Geosci. Model Dev. Discuss., 7, 1277–1315, 2014
www.geosci-model-dev-discuss.net/7/1277/2014/
doi:10.5194/gmdd-7-1277-2014

Geoscientific Model Development Discussions
Open Access

A multiresolution spatial parameterization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions

J. Ray[1], V. Yadav[2], A. M. Michalak[2], B. van Bloemen Waanders[3], and S. A. McKenna[4]

[1]Sandia National Laboratories, P.O. Box 969, Livermore CA 94551, USA
[2]Carnegie Institution for Science, Stanford, CA 94305, USA
[3]Sandia National Laboratories, P.O. Box 5800, Albuquerque NM 87185-0751, USA
[4]IBM Research, Smarter Cities Technology Centre, Bldg 3, Damastown Industrial Estate, Mulhuddart, Dublin 15, Ireland

Received: 28 December 2013 – Accepted: 20 January 2014 – Published: 6 February 2014

Correspondence to: J. Ray (jairay@sandia.gov)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The characterization of fossil-fuel CO_2 ($ffCO_2$) emissions is paramount to carbon cycle studies, but the use of atmospheric inverse modeling approaches for this purpose has been limited by the highly heterogeneous and non-Gaussian spatiotemporal variability of emissions. Here we explore the feasibility of capturing this variability using a low-dimensional parameterization that can be implemented within the context of atmospheric CO_2 inverse problems aimed at constraining regional-scale emissions. We construct a multiresolution (i.e., wavelet-based) spatial parameterization for $ffCO_2$ emissions using the Vulcan inventory, and examine whether such a parameterization can capture a realistic representation of the expected spatial variability of actual emissions. We then explore whether sub-selecting wavelets using two easily available proxies of human activity (images of lights at night and maps of built-up areas) yields a low-dimensional alternative. We finally implement this low-dimensional parameterization within an inversion, where a sparse reconstruction algorithm, an extension of Stagewise Orthogonal Matching Pursuit (StOMP), is used to identify the wavelet coefficients. We find that (i) the spatial variability of fossil fuel emission can indeed be represented using a low-dimensional wavelet-based parameterization, (ii) that images of lights at night can be used as a proxy for sub-selecting wavelets for such analysis, and (iii) that implementing this parameterization within the described inversion framework makes it possible to quantify fossil fuel emissions at regional scales under some simplifying conditions.

1 Introduction

The characterization of fossil-fuel CO_2 ($ffCO_2$) emissions is paramount to carbon cycle studies. $ffCO_2$ emissions are the largest net carbon flux at the atmosphere–surface interface (Friedlingstein, 2006) and spatially-disaggregated (or gridded) $ffCO_2$ emissions form a critical input into general circulation and integrated assessment models

GMDD
7, 1277–1315, 2014

(Andres et al., 2012). An understanding of fossil fuel emissions is clearly necessary for characterizing the anthropogenic climate impact. In addition, a process-level understanding of the terrestrial carbon sink requires the quantification of terrestrial biospheric fluxes at fine spatiotemporal scales, which, in turn, requires the differentiation between anthropogenic and biospheric fluxes at those scales.

Gridded inventory estimates of $ffCO_2$ emissions can be derived using socioeconomic data (Oda and Maksyutov, 2011; Rayner et al., 2010), and such "bottom-up" estimates have been proposed as a means of monitoring international agreements aimed at mitigating $ffCO_2$ emissions (Pacala et al., 2010). Gridded inventory estimates are derived from $ffCO_2$ budgets and produced by a few institutions (Andres et al., 2012). These budgets are compiled from national and provincial statistics on fossil-fuel production and consumption. These large-scale estimates can then be downscaled to finer spatiotemporal scales using easily-observed proxies of human activity (and consequently $ffCO_2$ emissions) such as images of lights at night (henceforth "nightlights"), population density, etc. (Oda and Maksyutov, 2011; Rayner et al., 2010; Doll et al., 2000). More sophisticated approaches to the fine-scale bottom-up estimation of $ffCO_2$ emissions have also begun to emerge, including, for example, the Vulcan inventory that includes estimates for the US at a 10 km and hourly resolution for 2002 (http://vulcan.project.asu.edu; Gurney et al., 2009). Such approaches rely on detailed reporting and monitoring data, which are not currently available for many regions of the world.

Although inventory estimates provide a key tool in the understanding of anthropogenic CO_2 emissions, their accuracy at large scales depends on the accuracy of reported national consumption data, e.g., the error in $ffCO_2$ emissions from China lies in the 15–20 % range (Gregg et al., 2008). When evaluated at finer spatiotemporal scales, their accuracy also depends on the method used to disaggregate national/provincial $ffCO_2$ emission budgets to finer spatial scales. Pregger et al. (2007) showed that two 0.5° inventories for Western Europe, for 2003, differed at the grid-cell level by 20 %, with a standard deviation of 40 %; at finer resolutions, the disagreement worsened. Sources

GMDD
7, 1277–1315, 2014

of errors in inventories are discussed in detail in Andres et al. (2012); Raupach et al. (2009); Rayner et al. (2010). These uncertainties lead to frequent corrections of sub-national inventories of combustion products (Streets et al., 2006) and model predictions of CO_2 concentrations that disagree with observations locally (Brioude et al., 2012).

Given both the benefits, and the limitations, of inventory-based estimates, interest has emerged in the development of "top-down" estimates of $ffCO_2$ emissions. These estimates rely on attributing the observed variability in CO_2 and related trace gas concentrations in the atmosphere to the underlying fossil fuel emissions, through the application of statistical atmospheric inversion methods. Some of the proposed approaches (e.g. Turnbull et al., 2011) have relied on observations of $\Delta^{14}CO_2$ measurements or other non-CO_2 tracers. One challenge with these approaches, however, is a combination of the limited number of available observations and the need to understand emission ratios for any co-emitted tracers. Atmospheric inversions relying on atmospheric CO_2 measurements, on the other hand, have primarily targeted biospheric CO_2 fluxes, often by first pre-subtracting the fossil fuel influence calculated from an inventory; see Ciais et al. (2010) for a review of atmospheric inversion methods. The measurements then consist of CO_2 concentration obtained from in situ or remote sensing observations, and estimates are obtained at a variety of spatiotemporal resolutions for either global or regional domains. The statistical approaches applied typically rely on Gaussian assumptions for flux residuals from prior estimates or other spatiotemporal patterns. Investigations aimed at the estimation of $ffCO_2$ emissions are less common, in part because the statistical assumptions used in inversions aimed at understanding biospheric fluxes are ill-suited to the highly heterogeneous and non-Gaussian spatiotemporal variability of $ffCO_2$ emissions. However, some estimates of $ffCO_2$ emissions at urban scales are beginning to emerge, including for Salt Lake City (McKain et al., 2012), Indianapolis (Gurney et al., 2012) and Sacramento (Turnbull et al., 2011).

The goal of the work presented here is to explore the possibility of defining an inversion framework that is specifically targeted at the characteristics of the spatiotemporal variability of $ffCO_2$ emissions at regional scales. We will model spatial variability

GMDD
7, 1277–1315, 2014

at $1^\circ \times 1^\circ$ resolution. Such a framework would require, among other things, a low-dimensional spatial parameterization of $ffCO_2$ emissions, given the data limitations associated with any atmospheric inversion system. We explore this topic through a sequence of three specific objectives:

1. *Identification of a low-dimensional parameterization for* $ffCO_2$ *emissions:* $ffCO_2$ emissions are strongly non-stationary (see Fig. 1), and any parameterization must be able to represent such variability. Wavelets, which are an orthogonal basis set with compact support, are widely used to model non-stationary fields e.g. images (Strang and Nguyen, 1997; Chan and Shen, 2005). We will examine a number of wavelet families to identify the wavelet type that can represent $ffCO_2$ emissions most efficiently i.e., with minimum error if only a limited number of wavelets were to be retained. The $ffCO_2$ emissions will be obtained from the Vulcan inventory (for the US only) as a realistic example of what the variability of true $ffCO_2$ emissions is likely to be. This objective will ultimately answer the question of whether a low-dimensional parameterization is possible for the type of spatial variability expected in real $ffCO_2$ emissions.

