proposed approach. All involve estimating biospheric fluxes at 3-hourly temporal resolution to avoid temporal aggregation errors (for details; see Gourdji et al, 2010), while FF fluxes are estimated at 8-day temporal resolution, in part due to the computational cost of imposing non-negativity constraints. Spatially, both FF and biospheric fluxes are estimated at 1° by 1° for the land area between 10° N to 70° N and 50° W to 170° W. All inversions are conducted for January 2008.

3.1 Data for inversion case studies

The sensitivity matrix (**H**) of the CO_2 observations to surface fluxes for inversions was obtained from the Stochastic Time-Inverted Lagrangian Transport (STILT; Lin et al., 2003) model (for details see; Gourdji et al., 2012, Shiga et al. 2014). Measurements of CO_2 concentrations from 35 in-situ towers in North America were used in the two real data inversion case studies. The names and locations of these towers, and measurement time of the CO_2 observations, are given in Appendix 1.

In the synthetic data inversion, the "ground-truth" for biospheric fluxes was obtained from the Carnegie Ames Stanford Approach (CASA) model as configured for the Global Fire Emissions Database (GFED) v2 project (Randerson et al., 1997;van der Werf et al., 2006) and run within the ambit of North American Carbon Program Regional Interim Synthesis (for details see; Huntzinger et al. 2012). The estimates for FF fluxes were obtained from the Vulcan (USA; 2002) and ODIAC (Canada, Mexico and Alaska; 2007) inventories (Gurney et al., 2009; Oda and Maksyutov, 2011). These were then scaled to 2008 to account for changes in the FF fluxes from those reported in these inventories. Since the CASA-GFED v2 biospheric fluxes were available only at monthly scale they were downscaled to 3-hourly temporal resolution by using net shortwave radiation and near-surface temperature data from the NASA Global Land Data Assimilation System (Olsen and Randerson, 2004; GLDAS; Rodell et al., 2004).

3.2 Real Data Case Studies

The real data case studies were designed to test the influence of a FF inventory in explaining variations in inferred FF fluxes and disaggregating them from biospheric fluxes. This is achieved by examining a

posteriori cross-covariances and results of model selection. Thus, in one case study, the model selection scheme (see; section 2.4) is allowed to select covariates and an error covariance model for FF fluxes from the full superset given in section 2.2 (henceforth, RD1), whereas in the second case study this superset excludes covariate and error covariance model based on FF inventory (henceforth, RD2). This distinction was made to explore the addition error/uncertainty incurred due to the lack of a detailed inventory datasets, which is a realistic constraint in many parts of the world.

3.3 Synthetic Data Case Study

The main goal of the synthetic data case study (henceforth, SD) was to evaluate the performance of the inversion method when true fluxes are known. Its results provide a two way indication of the performance of the proposed method in disaggregating fluxes, that is (1) through analysis of a posteriori cross-covariance between FF and biospheric fluxes, and (2) through comparison of the estimated fluxes with true fluxes (see section 4.3). Overall, this case study is similar to the RD2, as FF inventory estimates are not used as candidate covariates in $\mathbf{X}_{ff}$ or $\mathbf{Q}_{ff}$. This is because in this case the synthetic CO_2 observations are themselves generated using an inventory datasets, and using this same dataset in the inversion would have provided an unrealistic amount of information to the inversion. The synthetic CO_2 observations are corrupted with a zero-mean Gaussian white noise with variances equal to those in the model-data mismatch matrix ($\mathbf{R}$) in RD2. $\mathbf{R}$ in SD is fixed to equal that in RD2, whereas the $\mathbf{Q}_{ff}$ and $\mathbf{Q}_{bio}$ covariance parameters and covariates are obtained from the procedure described in section 2.3. The quality of disaggregation is examined by comparing the inferred fluxes with the true fluxes i.e., CASA-GFED v2 biospheric and Vulcan and ODIAC FF fluxes.