2. *Evaluation of the use of a low-dimensional parameterization in combination with easily-available proxies of anthropogenic emissions:* For most areas of the world, fine-scale estimates of $ffCO_2$ emissions are based on downscaling of national inventories using easily observed proxies of human activity, such as maps of night-lights or of built-up areas (BUA). In this second objective, we will use the wavelets selected in the first objective, and sub-select them using these two proxies of human activity for the United States. We will then evaluate the degree to which the remaining wavelets can be used to represent the complexity of spatial patterns in $ffCO_2$ emissions. The Vulcan inventory will be used for this purpose too. This objective will answer the question of whether an easily observable proxy can be used to reduce the dimensionality of a wavelet-based spatial parameterization for

GMDD
7, 1277–1315, 2014

$ffCO_2$ emission fields. The set of wavelets selected in this manner form a random field model, which we will refer to as the Multiscale Random Field (MsRF) model.

3. *Evaluation of the parameterization in an atmospheric inversion, using sparse reconstruction:* In the third objective, we will use the reduced basis identified in Objective 2 within a synthetic-data atmospheric inversion aimed at characterizing the spatiotemporal variability of US $ffCO_2$ emissions. A new sparsity-enforcing optimization method that preserves the non-negative nature of $ffCO_2$ emissions will be used to solve the inverse problem. (The terms *sparsity* and *sparsity-enforcement* are defined in Sect. 2). The new sparse reconstruction method is used to ensure an unique solution and to guard against overfitting (fitting to observational noise). The optimization procedure will identify the subset of wavelets in the MsRF that can be actually estimated from the observations, while "turning off" the rest. In doing so, it will ensure that the MsRF, as designed, has sufficient flexibility to extract information on $ffCO_2$ from the observations. For simplicity, the synthetic data inversion will focus only on $ffCO_2$ emissions. We recognize that an ultimate application with real data would require a combination of methods to capture both the biospheric and fossil fuel signals, or would require the pre-subtraction of the influence of biospheric fluxes on observations. For the purposes of the work presented here, however, the question that we aim to answer is whether an inversion approach based on a low-dimensional parameterization is feasible, even under idealized conditions.

The paper is structured as follows. In Sect. 2, we review the use of wavelet modeling in inverse problems. In Sect. 3, we investigate families of wavelets for modeling $ffCO_2$ emissions and construct two MsRF models, based on nightlights and maps of BUA. In Sect. 4, we describe the inverse problem and the numerical method used to solve it. In Sect. 5 we perform inversion tests with synthetic data. Conclusions are in Sect. 6.

GMDD
7, 1277–1315, 2014

2 Wavelet modeling in inverse problems

Wavelets are a family of orthogonal bases with compact support (Williams and Amaratunga, 1994; Walker, 2008). They are generated using a scaling function ϕ' which obeys the recursive relationship: $\phi'(x) = \sum_i c_i \phi'(2x - i)$. A wavelet ϕ is generated from the scaling function by taking differences e.g., $\phi(x) = \sum_i (-1)^i c_{1-i} \phi'(2x - i)$. The choice of the filter coefficients c_i and ϕ' determine the type of the resulting wavelets. The simplest type are the Haars, which are symmetric, but not differentiable. Wavelets can be shrunk and translated i.e., $\phi_{s,i} = 2^{\frac{s}{2}} \phi(2^s x - i)$, where s is the dilation scale and i refers to translation (location). This allows them to model complex, non-stationary functions efficiently. For each increment in scale, the support of the wavelet halves. Wavelets are defined on dyadic (power-of-two) hierarchical or multi-resolution grids.

Consider a domain of size D, discretized by a hierarchy of meshes with resolutions $\Delta D/D = \{1, 1/2, 1/2^2, \ldots 1/2^M\}$. Wavelets are defined on each of the levels of the hierarchical mesh and can be positioned at any even-numbered grid-cell $2i, 0 \le 2i \le 2^s - 1$, on any scale s of the hierarchical mesh. An arbitrary 1-D function $g(x)$ can be represented as $g(x) = w'\phi'(x) + \sum_{s=1}^{M} \sum_{i=0}^{2^{(s-1)}-1} w_{s,i} \phi_{s,i}(x)$, where the coefficients (or weights) $w_{s,i}$ and w' are obtained, via projections of $g(x)$, using fast wavelet transforms. In 2-D, a function $g(x,y)$, defined on a $D \times D$ domain with a hierarchical $2^M \times 2^M$ mesh, can be wavelet-transformed by applying 1-D wavelet transforms repeatedly, e.g., first by rows and then by columns. Wavelets of scale s have a support $2^{M-s} \times 2^{M-s}, 0 \le s \le M$ and can be positioned (in 2-D space) at location $(i,j), 0 \le (i,j) < 2^s$. A 2-D wavelet transform results in $2^M \times 2^M$ wavelet coefficients. In general,

$$g(x,y) = w'\phi'(x,y) + \sum_{s=1}^{M} \sum_{i,j}^{\mathfrak{W}^{(s)}} w_{s,i,j} \phi_{s,i,j}(x,y) \quad (1)$$

GMDD
7, 1277–1315, 2014

where $\mathfrak{W}^{(s)}, |\mathfrak{W}^{(s)}| = (4^s - 4^{s-1})$, is the set of (i, j) indices of wavelet coefficients on scale s. A large number of fine-scale (i.e., high s) wavelets model fine spatial details.

2.1 Wavelet-based random field models

Wavelets are often used to represent fields, most commonly, to represent images e.g., the JPEG2000 standard (Taubman and Marcellin, 2002). Most fields/images are compressible in a judiciously chosen wavelet basis i.e., most of the wavelet coefficients are small and can be discarded to form a reduced-rank approximation of the field (Welstead, 1999). In Auger and Tangborn (2002), a reduced-rank wavelet model was developed for global, time-variant CH_4 concentration, which were updated with limited observations using a Kalman filter. The selection of wavelets to form the random field (RF) model was done empirically, by decimating the fine-scale wavelets. The construction of the RF model can be performed more rigorously if a prior model is available. In Jafarpour (2011) wavelets were used in the reconstruction of permeability fields from limited measurements of flow through a porous medium. An ensemble of permeability field realizations, drawn from the prior distribution, were used to develop a multivariate Gaussian prior distribution for the wavelet coefficients. The RF model was created by discarding wavelet coefficients with small means. In Romberg et al. (2001), the authors constructed a hidden Markov model to encode the relationship between the wavelet coefficients on adjacent scales of the wavelet tree. This RF model was used to reconstruct compressively sensed signals (Duarte et al., 2008) and images (He and Carin, 2009). Thus the use of wavelet-based RF models to parameterize and reconstruct complex fields from limited measurements is quite common.

The RF model need not be constructed offline using a prior; it can also be constructed during the inversion, in a data-driven manner. This occurs in the compressive sensing (CS) of signals and images (Candes and Wakin, 2008). In this approach, all the wavelets in a field's representation are retained and the ones that cannot be estimated from available observations are identified and removed. Let **g** be an image of

GMDD
7, 1277–1315, 2014

size N that can be represented sparsely using $L \ll N$ wavelets. Let $\mathbf{g}'$, of size N_m, $L < N_m \ll N$, be its compressed measurement, obtained by projecting $\mathbf{g}$ onto a set of random vectors $\boldsymbol{\psi}_j$, i.e., $\mathbf{g}' = \boldsymbol{\Psi}\mathbf{g} = \boldsymbol{\Psi}\boldsymbol{\Phi}\mathbf{w}$. Here the rows of $\boldsymbol{\Psi}$ consist of the random vectors $\boldsymbol{\psi}_j$ and columns of $\boldsymbol{\Phi}$ consist of wavelets ϕ_i. $\boldsymbol{\Phi}$ is a $N \times N$ matrix while $\boldsymbol{\Psi}$ is $N_m \times N$. The bulk of the theory that allows reconstruction of the image/field with such observations was established in Candes and Tao (2006); Donoho (2006); Candes et al. (2006). In CS, the reconstruction of $\mathbf{g}$ (alternatively, $\mathbf{w}$) can be performed using a number of methods. It is posed as an optimization problem

$$\underset{\mathbf{w}\in\mathbb{R}^N}{\text{minimize}} \|\mathbf{w}\|_1, \quad \text{subject to } \|\mathbf{g}' - \mathbf{A}\mathbf{w}\|_2 < \epsilon_2, \quad \mathbf{A} = \boldsymbol{\Psi}\boldsymbol{\Phi}. \tag{2}$$

Thus we enforce sparsity in $\mathbf{w}$ with its ℓ_1 norm, under the constraint that the ℓ_2 norm of the misfit between $\mathbf{g}'$ and the image reconstructed from $\mathbf{w}$ is kept within a bound. Some of the methods to solve this problem are Basis Pursuit (Chen et al., 1998), Matching Pursuit (Mallat and Zhang, 1993), Orthogonal Matching Pursuit (Tropp and Gilbert, 2007) and Stagewise Orthogonal Matching Pursuit (StOMP; Donoho et al., 2012).

2.2 Wavelets and sparsity in inverse problems

Sparsity is often used to solve inverse problems in physics, with the $\boldsymbol{\Psi}$ operator representing the physical process. In Li and Jafarpour (2010), sparsity was used to estimate a permeability field (represented by wavelets) from fluid transport observations. A good review of the use of sparsity in permeability field reconstruction is available in Jafarpour (2013). Seismic tomography, which estimates subsurface geologic facies, also has exploited sparsity for reconstruction. This has been demonstrated with wavelet representations of the subsurface and nonlinear forward models (Loris et al., 2007; Simons et al., 2011). Gholami and Siahkoohi (2010) used a split Bregman iteration (Goldstein and Osher, 2009) to solve a seismic tomography problem, imposing sparsity via soft thresholding (Donoho, 1995). The authors also expanded the imposition of sparsity from the wavelet space to the finite-difference space i.e., they used an ℓ_1 norm to

GMDD
7, 1277–1315, 2014

sparsify deviations of the solution from a "best-guess", in the absence of constraining observations. In this manner, both the prior/guessed solution and sparsity are used to regularize the inversion.

To summarize, wavelet-based RF models are routinely used in inverse problems. Their dimensionality can be reduced a priori using prior information. Data-driven dimensionality reduction can also be performed by enforcing sparsity during the inversion. This has been demonstrated with nonlinear problems too.

3 Constructing a multiscale random field model

We seek an approximate representation of $ffCO_2$ emissions, which is low dimensional or sparse i.e., many of the $w_{s,i,j}$ in Eq. (1) may be set to zero. For this purpose, we use $ffCO_2$ emissions from the Vulcan inventory, coarsened to 1° resolution and averaged over a year to yield $\overline{\mathbf{f}_V}$ (see Fig. 1 for a plot). The emissions are described on a $2^M \times 2^M$ grid, $M = 6$. The rectangular domain extents are given by the corners (24.5° N, 63.5° W) and (87.5° N, 126.5° W). $ffCO_2$ emissions are restricted to $\mathcal{R}$, the lower 48 states of the US.

3.1 Choosing a wavelet

We investigate a number of wavelet families (Haars, Daubechies, Symlets and Coiflets) in order to determine which provides the sparsest representation of $\overline{\mathbf{f}_V}$. $\overline{\mathbf{f}_V}$ is first subjected to a wavelet transform using a chosen wavelet. At each scale s, we remove wavelets which contribute little to $\overline{\mathbf{f}_V}$. We set "small" wavelet coefficients ("small" is defined as the ratio $|w_{s,i,j}/w_{\max,s}| \leq 10^{-3}$, where $w_{\max,s}$ is the wavelet coefficient on scale s with the largest magnitude) to zero. We refer to the fraction of zero wavelet coefficients at scale s as its *sparsity*. Figure 2 plots the sparsity at scale $s = 4, 6$ for a large combination of wavelet families and orders. We see that the Haar wavelet (also called the Daubechies, order 2 wavelet) provides the sparsest representation, making it

GMDD
7, 1277–1315, 2014

a candidate for developing a low-dimensional MsRF for $ffCO_2$ emissions. This is a consequence of the spatial distribution of $\overline{\mathbf{f}_V}$ – the map of $\overline{\mathbf{f}_V}$ (Fig. 1b) is largely empty west of -100° W, which manifests itself as small coefficients for the wavelets whose support lie in that region. This favors simpler (non-smooth) and low-order wavelets.

Next we investigate the variation of the magnitude of wavelet coefficients, as a function of the type of wavelets used to model $\overline{\mathbf{f}_V}$. We select 5 wavelet types e.g., Haars, Daubechies, order 4 and 8 and Symlet, order 4 and 6, and perform a wavelet transform of $\overline{\mathbf{f}_V}$. At each scale s, we set the "small" wavelet coefficients to zero. In Fig. 3 we plot the average and standard deviation of the non-zero wavelet coefficients. The means of the wavelet coefficients at the finer scales are small, regardless of the wavelet type. We see that while Haars provide the sparsest representation, the non-zero wavelet coefficients tend to have large magnitudes. In contrast, smoother wavelets with broader support e.g., Daubechies, 8th order, have more non-zero wavelet coefficients, but with smaller wavelet coefficients. This is a consequence of the spatial distribution of $\overline{\mathbf{f}_V}$ (Fig. 1) which has sharp gradients, placing smooth wavelets at a disadvantage. In Fig. 3, we also see that the means and standard deviations shrink, especially after scale $s = 3$; further, the distributions of wavelet coefficients arising from the different wavelet types begin to resemble each other. This arises from the fact that there are sharp boundaries around the areas where $ffCO_2$ emissions occur; when subjected to a wavelet transform, the region in the vicinity of a sharp boundary gives rise to large wavelet coefficients down to the finest scale. Thus the few non-zero wavelet coefficients at the finer scales assume similar values, irrespective of the wavelet type.