3.4 Framework for Evaluating Case Studies

The Frobenius norm (for description see; Golub and Van Loan, 2012) of FF and biospheric a posteriori cross-covariances is computed to check for the quality of the separation of the estimated fluxes. To compute the Forbenius norm of cross-covariances, firstly a posteriori covariance from inversions are aggregated to monthly resolution at grid-scale. This covariance can be given as:

$$\overline{\mathbf{V}} = \begin{bmatrix} \overline{\boldsymbol{V}}_{bio} & \overline{\boldsymbol{V}}_{ff,bio}^{T} \\ \overline{\boldsymbol{V}}_{ff,bio} & \overline{\boldsymbol{V}}_{ff} \end{bmatrix} \tag{12}$$

where $\overline{\mathbf{V}}$ is the a posteriori covariance of the fluxes aggregated to monthly temporal resolution, $\overline{\boldsymbol{V}}_{bio}$ and $\overline{\boldsymbol{V}}_{ff}$ are the posterior covariances of the biospheric and FF fluxes at monthly resolution, respectively, and $\overline{\boldsymbol{V}}_{ff,bio}$ represents their cross-covariance. The Forbenius norm is computed for the $\overline{\boldsymbol{V}}_{ff,bio}$ as:

$$\left\|\overline{\boldsymbol{V}}_{ff,bio}\right\|_{F} = \sqrt{Trace\left(\overline{\boldsymbol{V}}_{ff,bio}^{T}\overline{\boldsymbol{V}}_{ff,bio}\right)} \tag{13}$$

where $\|\ \ \|$ stands for the norm, and all other terms are as defined earlier. A smaller Frobenius norm of $\overline{\boldsymbol{V}}_{ff,bio}$ indicates better separation of the two signals and lower covariance between the disaggregated fluxes; whereas the opposite is implied by a larger Frobenius norm.

4 Results and Discussion

As described in the introduction, the quantification of fossil fuel emissions from atmospheric observations depends on two types of advances: the first is the availability of an observational network that is sufficiently sensitive to FF emissions, and the second is a methodological framework for isolating the biospheric and FF components of terrestrial fluxes. An approach for fulfilling the second of these needs

is presented here. The presented approach is evaluated with case studies in January, because this is a month during which even the limited existing observational network is relatively conducive to detecting FF emissions (Shiga et al. 2014). Similarly, the approach is evaluated within four regions of the United States (Figure 1), again because these are regions for which the existing observational network is relatively more effective at detecting FF emissions (Shiga et al. 2014).

For the RD1 case study, the fossil fuel inventory is selected both as the spatial trend of the FF emissions ($\mathbf{X}_{ff}$), and as the dataset used to populate the error covariance matrix ($\mathbf{Q}_{ff}$). Intuitively, in the context of the inversion case studies, the choice of a FF trend and error covariance model by BIC implies that among all candidate models it is best suited for: (1) describing the variance in the spatial distribution of FF emissions, (2) identifying the FF signal in the CO_2 observations, (3) separating FF and biospheric fluxes, and (4) computing estimates of FF and biospheric fluxes. The selection of the FF inventory in the RD1 case is expected, because this inventory is indeed expected to be more representative of the true FF emissions patterns relative to the other candidate variables.

Results from RD1 confirm that the statistical framework presented here can be used to disaggregate biospheric and FF terrestrial CO_2 fluxes. The success of the disaggregation of FF and biospheric fluxes in RD1 can be evaluated by examining the posterior cross-covariance and cross-correlation (Figure 2; also see Appendix 2) between these flux estimates at a-posteriori-aggregated spatial (i.e. regional) and temporal (i.e. monthly) scales. The cross-covariances are generally small relative to the magnitude of the fluxes (Figure 3), and the cross-correlations are low, except for the Midwest. Consistently, the Forbenius norm (equation 13, Table 1) was > 15 times lower than the inversions conducted for the summer month of July (results not shown or discussed), indicating small cross-covariances in the FF and biospheric flux uncertainties at the grid scale as well.

For the RD2 case study, the fossil fuel inventory is made unavailable for the variable selection for both the trend ($\mathbf{X}_{ff}$) and prior error covariance ($\mathbf{Q}_{ff}$) models, and population density and % urban land cover ($\mathbf{X}_{ff}$) and the maximum value of night lights intensity within each 1° by 1° inversion gridcell ($\mathbf{Q}_{ff}$) are selected as alternatives (Table 1). The impact of using these datasets, which are less directly representative of the underlying FF emissions, is seen via increased cross-covariances (Figure 2) and cross-correlations in the monthly regional posterior uncertainties of the biospheric and FF fluxes. The Forbenius norm as in RD1 is low (Table 1) and the RD2 flux estimates (Figure 3 and 4) confirm the impact of using the alternate covariates, showing similar uncertainties on the total flux relative to RD1, but increased uncertainties on the contributions from FF and biospheric fluxes. This case study illustrates the impact of the statistical model on the ability to isolate the FF signal.