Finally, we check the accuracy of a Haar representation of $\overline{\mathbf{f}_V}$ at various levels of sparsity. Again, we define a "small" wavelet as $|w_{s,i,j}/w_{\max,s}| \leq \alpha$. We perform a wavelet transform of $\overline{\mathbf{f}_V}$ using Haar wavelets and sparsify (set small wavelets to zero) using $10^{-6} \leq \alpha \leq 1$. We then perform an inverse transform to reconstruct a "sparsified" $\overline{\mathbf{f}_V}'$. In Fig. 4, we plot overall sparsity, reconstruction error $\epsilon = \|\overline{\mathbf{f}_V}' - \overline{\mathbf{f}_V}\|_2 / \|\overline{\mathbf{f}_V}\|_2$ and the Pearson correlation between the true and reconstructed $\overline{\mathbf{f}_V}$ as a function of α. We define the

GMDD
7, 1277–1315, 2014

Pearson correlation between $\overline{\mathbf{f}_V}'$ and $\overline{\mathbf{f}_V}$ as

$$\rho\left(\overline{\mathbf{f}_V}', \overline{\mathbf{f}_V}\right) = \frac{\mathrm{cov}(\overline{\mathbf{f}_V}, \overline{\mathbf{f}_V}')}{\sigma_{\overline{\mathbf{f}_V}} \sigma_{\overline{\mathbf{f}_V}'}},$$

where $\sigma^2_{\overline{\mathbf{f}_V}}$ and $\sigma^2_{\overline{\mathbf{f}_V}'}$ are the variances of the true and reconstructed fluxes and $\mathrm{cov}(Z_1, Z_2)$ is the covariance between two random variables Z_1 and Z_2. Here $\|\cdot\|_2$ denotes the ℓ_2 norm. We find that for $\alpha < 10^{-2}$ there is practically no error (as measured by these metrics) though we achieve a sparsification of about 80 %. Even at a sparsity of around 90 %, the error is less that 10 %. Thus a small collection of Haar wavelets have the ability to reproduce $\overline{\mathbf{f}_V}$ with an acceptable degree of error. This low-dimensional character of a Haar representation of $\overline{\mathbf{f}_V}$ can be invaluable in an inverse problem predicated on sparse observations, and henceforth, we will proceed with Haar wavelets as the basis set of choice for representing $ffCO_2$ emissions.

Note that the sparse nature of $ffCO_2$ emissions, at a 1° resolution, is due to its correlation with human activity and static sources (e.g., power generation, highways etc.). Thus, while we may have arrived at our sparse models using an annual average of the Vulcan inventory, the same spatially sparse models can be used to represent $ffCO_2$ emission at a weekly resolution (as will be done later in the paper). Note that the location and strengths of intense sources of $ffCO_2$ emissions, such as powerplants can be found at the CARMA (Carbon Monitoring for Action) website (http://carma.org).

3.2 Constructing a random field model

We seek a spatial parameterization for $ffCO_2$ emissions, of the form

$$\mathbf{f} = w'\phi' + \sum_{s=1}^{M} \sum_{i,j} w_{s,i,j}\phi_{s,i,j}, \quad \{s,i,j\} \in W^{(s)} \tag{3}$$

GMDD
7, 1277–1315, 2014

where $W^{(s)}$ is a set containing a small number of Haar bases. We will investigate the usefulness of an easily observed proxy $\mathbf{X}$ of human activity (which correlates with $ffCO_2$ emissions) to select the components of $W^{(s)}$. Radiance calibrated nightlights (http://www.ngdc.noaa.gov/dmsp/download_radcal.html; Cinzano et al., 2000) have been used for constructing $ffCO_2$ inventories (Doll et al., 2000) and are an obvious choice for $\mathbf{X}$. However, nightlight radiances are also affected by economic factors (Raupach et al., 2009), and we will explore maps of built-up area as an alternative (http://www.sage.wisc.edu/atlas/maps.php?datasetid=18&includerelatedlinks=1&dataset=18; B. Miteva, personal communication, 2013). The map of BUA uses nightlight radiances in its computations, and so these are *not* independent proxies; however the BUA map also includes information from IGBP (International Geosphere–Biosphere Programme, http://www.igbp.net/) land-cover map. The two choices for $\mathbf{X}$ will be compared with respect to (1) sparsity, (2) the correlation between $\mathbf{X}$ and $\overline{\mathbf{f}_V}$ and (3) the ability of $W^{(s)}$ to capture $\overline{\mathbf{f}_V}$.

In Fig. 5 (top row), we plot maps of the two proxies, coarsened to 1° resolution. Comparing with Fig. 1 (right), we see that they bear a strong resemblance to $\overline{\mathbf{f}_V}$. We subject $\mathbf{X}$ to a wavelet transform and set all wavelet coefficients $|w_{s,i,j}| < \delta$ to zero, where δ is a user-defined threshold. The bases with non-zero coefficients are selected to constitute $W^{(s)}$. We reconstitute a "sparsified" proxy, $\mathbf{X}^{(s)}$, using just the bases in $W^{(s)}$, and compute the correlation between $\mathbf{X}^{(s)}$ and $\overline{\mathbf{f}_V}$. Finally, we project $\overline{\mathbf{f}_V}$ onto $W^{(s)}$, obtain its "sparsified" form $\overline{\mathbf{f}_V}^{(s)}$, and compute the error $\epsilon_f = \|\overline{\mathbf{f}_V}^{(s)} - \overline{\mathbf{f}_V}\|_2 / \|\overline{\mathbf{f}_V}\|_2$. In Fig. 5 (middle row), we plot the sparsity, correlation and ϵ_f for various values of δ. We do so for both nightlights and BUA. For nightlights, we achieve a sparsity of around 0.75 for $\delta < 10^{-2}$ i.e., we need to retain only a quarter of the Haar bases to represent nightlights. The nightlights so represented bear a correlation of around 0.7 with $\overline{\mathbf{f}_V}$, and achieve an error ϵ_f of around 0.1. Note that this measure of error reflects the inability of the MsRF to represent fine-scale details, and not spatially-aggregated quantities, which are represented more accurately. In contrast, using BUA as a proxy, we see that while

GMDD
7, 1277–1315, 2014

the sparsity achieved is similar, the correlation between $\mathbf{X}^{(s)}$ and $\overline{\mathbf{f}_V}$ is slightly higher. The behavior of ϵ_f is similar, except the error increases faster with δ, as compared to nightlights. However both nightlights and BUA maps show significant correlation with $\overline{\mathbf{f}_V}$ and the sparsified set of Haar bases that they (i.e., the proxies) provide (using $\delta = 10^{-2}$ in both the cases) allow us to construct a low dimensional parameterization of $ffCO_2$ emissions.

Finally, we use $\mathbf{X}^{(s)}$ to create a "prior model" $\mathbf{f}_{pr} = c\mathbf{X}^{(s)}$ for $ffCO_2$ emissions, $\mathbf{f}$. c is computed such that

$$\int_{\mathcal{R}} \overline{\mathbf{f}_V} dA = \int_{\mathcal{R}} \mathbf{f}_{pr} dA = c \int_{\mathcal{R}} \left(w'_{(X)} \phi' + \sum_{l,i,j} w_{(X),s,i,j} \phi_{l,i,j} \right) dA, \quad \{l,i,j\} \in W^{(s)} \tag{4}$$

where $\mathcal{R}$ denotes the lower 48 states of USA and $\mathbf{w}_{(X)} = \{w_{(X),s,i,j}\}$ are coefficients from a wavelet transform of $\mathbf{X}$. This implies that c is calculated such that both $\overline{\mathbf{f}_V}$ and $\mathbf{f}_{pr}$ provide the same value for the total emissions for the US. In Fig. 5 (bottom row), we plot the error $(\mathbf{f}_{pr} - \overline{\mathbf{f}_V})$. We see that neither nightlights nor the BUA map provide a $\mathbf{f}_{pr}$ that is an accurate representation of $\overline{\mathbf{f}_V}$, though they share similar spatial patterns i.e., $\mathbf{f}_{pr}$ may be used to provide a guess for $\mathbf{f}$ in an inverse problem, but, by itself, is a poor predictor, regardless of the proxy $\mathbf{X}$ used to create it.