For the SD case study, the fossil fuel inventory is also made unavailable for the variable selection, because it is in fact used to create the synthetic observations. The selected alternate covariates are different from the RD2 case, with night light intensity and population density ($\mathbf{X}_{ff}$) and the variance of population density within each 1° by 1° inversion gridcell ($\mathbf{Q}_{ff}$) being selected instead (Table 1). These differences reflect the many differences between the RD and SD setups, including among others the nature of the true FF fluxes and the impact of transport model errors. The impact of using these datasets on the posterior cross-covariances (Figure 2) and cross-correlations (Figure 3) in the biospheric and FF uncertainties is similar to that observed in RD2, however with a lower Forbenius norm of $\mathbf{V}_{ffbio}$.

For the SD case study, the FF, biospheric, and total fluxes can also be compared to their "true" values (Figure 4). Results confirm that, although the separation of FF and biospheric flux becomes more uncertain in the absence of a good inventory, the separation is still relatively robust, in the sense that the true fluxes lie within the range of the posterior uncertainties. The poorest performance is in the Midwest, which is also the region with the highest cross-covariance (Figure 2) and cross-correlation (Figure 3) in the posterior uncertainties. Another indication of the good overall performance of the flux disaggregation is the low RMSE of the 1° by 1° fluxes at the native temporal resolution of the inversion (3-hourly for

biospheric fluxes, 8-day for FF fluxes), namely 0.33 μmol $m^2 sec^{-1}$ for FF emissions and 0.22 μmol $m^2 sec^{-1}$ for biospheric fluxes, relative to the magnitude of the fluxes (Figure 4).

5 Conclusions

With increasing attention being placed on the accurate monitoring of FF emissions, the ability to provide a top-down verification of inventory-based estimates of FF emissions by disaggregating FF and biospheric fluxes from inverse models is a very promising development. The sparsity of in-situ measurement network, relative contribution of FF flux to the total CO_2 flux (i.e. FF and biospheric fluxes) and large model-data mismatch severely limits the capability of inverse models in accurately estimating FF emissions to well-instrumented regions and winter months. The presented analyses demonstrate that with the North American in-situ measurement network in January 2008 the proposed method is successful in separating FF and biospheric fluxes at sub-continental scales. This confirms the potential of using a statistical approach, based on the unique spatiotemporal signature of FF emissions, to isolate and estimate FF emissions using atmospheric CO_2 observations.

This work demonstrates the ability to estimate FF emissions through a statistical disaggregation of biospheric and FF fluxes based on different error covariance models and flux covariates for the FF and biospheric flux components. The use of a temporally varying FF inventory data as the basis for an error covariance model and use as a flux covariate is shown to provide a relatively superior disaggregation over other static FF related variables. Nevertheless, synthetic data case studies show that even without an explicit FF inventory available for the error covariance model or flux covariate, other static FF related variables (covariates: mean nightlights, mean population density, error covariance model: variance of population density) are can provide sufficient information to adequately disaggregate and estimate FF and biospheric fluxes. Additionally, this work demonstrates a process for selecting between various error covariance models, which can also prove to be useful in optimally defining error covariance models in tracer-based inversions that attempt to disaggregate component fluxes.

The ability to accurately disaggregate and estimate FF and biospheric fluxes simultaneously using atmospheric data is a continuing challenge. This pursuit relies heavily on external conditions including but not limited to the representativeness and density of the observational network as well as transport model accuracy. Nevertheless, the methodological advances presented here, specifically the utilization of the unique spatiotemporal structure of FF emissions, offers an approach to optimally exploit the information content of the available data to help provide a complimentary approach to estimating FF emission using atmospheric CO_2 data.

Acknowledgements

This work was supported by Sandia National Laboratories' LDRD (Laboratory Directed Research and Development) funds, sponsored by the Geosciences Investment Area. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the US Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Figures, Table and Appendix Captions

Table 1: Covariates and error covariance model selected by BIC selected for $\mathbf{X}_{ff}$ and $\mathbf{Q}_{ff}$ for the three case studies

Case Studies	**Covariates**					**Frobenius Norm** (μmol $m^2 sec^{-1}$)^2	**FF Covariance Model**
	Mean Night	**Mean Population**	**% Built Up Area**	**% Urban**	**FF Inventory**		
RD1					✓	3.47	Mean (FF
RD2		✓		✓	N/A	4.48	Maximum (Night Lights Intensity)
SD	✓	✓			N/A	3.86	Variance (% Population Density)

Figure 1: Regional classification map for aggregating fluxes and a posteriori cross-covariances of biospheric and fossil fuel fluxes.

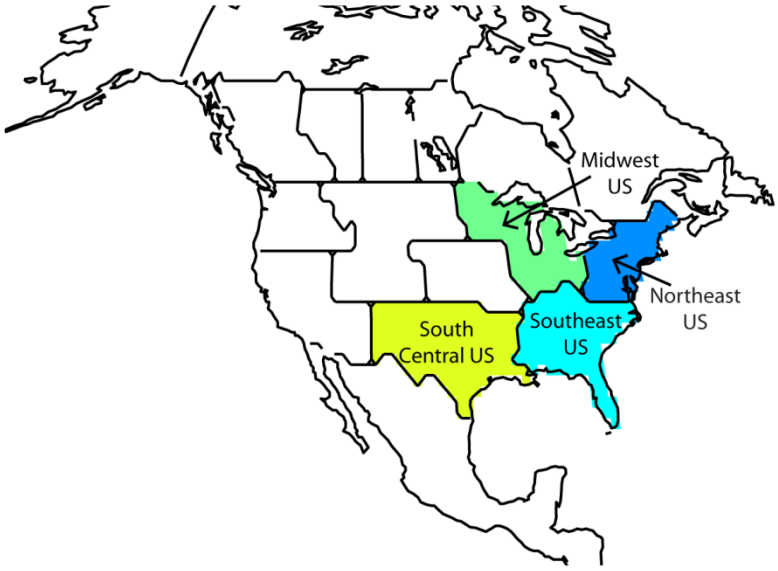
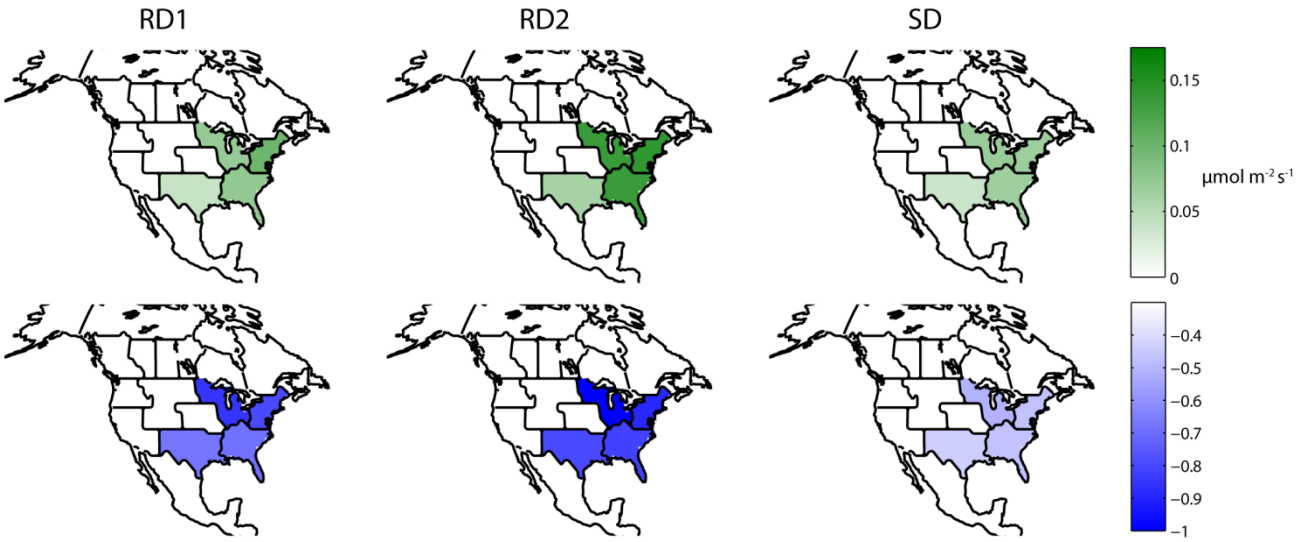


Figure 2: Row 1 represents the aggregated a posteriori cross-covariances*. Row 2 shows the aggregated correlation between the a posteriori biospheric and fossil fuel flux covariances for the three inversion case studies. Smaller covariance and correlation implies better separation between fossil fuel and biospheric flux estimates.



*Note: Shown here is square root of the absolute value of the cross-covariance ($\overline{\boldsymbol{V}}_{ff,bio}$) to be comparable to uncertainty bounds from Figure 3 and 4.