4 Formulation of the estimation problem

In this section, we pose and solve an inverse problem to estimate $ffCO_2$ emissions using the MsRF developed in Sect. 3. The method is new, and uses a sparse reconstruction method that is summarized in Sects. 4.1 and 4.2; full details are in Ray et al. (2013). The inversion technique is most relevant in situations where accurate, finely-gridded $ffCO_2$ inventories are unavailable, and one has to take recourse to easily observable proxies for information on the spatial pattern of $ffCO_2$ emissions.

GMDD
7, 1277–1315, 2014

The inverse problem is predicated on synthetic observations, $\mathbf{y}^{\text{obs}}$, of CO_2 concentrations measured at 35 towers. These are plotted as markers in Fig. 6; see Ray et al. (2013) for their precise locations and names. The measurements are related to $ffCO_2$ emissions described on a finely gridded domain as

$$\mathbf{y}^{\text{obs}} = \mathbf{y} + \epsilon = \mathbf{H}\mathbf{f} + \epsilon, \tag{5}$$

where $\mathbf{H}$ is the transport or sensitivity matrix, obtained from a transport model, $\mathbf{y}$ is the CO_2 concentration predicted by the atmospheric model, which differs from its measured counterpart by an error ϵ. The $ffCO_2$ emissions $\mathbf{f}$ are defined on a grid, and are assumed to be non-zero only within $\mathcal{R}$.

The estimation of CO_2 fluxes, typically biospheric (Nassar et al., 2011; Chatterjee et al., 2012; Gourdji et al., 2012), is usually posed as the minimization of an objective function J,

$$J = (\mathbf{y}^{\text{obs}} - \mathbf{H}\mathbf{f})^{\mathrm{T}}\mathbf{R}_{\mathrm{e}}^{-1}(\mathbf{y}^{\text{obs}} - \mathbf{H}\mathbf{f}) + (\mathbf{f} - \mathbf{f}_{\mathrm{m}})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{f} - \mathbf{f}_{\mathrm{m}}), \tag{6}$$

where $\mathbf{f}_{\mathrm{m}}$ are "prior" (or guessed) fluxes and $\mathbf{R}_{\mathrm{e}}^{-1}$ is a diagonal matrix containing the standard deviation of Gaussian noise used to model measurement error. The discrepancy between the "true" and prior fluxes is modeled as a multivariate Gaussian field, whose covariance $\mathbf{Q}$ is calculated offline. In contrast, in our $ffCO_2$ inversion, $\mathbf{f}$ will be modeled using the MsRF rather than a multivariate Gaussian field. Further, the second term in Eq. (6) is omitted and the effect of the "guessed" or "prior" emissions $\mathbf{f}_{\mathrm{pr}}$ is introduced in a manner that is amenable to sparse reconstruction (see Sect. 4.1). The calculation of the sensitivities $\mathbf{H}$ is described in detail in Gourdji et al. (2012). The elements of the $\mathbf{H}$ matrix are calculated using the Stochastic Time-Inverted Lagrangian Transport Model (Lin et al., 2003), with wind fields from the Weather Research & Forecasting model (Skamarock and Klemp, 2008), version 2.2, driven by 2008 meteorology. Concentration footprints (or sensitivities) were calculated at 3 h intervals by integrating the trajectories over a North American $1^\circ \times 1^\circ$ grid as described in Lin et al. (2003).

GMDD
7, 1277–1315, 2014

The sensitivity of the CO_2 concentration at each observation location due to the flux at each grid-cell (the "footprint") is calculated in units of ppmv μmol^{-1} m^2 s (ppmv: parts per million by volume). ffCO_2 emissions were averaged over 8 day intervals and the sensitivity of **y** to the 8-day-averaged emissions were obtained from the 3 h sensitivities described above by simply adding the $8 \times 24/3 = 64$ sensitivities that span the 8 day period. Thereafter, the grid-cells outside $\mathcal{R}$ were removed to obtain the **H** matrix used in this study. The size of the **H** matrix is $(K_s N_s) \times (N_{\mathcal{R}} K)$, where K_s is the number of tower measurements per year, N_s is the number of sensors/towers, $N_{\mathcal{R}}$ is the number of grid-cells in $\mathcal{R}$, the part of the domain covered by the lower 48 states of the US and K is the number of 8 day periods that constitute the duration over which the emissions are estimated.

4.1 Posing and solving the inverse problem

We denote the spatial distribution of emissions during an arbitrary 8 day period k as $\mathbf{f}_k$. We seek emissions over an entire year i.e., we seek $\mathbf{F} = \{\mathbf{f}_k\}, k = 1 \ldots K$. We will model emissions on the $2^M \times 2^M, M = 6$ mesh with wavelets:

$$\mathbf{f}_k = w'_k \phi' + \sum_{s=1}^{M} \sum_{i,j} w_{s,i,j,k} \phi_{s,i,j}, \quad \{s,i,j\} \in W^{(s)}$$
$$= \mathbf{\Phi} \mathbf{w}_k. \tag{7}$$

Note that $\mathbf{\Phi}$ comprises of only those wavelets selected using **X** and contained in $W^{(s)}$. For the entire year, the expression for emissions becomes $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_K\} = \{\mathbf{\Phi}\mathbf{w}_1, \mathbf{\Phi}\mathbf{w}_2, \ldots \mathbf{\Phi}\mathbf{w}_K\} = \tilde{\mathbf{\Phi}}\mathbf{w}$. Since $\mathbf{\Phi}\mathbf{w}_k$ models the emissions over all grid-cells i.e., over the rectangular region given by the corners (24.5° N, 63.5° W) and (87.5° N, 126.5° W), and not just $\mathcal{R}$, **F** contains emissions over the lower 48 states, as well as the region outside it (where we have assumed that the emissions are non-existent). We separate out the two fluxes by permuting the rows of $\tilde{\mathbf{\Phi}}$

GMDD
7, 1277–1315, 2014

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_{\mathcal{R}} \\ \mathbf{F}_{\mathcal{R}'} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{\Phi}}_{\mathcal{R}} \\ \tilde{\mathbf{\Phi}}_{\mathcal{R}'} \end{pmatrix} \mathbf{w},$$

where $\tilde{\mathbf{\Phi}}_{\mathcal{R}}$ and $\tilde{\mathbf{\Phi}}_{\mathcal{R}'}$ are $(N_{\mathcal{R}}K) \times (LK)$ and $(N_{\mathcal{R}'}K) \times (LK)$ matrices respectively. Here L is the number of wavelets in $W^{(s)}$ and $N_{\mathcal{R}'}$ is the number of grid-cells in $\mathcal{R}'$, the region outside $\mathcal{R}$ but inside the rectangular domain. The modeled concentrations at the measurement towers, caused by $\mathbf{F}_{\mathcal{R}}$, can be written as $\mathbf{y} = \mathbf{H}\mathbf{F}_{\mathcal{R}}$. For arbitrary $\mathbf{w}$, $\mathbf{F}_{\mathcal{R}'}$, the emissions in the region outside $\mathcal{R}$, are not zero. Consequently, it will be necessary to specify $\mathbf{F}_{\mathcal{R}'} = 0$ as a constraint in the inverse problem.

Specifying the constraint $\mathbf{F}_{\mathcal{R}'} = 0$ directly is not very efficient since it leads to $N_{\mathcal{R}'}K$ constraints. In a global inversion, or at resolutions higher than $1^\circ \times 1^\circ$, this could get very large. Consequently, we adapt an approach from compressive sensing to enforce this constraint approximately. Consider a $M_{\mathrm{cs}} \times (N_{\mathcal{R}'}K)$ matrix $\mathbf{R}$, whose rows are direction cosines of random points on the surface of $N_{\mathcal{R}'}K$-dimensional unit sphere. This is called a uniform spherical ensemble (Tsaig and Donoho, 2006). The projection of the emission field $\mathbf{F}_{\mathcal{R}'}$ on $\mathbf{R}$ i.e., $\mathbf{R}\mathbf{F}_{\mathcal{R}'}$ compressively samples $\mathbf{F}_{\mathcal{R}'}$. M_{cs} is the number of such projections or compressive samples. Setting this projection to zero during inversion allows us to enforce zero emissions outside $\mathcal{R}$. However, to do so, we add only M_{cs} constraint equations. The computational savings afforded by imposing the $\mathbf{F}_{\mathcal{R}'} = 0$ constraint in this manner is investigated in Ray et al. (2013).