Figure 3: Estimates of the fossil fuel, biospheric, and total flux with 1σ (first hash mark) and 2σ uncertainty bounds for the regions shown in Figure 1 for the two real data cases, here diamonds represent RD1 and circles represent RD2.

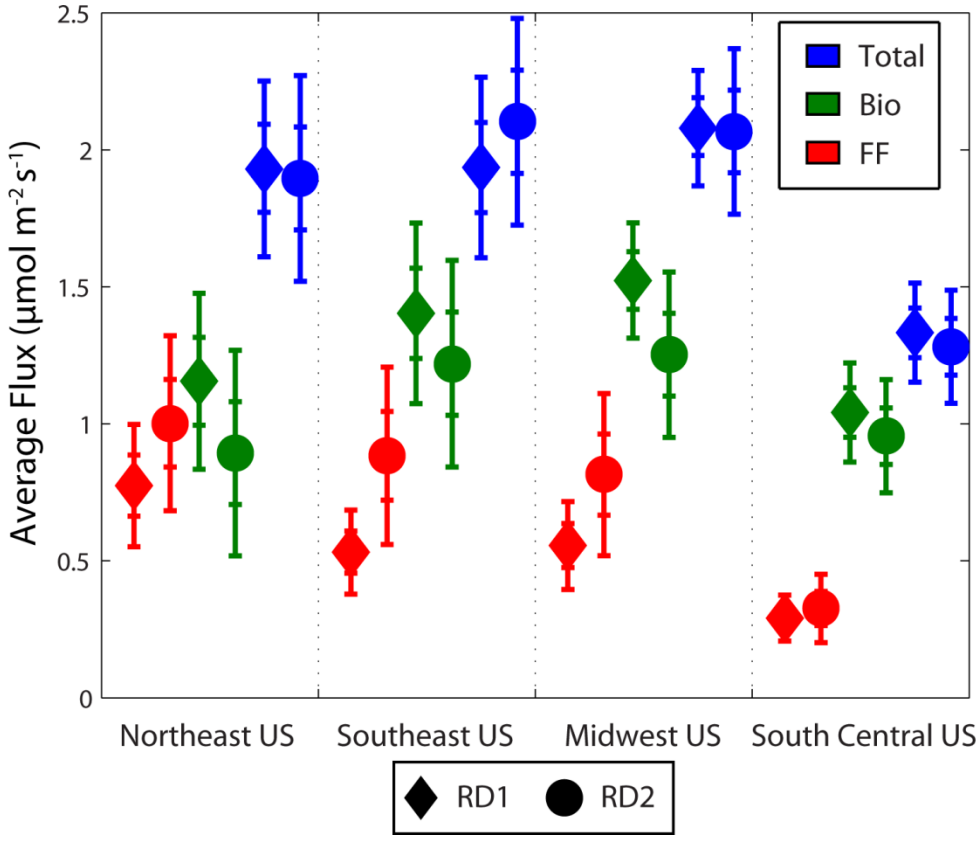
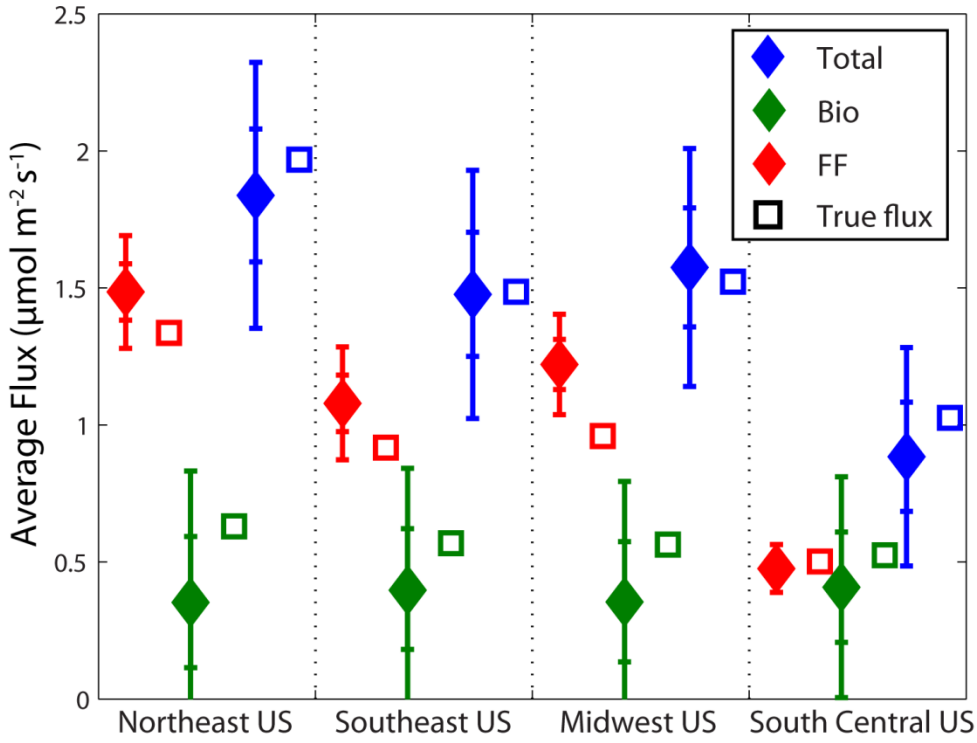


Figure 4: Estimates of fossil fuel, biospheric and total fluxes with 1σ (first hash mark) and 2σ uncertainty bounds for the regions shown in Figure 1 for the synthetic data case.



Appendix 1: Locations and measurement times of CO_2 concentrations across study sites (in-situ towers). Note this table has been reproduced from Shiga et al. 2014.

Tower	Name	Latitude	Longitude	Time of Day (Local Time, hours)	Height [m]	$\overline{\sigma_R}$ [ppm]
LEF	Park Falls	45.95	-90.27	1 4 7 10 13 16 19 22	396	2.2
WKT	Moody	31.32	-97.33	1 4 7 13 16 19 22	457	2.5
WBI	West Branch	41.73	-91.35	1 4 7 10 13 16 19 22	379	3.3
BAO	Boulder Atmospheric Observatory	40.05	-105.01	1 4 7 13 16 19 22	300	2.3
SCT	South Carolina Tower	33.41	-81.83	1 4 7 13 16 19 22	305	3.9
WGC	Walnut Grove	38.27	-121.49	1 4 7 13 16 19 22	483	4.7
AMT	Argyle	45.03	-68.68	13 16 19	107	3.3
BRW	Barrow	71.32	-156.61	1 4 7 10 13 16 19 22	17	0.5
FRD	Fraserdale	49.88	-81.57	13 16 19	40	1.4
CDL	Candle Lake	53.99	-105.12	13 16 19	30	1.3
SBL	Sable Island	43.93	-60.02	1 4 7 10 13 16 19 22	25	1.5
EGB	Egbert	44.23	-79.78	13	3	4.6
ETL	East Trout Lake	54.35	-104.99	10 13 16 19	105	2.2
LLB	Lac LaBiche	54.95	-112.45	13	10	4.4
CHI	Chibougamau	49.69	-74.34	13 16 19	30	2.1
HFM	Harvard Forest	42.54	-72.17	13 16 19	30	5.5
ARM	Southern Great Plains	36.8	-97.5	13 16 19	60	3.7
CVA	Canaan Valley	39.06	-79.42	13	7	4.6
MM	Morgan Monroe	39.32	-86.41	13 16 19	48	5
OZA	Ozark	38.74	-92.2	13 16 19	30	3.3
KEW	Kewanee	41.28	-89.97	13 16 19	140	4.2
CEN	Centerville	40.79	-92.88	13 16 19	110	3.8
MEA	Mead	41.14	-96.46	13 16 19	122	4.4
ROL	Round Lake	43.53	-95.41	13 16 19	110	4.1
GAL	Galesville	44.09	-91.34	13 16 19	122	3.2
SNP	Shenandoah National Park	38.62	-78.35	1	17	3
SPL	Storm Peak Lab	40.45	-106.73	1	9	1.2
NWR	Niwot Ridge	40.05	-105.58	1	5	2.6

HDP	Hidden Peak Snowbird	40.56	-111.65	1	18	0.9
FIR	Fir	44.65	-123.55	13 16 19	38	5
MET	Metolius	44.45	-121.56	13 16 19	34	2
YAH	Yaquina Head	44.67	-124.07	13 16 19	13	3.8
MAP	Mary's Peak	44.5	-123.55	13 16 19	8	2.2
NGB	NGBER	43.47	-119.69	13 16 19	7	1.2
LJA	La Jolla	32.87	-117.26	1 4	5	4.3

Appendix 2: Metrics computed from a posteriori covariance for regions shown in figure 1. Note a posteriori cross-correlation coefficient and cross-covariance have also been shown in figure 2. Correlation coefficient shown in figure 2 is computed by diving a posteriori cross-covariance by the product of a posteriori standard deviation of biospheric and fossil fuel fluxes.