The equivalent of Eq. (5) is written as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}^{\mathrm{obs}} \\ 0 \end{pmatrix} \approx \begin{pmatrix} \mathbf{H}\tilde{\mathbf{\Phi}}_{\mathcal{R}} \\ \mathbf{R}\tilde{\mathbf{\Phi}}_{\mathcal{R}'} \end{pmatrix} \mathbf{w} = \mathbf{G}\mathbf{w}. \quad (8)$$

We incorporate the spatial patterns in $\mathbf{X}$ into the estimation procedure by using $\mathbf{w}_{(X)}$ to normalize $\mathbf{w}$. Other, less effective, methods were investigated and discarded in

GMDD
7, 1277–1315, 2014

Ray et al. (2013). We rewrite Eq. (8) as

$$\mathbf{Y} \approx \mathbf{G}\,\mathrm{diag}\left(\mathbf{w}_{(X)}\right)\,\mathrm{diag}\left(\mathbf{w}_{(X)}^{-1}\right)\mathbf{w} = \mathbf{G}'\mathbf{w}' = \begin{pmatrix}\mathbf{H}\widetilde{\mathbf{\Phi}}'_{\mathcal{R}} \\ \mathbf{R}\widetilde{\mathbf{\Phi}}'_{\mathcal{R}'}\end{pmatrix}\mathbf{w}', \tag{9}$$

where $\mathbf{w}' = \{w_{s,i,j}/(c\ w_{(X),s,i,j})\}, \{s,i,j\} \in W^{(s)}$, is the normalized set of wavelet coefficients, $\widetilde{\mathbf{\Phi}}'_{\mathcal{R}} = \widetilde{\mathbf{\Phi}}_{\mathcal{R}}\,\mathrm{diag}(\mathbf{w}_{(X)})$ and $\widetilde{\mathbf{\Phi}}'_{\mathcal{R}'} = \widetilde{\mathbf{\Phi}}_{\mathcal{R}'}\,\mathrm{diag}(\mathbf{w}_{(X)})$.

The under-determined system Eq. (9) is solved using optimization. Given the small number of towers (35) and their location (the towers were sited with biospheric fluxes in mind), it may not be possible to estimate all the elements of $\mathbf{w}'$, especially those that contribute to the fine-scale details of $\mathbf{F}_{\mathcal{R}}$. Further, a priori, we do not know the identity of these "un-estimatable" wavelet coefficients in $\mathbf{w}'$. Consequently, we employ a sparse reconstruction method, based on ℓ_1 minimization, that identifies and estimates the elements of $\mathbf{w}'$ that can be constrained by $\mathbf{y}^{\mathrm{obs}}$, while setting the rest to zero. We cast the optimization problem as

$$\underset{\mathbf{w}' \in \mathbb{R}^N}{\text{minimize}}\ \|\mathbf{w}'\|_1, \quad \text{subject to } \|\mathbf{Y} - \mathbf{G}'\mathbf{w}'\|_2^2 < \epsilon_2. \tag{10}$$

This is of the same form as Eq. (2) and is solved using StOMP (Donoho et al., 2012).

4.2 Imposing non-negativity on $ffCO_2$ fluxes

Estimates of $\mathbf{w}'$ calculated by StOMP do not necessarily provide non-negative estimates of $\mathbf{F}_{\mathcal{R}} = \widetilde{\mathbf{\Phi}}_{\mathcal{R}}\mathbf{w}$. In practice negative $ffCO_2$ emissions occur in only a few grid-cells and are usually small in magnitude. We devised an iterative method to impose non-negativity as a post-processing step. We present a summary here; details are in Ray et al. (2013).

We use the StOMP solution to generate $\mathbf{F} = \widetilde{\mathbf{\Phi}}\mathbf{w}$, discard $\mathbf{F}_{\mathcal{R}'}$ and manipulate the emissions $\mathbf{E} = \{E_i\}, i = 1 \ldots N_{\mathcal{R}}K$ in $\mathcal{R}$ directly. We start with a guessed $\mathbf{E}$ ($= |\mathbf{F}_{\mathcal{R}}|$) and

GMDD
7, 1277–1315, 2014

at the mth iteration calculate an increment $\Delta\mathbf{E}^{(m-1)}$ to the current iterate $\mathbf{E}^{(m-1)}$

$$\mathbf{y}^{\text{obs}} - \mathbf{H}\mathbf{E}^{(m-1)} = \Delta\mathbf{y} \approx \mathbf{H}\Delta\mathbf{E}^{(m-1)} \tag{11}$$

This is an under-determined problem, and we seek the sparsest set of increments $\Delta\mathbf{E}^{(m-1)}$ using StOMP. The increment is used to calculate a correction $\boldsymbol{\xi} = \{\xi_i\}, i = 1 \ldots N_{\mathcal{R}}K, |\xi_i| \leq 1$ and update $\mathbf{E}^{(m-1)}$

$$\xi_i = \text{sgn}\left(\frac{\Delta E_i^{(m-1)}}{E_i^{(m-1)}}\right) \max\left(1, \left|\frac{\Delta E_i^{(m-1)}}{E_i^{(m-1)}}\right|\right),$$
$$E_i^{(m)} = E_i^{(m-1)} \exp(\xi_i). \tag{12}$$

The iteration is stopped when $\|\mathbf{y}^{\text{obs}} - \mathbf{H}\mathbf{E}\|_2 / \|\mathbf{y}^{\text{obs}}\|_2 \leq \epsilon_3$ for a small value of ϵ_3.

5 Numerical tests

Numerical tests are performed for the domain between the corners (24.5° N, 63.5° W) and (87.5° N, 126.5° W). It is discretized by a $2^M \times 2^M$, $M = 6$ mesh, with 4096 grid cells. Of these, $N_{\mathcal{R}} = 816$ cells lie inside $\mathcal{R}$, while the rest, $N_{\mathcal{R}'} = 4096 - 816 = 3280$ lie outside in $\mathcal{R}'$. We estimate emissions over $k = 1 \ldots K, K = 45$, i.e., for $45 \times 8 = 360$ days (approximately a year). We generate synthetic observations $\mathbf{y}^{\text{obs}}$ using the $ffCO_2$ emissions in Vulcan, which provides them only in $\mathcal{R}$. Hourly Vulcan fluxes are coarsened from 0.1° resolution to 1°, and averaged to 8 day periods. These fluxes are multiplied by $\mathbf{H}$ to obtain $ffCO_2$ concentrations at the $N_s = 35$ measurement towers. Observations are available every 3 h and span a full year, i.e. we collect $K_s = 24/3 \times 360 = 2880$ observations per tower. A measurement error $\epsilon \sim N(0, \sigma^2)$ is added to the concentrations to obtain $\mathbf{y}^{\text{obs}}$, as used in Eq. (8). The same σ is used for all towers and is set to 0.1 ppmv, as in Chatterjee et al. (2012).