I. A posteriori Correlation Coefficient between Biospheric and Fossil Fuel Fluxes			
	RD1	RD2	PD
Northeast	-0.56	-0.67	-0.19
Southeast	-0.44	-0.59	-0.19
Midwest	-0.63	-0.82	-0.24
South central	-0.43	-0.56	-0.16
II. A Posteriori Cross Covariance between Biospheric and Fossil Fuel Fluxes; **Units**: $(\mu mol\ m^2 sec^{-1})^2$			
	RD1	RD2	PD
Northeast	-0.010	-0.020	-0.005
Southeast	-0.006	-0.018	-0.004
Midwest	-0.005	-0.018	-0.005
South central	-0.002	-0.004	-0.001
III. A Posteriori Standard Deviation of Fossil Fuel Fluxes **Units**: $\mu mol\ m^2 sec^{-1}$			
	RD1	RD2	PD
Northeast	0.112	0.160	0.103
Southeast	0.077	0.162	0.103
Midwest	0.080	0.148	0.092
South central	0.042	0.062	0.044
IV. A Posteriori Standard Deviation of Biospheric Fluxes **Units**: $\mu mol\ m^2 sec^{-1}$			
	RD1	RD2	PD
Northeast	0.160	0.188	0.240
Southeast	0.165	0.188	0.222
Midwest	0.105	0.151	0.220
South central	0.090	0.103	0.201

References

Andres, R. J., Fielding, D. J., Marland, G., Boden, T. A., Kumar, N., and Kearney, A. T.: Carbon dioxide emissions from fossil-fuel use, 1751-1950, Tellus B, 51, 759-765, 10.1034/j.1600-0889.1999.t01-3-00002.x, 1999.

Andres, R. J., Gregg, J. S., Losey, L., Marland, G., and Boden, T. A.: Monthly, global emissions of carbon dioxide from fossil fuel consumption, Tellus Series B-Chemical and Physical Meteorology, 63, 309-327, 10.1111/j.1600-0889.2011.00530.x, 2011.

Casella, G., Berger, R. L,: Stalistical lnference, Duxbury, Pacific Grove, CA, 2002.

Duren, R. M., and Miller, C. E.: Measuring the carbon emissions of megacities, Nature Climate Change, 2, 560-562, 10.1038/nclimate1629, 2012.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., and Davis, E. R.: Mapping city lights with nighttime data from the DMSP Operational Linescan System, Photogrammetric Engineering and Remote Sensing, 63, 727-734, 1997.

Gill, P. E., Murray, W., and Wright, M. H.: Practical optimization, 1981.

Golub, G. H., and Van Loan, C. F.: Matrix computations, JHU Press, 2012.

Gourdji, S. M., Hirsch, A. I., Mueller, K. L., Yadav, V., Andrews, A. E., and Michalak, A. M.: Regional-scale geostatistical inverse modeling of North American CO(2) fluxes: a synthetic data study, Atmospheric Chemistry and Physics, 10, 6151-6167, 10.5194/acp-10-6151-2010, 2010.

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., and Michalak, A. M.: North American CO2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457-475, 10.5194/bg-9-457-2012, 2012.

Gurney, K. R., 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 CO2 sources and sinks using atmospheric transport models, Nature, 415, 626-630, 10.1038/415626a, 2002.

Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., and de la Rue du Can, S.: High Resolution Fossil Fuel Combustion CO2 Emission Fluxes for the United States, Environ. Sci. Technol., 43, 5535-5541, 10.1021/es900806c, 2009.

Gurney, K. R., Razlivanov, I., Song, Y., Zhou, Y., Benes, B., and Abdul-Massih, M.: Quantification of Fossil Fuel CO2 Emissions on the Building/Street Scale for a Large U.S. City, Environ. Sci. Technol., 46, 12194-12202, 10.1021/es3011282, 2012.

Kitanidis, P. K.: Quasi-Linear Geostatistical Theory for Inversing, Water Resources Research, 31, P.-2411, 199510.1029/95wr01945, 1995.

Kort, E. A., Frankenberg, C., Miller, C. E., and Oda, T.: Space-based observations of megacity carbon dioxide, Geophysical Research Letters, 39, 10.1029/2012gl052738, 2012.

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, Journal of Geophysical Research: Atmospheres, 108, n/a-n/a, 10.1029/2002jd003161, 2003.

Mardia, K. V., Kent, J. T., Bibby, J. M.,: Multivariate Analysis, Probability and Mathematical Statistics, Academic Press, San Diego, CA, 518 pp., 1979.

Michalak, A. M., and Kitanidis, P. K.: A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification, Water Resources Research, 39, 1033, 10.1029/2002WR001480, 2003.

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res, 109, D14109, 2004.

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., and Tans, P. P.: Linking emissions of fossil fuel CO2 and other anthropogenic trace gases using atmospheric 14CO2, Journal of Geophysical Research, 117, 23-PP., 201210.1029/2011jd017048, 2012.

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

Olsen, S. C., and Randerson, J. T.: Differences between surface and column atmospheric CO2 and implications for carbon cycle research, Journal of Geophysical Research: Atmospheres, 109, n/a-n/a, 10.1029/2003jd003968, 2004.

Pacala, S., Breidenich, C., Peter, G. B., Fung, I., Gunson, M., Heddle, G., Law, B. E., Marland, G., Paustian, K., Prather, M., Randerson, J. T., Tans, P. P., and Wofsy, S. C.: Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements. Committee on Methods for Estimating Greenhouse Gas Emissions, National Research Council Report, The National Academies Press, Washington, D.C., 124 pp., 2010.

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 Biogeochemical Cycles, 11, 10.1029/97gb02268, 1997.

Ray, J., Yadav, V., Michalak, A. M., van Bloemen Waanders, B., and McKenna, S. A.: A multiresolution spatial parameterization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions, Geosci. Model Dev. Discuss., 7, 1277-1315, 10.5194/gmdd-7-1277-2014, 2014.

Rayner, P. J., Enting, I. G., Francey, R. J., and Langenfelds, R.: Reconstructing the recent carbon cycle from atmospheric CO2, delta13C and O2/N2 observations*, Tellus B, 51, 213-232, 10.1034/j.1600-0889.1999.t01-1-00008.x, 1999.

Rodell, M., Houser, P., Jambor, U. e. a., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., and Bosilovich, M.: The global land data assimilation system, Bulletin of the American Meteorological Society, 85, 2004.

Schneider, A., Friedl, M. A., and Potere, D.: A new map of global urban extent from MODIS satellite data, Environ. Res. Lett., 4, 044003, 2009.

Schwarz, G.: Estimating the Dimension of a Model, The Annals of Statistics, 6, 461-464, 1978.

Tans, P. P., Fung, I. Y., and Takahashi, T.: Observational Contrains on the Global Atmospheric Co2 Budget, Science, 247, 1431-1438, 10.1126/science.247.4949.1431, 1990.

Turnbull, J. C., Miller, J. B., Lehman, S. J., Tans, P. P., Sparks, R. J., and Southon, J.: Comparison of (CO2)-C-14, CO, and SF6 as tracers for recently added fossil fuel CO2 in the atmosphere and implications for biological CO2 exchange, Geophysical Research Letters, 33, 10.1029/2005gl024213, 2006.

Turnbull, J. C., Karion, A., Fischer, M. L., Faloona, I., Guilderson, T., Lehman, S. J., Miller, B. R., Miller, J. B., Montzka, S., Sherwood, T., Saripalli, S., Sweeney, C., and Tans, P. P.: Assessment of fossil fuel carbon dioxide and other anthropogenic trace gas emissions from airborne measurements over

Sacramento, California in spring 2009, Atmospheric Chemistry and Physics, 11, 705-721, 10.5194/acp-11-705-2011, 2011.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano, J. A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmospheric Chemistry and Physics, 6, 3423-3441, 2006.

Wu, L., Bocquet, M., Lauvaux, T., Chevallier, F., Rayner, P., and Davis, K.: Optimal representation of source-sink fluxes for mesoscale carbon dioxide inversion with synthetic data, Journal of Geophysical Research: Atmospheres (1984–2012), 116, 2011.

Yadav, V., Mueller, K. L., Dragoni, D., and Michalak, A. M.: A geostatistical synthesis study of factors affecting gross primary productivity in various ecosystems of North America, Biogeosciences, 7, 2655-2671, 10.5194/bg-7-2655-2010, 2010.

Yadav, V., Mueller, K. L., and Michalak, A. M.: A backward elimination discrete optimization algorithm for model selection in spatio-temporal regression models, Environmental Modelling & Software, 42, 88-98, 10.1016/j.envsoft.2012.12.009, 2013.