GMDD
7, 1277–1315, 2014

We solve Eq. (10) and enforce non-negativity on $\mathbf{F}_{\mathcal{R}}$ to obtain $\mathbf{E}$. The coefficients $\mathbf{w}_{(\mathbf{X})}$ used in Eq. (9) are obtained from a wavelet decomposition of $\mathbf{f}_{\mathrm{pr}}$ based on nightlights (Sect. 3). The constant c in Eq. (4) is obtained by using fluxes from the Emission Database for Global Atmospheric Research (EDGAR, http://edgar.jrc.ec.europa.eu; Olivier et al., 2005) for 2005 i.e., instead of using emissions from Vulcan to calculate $\overline{\mathbf{f}_{\mathrm{V}}}$, we use EDGAR. The following parameters are used in the inversion process (Sect. 4.2): $\epsilon_2 = 10^{-5}, \epsilon_3 = 5.0 \times 10^{-4}, M_{\mathrm{cs}} = 13\,500$ i.e., 300 compressive samples for each 8 day period. The initial guess for $\mathbf{w}'$ in Eq. (10) is zero.

In Fig. 6, we plot the true and reconstructed emissions for the 33rd 8 day period ($k = 33$). We see that the reconstruction in the NE quadrant is qualitatively similar to the true emissions. In contrast, the reconstruction on the West coast contains significant inaccuracies. For example, we see that the Los Angeles–San Diego region (southwest quadrant) is estimated wrongly. The estimated emissions in the center of the country (Continental Divide and Great Plains, in the western quadrants) show similar errors, as well as far more structure than the true ffCO_2 emissions. The region around the Gulf of Mexico is also not well estimated. The quality of the reconstruction in the various regions correlate with the density of observations towers, though the wind fields too play an important part. In the regions where the observations are not very informative, the impact of normalization by $\mathbf{f}_{\mathrm{pr}}$ is clear as some of its structure is retained in the estimated emissions. These errors are almost entirely at fine spatial scales.

In Fig. 7 (left) we plot a time-series of errors defined as a percentage of total, country-level Vulcan emissions. Percent errors in reconstructed emissions and $\mathbf{f}_{\mathrm{pr}}$ are calculated using Eq. (13).

GMDD
7, 1277–1315, 2014

$$\text{Error}_k\ (\%) = \frac{100}{K}\sum_{k=1}^{K}\frac{E_k - E_{\text{V},k}}{E_{\text{V},k}}, \quad \text{where } E_k = \int_{\mathcal{R}} \mathbf{E}_k \mathrm{d}A \text{ and } E_{\text{V},k} = \int_{\mathcal{R}} \mathbf{f}_{\text{V},k}\mathrm{d}A,$$

$$\text{Error}_{\text{pr},k}\ (\%) = \frac{100}{K}\sum_{k=1}^{K}\frac{E_{\text{pr}} - E_{\text{V},k}}{E_{\text{V},k}}, \quad \text{where } E_{\text{pr}} = \int_{\mathcal{R}} \mathbf{f}_{\text{pr}}\mathrm{d}A. \tag{13}$$

Here, $\mathbf{f}_{\text{V},k}$ are Vulcan emissions averaged over the kth 8 day period and $\mathbf{E}_k$ are the non-negativity enforced emission estimates in the same time period. A positive error denotes an overestimation by the inverse problem. We see 25 % errors in $\mathbf{f}_{\text{pr}}$. The large error is a consequence not only of the disagreement between EDGAR (from 2005) and Vulcan (from 2002), but also the manner in which they account for emissions. As can be seen, assimilation of $\mathbf{y}^{\text{obs}}$ reduces the error significantly vis-à-vis $\mathbf{f}_{\text{pr}}$. The least accurate reconstructions are during spring (k = 10–15). In order to check the accuracy of the spatial distribution of $\mathbf{E}_k$, we calculate the Pearson correlations $\rho(\mathbf{E}_k, \mathbf{f}_{\text{V},k})$ and $\rho(\mathbf{f}_{\text{pr}}, \mathbf{f}_{\text{V},k})$. We see that data assimilation results in a clear increase in the correlation. When the emissions are aggregated/averaged over 32 day periods, the correlation increases to about 0.85, whereas the "prior" correlation was around 0.7. Thus the ffCO_2 emissions obtained using a nightlight proxy are substantially improved by the incorporation of $\mathbf{y}^{\text{obs}}$. Only about half the wavelet coefficients could be estimated; the rest were set to zero by the sparse reconstruction technique (Ray et al., 2013).

We next investigate the effect of using BUA maps, instead of nightlights, as the proxy. Changing the proxy results in a different set of wavelets being chosen (nightlights resulted in a $W^{(s)}$ of 1031 wavelets; the corresponding number for BUA was 1049); further one was *not* a strict subset of the other. It also results in a different normalization in Eq. (9). The inversion was performed in a manner identical to that adopted for the nightlight proxy. In Fig. 7 (left) we see that the ffCO_2 emissions developed using nightlights and BUA as proxies are similar, as measured by reconstruction error (Eq. 13), though the BUA reconstruction error tends to be slightly smaller. The aggregated error

GMDD
7, 1277–1315, 2014

between the true and "prior" fluxes remain unchanged (nightlights vs. BUA) since it just reflects the difference between EDGAR (in 2005) and Vulcan (in 2002) inventories. In Fig. 7 (right) we plot the spatial correlation between the true, reconstructed and "prior" fluxes. The correlation between true and reconstructed emissions (from BUA) tends to be worse than the nightlight reconstruction. The correlation of $\mathbf{f}_{pr}$ with true emissions from Vulcan are different for nightlights and BUA reflecting the distinct spatial difference between them as seen in Fig. 5. This results in the difference between the two dashed lines. Averaging over 32 day intervals improves the correlation and makes them almost indistinguishable from those obtained using nightlights.

In Fig. 8 we investigate the differences between the nightlight- and BUA-based reconstructions at the quadrant level. We see in Fig. 8 (left) that the difference between nightlight- and BUA-based reconstruction errors in the NE quadrant are smaller than those for the NW quadrant. Thus, while the $\mathbf{f}_{pr}$ from nightlights and BUA are quite different (see the last row of Fig. 5), the estimated emissions are well informed by $\mathbf{y}^{obs}$ in the NE quadrant and the impact of the proxies is small. This is not the case for the NW quadrant, where the reconstruction based on BUA is clearly much worse than the nightlight-based reconstruction. This is not surprising given the paucity of towers there (see Fig. 6), which increases the impact of $\mathbf{f}_{pr}$. In Fig. 8 (right) we plot the correlation of the reconstructed and true emissions in the NE and NW quadrants. We see that there is little to choose between the correlations generated using nightlight- vs. the BUA-based emission estimates. Again, due to the larger density of towers in the NE, the correlations are higher there. Thus, while Fig. 5 (middle row) showed that BUA-had a slightly better correlation with true (Vulcan) emissions, its larger errors, as seen in Fig. 5 (bottom row) lead to a less accurate reconstruction. This result is also a testament to the inadequacy of $\mathbf{y}^{obs}$ over the whole country; had there been sufficient data to inform $\mathbf{E}$, the impact of $\mathbf{f}_{pr}$ would have been minimal.

Next, in Fig. 9 (left) we compare the estimated emissions for the 34th 8 day period developed from the two competing prior models. We plot the difference between the two estimates; it shows differences spread over a large area, though their magnitudes

GMDD
7, 1277–1315, 2014

are not very big. Thus the "prior" model has a measurable impact on the spatial distribution of the emissions. In Fig. 9 (right) we plot **y** predicted by the reconstructed emissions (from nightlights as priors) at 3 towers. The towers were chosen to represent the range of the ffCO_2 signal strengths encountered in our test cases. We see that the ffCO_2 concentrations are well reproduced by the estimated emissions. Further, note that the measurement noise (σ = 0.1 ppmv) is relatively large compared to some of the observations. Thus, the lack of fidelity at the smaller scales (seen in Fig. 6) do not substantially impact the measurements. This is due to the weak strengths of the erroneous emission sources (while a few may be intense, they are present only over a small area) and their distance from the towers.

6 Conclusions

We have devised a multiresolution parametrization (also known as a multiscale random field or MsRF model) for modeling ffCO_2 emissions at 1° resolution. The MsRF models emissions in the lower 48 states of the US and is designed for use in atmospheric inversions. The parameterization employs Haar wavelets which provide a sparser representation than other smoother wavelets with wider support. This is the first "abstract" parameterization i.e., a RF model for spatially resolved ffCO_2 emissions.

The dimensionality of the MsRF was reduced by judiciously selecting its component Haar wavelets using proxies of human activity, and therefore indicative of ffCO_2 emissions. We developed two MsRFs based on images of lights at night and maps of built-up areas. The former had a slightly lower dimensionality but was not a strict subset of the latter. The MsRF models were also used to develop two approximate emission models that differed in their fine-spatial details.

The MsRF model was tested in a synthetic-data inversion. Time-dependent ffCO_2 emissions, averaged over 8 day periods, were estimated for a 360 day period from measurements of ffCO_2 concentrations at 35 towers. These observations were sufficient only for estimating about half the wavelets retained in the MsRF model. We used

GMDD
7, 1277–1315, 2014

a sparse reconstruction technique, namely Stagewise Matching Orthogonal Pursuit (StOMP), to identify and estimate wavelet coefficients in MsRF that could be informed by the available data. The StOMP estimates were not necessarily non-negative (as $ffCO_2$ emissions are required to be) and we devised an iterative, post-processing procedure to impose non-negativity. Furthermore, the MsRF actually models emissions in a rectangular region and constraints had to be imposed during the inversion to ensure that emissions were restricted to the lower 48 states. To our knowledge, this is the first instance of using a sparse reconstruction method in atmospheric CO_2 inversions. The algorithmic details of the inversion procedure are in Ray et al. (2013). This is also the first use of an MsRF model in $ffCO_2$ emission estimation.

The inversions using the competing MsRF models (nightlights- vs. built-up-area-based) resulted in emission estimates that differed in their fine-scale details. Country-level estimates of emissions, and their correlation with the true emissions, differed little when developed using either of the two MsRFs. Further, emissions obtained by dividing the US into quadrants showed large errors vis-à-vis true emissions. This is a testament to the inadequacy of the observations to provide fine-scale coverage over the entire US.

Primarily, our work demonstrated that observations of $ffCO_2$ concentrations could be used to update models of $ffCO_2$ emissions and improve their accuracy (see Fig. 7). We assumed that these measurements could be obtained from existing towers, which were sited with a view of reconstructing biospheric, rather than $ffCO_2$, fluxes. This accounts for some of our inability to estimate fine-scale spatial structure in the emissions.

Our inversion formulation suffers from three drawbacks. It requires knowledge of $ffCO_2$ concentrations at the measurement towers, typically estimated from $^{14}CO_2$ or CO measurements. Such measurements are more difficult and expensive than CO_2 concentration measurements. The second drawback is the deterministic nature of the reconstruction – we do not provide error bounds on the estimates of the MsRF parameters (the wavelet coefficients). This can be rectified by adopting a Bayesian approach e.g., Kalman filters, which is underway. Finally, we have demonstrated our MsRF model

GMDD
7, 1277–1315, 2014

by neglecting biospheric CO_2 fluxes; in an inversion based on true CO_2 concentration observations, they would have to be modeled too. The success at modeling them using multiGaussian fields indicates that they could be accommodated by augmenting our collection of wavelets with the Karhunen–Loève modes of the multiGaussian fields currently used for the purpose. This would render our MsRF a dictionary rather than an orthogonal basis set. Sparse reconstruction methods that use dictionaries exist; see Jafarpour (2013) for examples and references on the inference of permeability fields from observations of water transport.

In conjunction with this paper, we are also providing the MATLAB® code required to (1) reduce the dimensionality of the MsRF using nightlights and (2) perform the inversion using synthetic observations at our website (Ray, 2013). The website also contains links to the (free) MATLAB toolkits that our code depends on, along with a user's manual.

Acknowledgements. Thanks are due to Sharon Gourdji of Stanford University, and Kim Mueller and Abhishek Chatterjee of University of Michigan, Ann Arbor for fruitful discussions regarding algorithms, the challenges of working with field observations as well as some of the data used in this work. 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.

References

Andres, R. J., Boden, T. A., Bréon, F.-M., Ciais, P., Davis, S., Erickson, D., Gregg, J. S., Jacobson, A., Marland, G., Miller, J., Oda, T., Olivier, J. G. J., Raupach, M. R., Rayner, P., and Treanton, K.: A synthesis of carbon dioxide emissions from fossil-fuel combustion, Biogeosciences, 9, 1845–1871, doi:10.5194/bg-9-1845-2012, 2012. 1279, 1280

GMDD
7, 1277–1315, 2014

Auger, L. and Tangborn, A. V.: A wavelet-based reduced rank Kalman filter for assimilation of stratospheric chemical tracer observations, Mon. Weather Rev., 132, 1220–1237, 2002. 1284

Brioude, J., Petron, G., Frost, G. J., Ahmadov, R., Angevine, W. M., Hsie, E.-Y., Kim, S.-W., Lee, S.-H., McKeen, S. A., Trainer, M., Fehsenfeld, F. C., Holloway, J. S., Peischl, J., Ryerson, T. B., and Gurney, K. R.: A new inversion method to calculate emission inventories without a prior at mesoscale: application to the anthropogenic CO_2 emissions from Houston, Texas, J. Geophys. Res., 117, D05312, doi:10.1029/2011JD016918, 2012. 1280

Candes, E. and Tao, T.: Near optimal signal recovery from random projections: universal encoding strategies?, IEEE T. Inform. Theory, 52, 5406–5425, 2006. 1285

Candes, E. and Wakin, M.: An introduction to compressive sampling, IEEE Signal Proc. Mag., 25, 21–30, 2008. 1284

Candes, E., Romberg, J., and Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE T. Inform. Theory, 52, 489–509, 2006. 1285

Chan, T. F. and Shen, J. J.: Image Processing and Analysis – Variational, PDE, Wavelet, and Stochastic Methods, Society of Applied Mathematics, 2005. 1281

Chatterjee, A., Michalak, A. M., Mueller, K. L., Anderson, J. L., and Yadav, V.: Towards reliable ensemble Kalman filter estimates, J. Geophys. Res.-Atmos., 117, D22306, doi:10.1029/2012JD018176, 2012. 1291, 1295

Chen, S., Donoho, D., and Saunders, M.: Atomic decomposition by basis pursuit, SIAM J. Sci. Comput., 20, 33–61, 1998. 1285

Ciais, P., Rayner, P., Chevallier, F., Bousquet, P., Logan, M., Peylin, P., and Ramonet, M.: Atmospheric inversions for estimating CO_2 fluxes: methods and perspectives, Climate Change, 103, 69–92, 2010. 1280

Cinzano, P., Falchi, F., Elvidge, C. D., and Baugh, K. E.: The artificial night sky brightness mapped from DMSP satellite Operational Linescan System measurements, Mon. Not. R. Astron. Soc., 318, 641–657, 2000. 1289

Doll, C. N. H., Muller, J.-P., and Elvidge, C. D.: Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions, Ambio, 29, 157–162, 2000. 1279, 1289

Donoho, D. L.: Denoising by soft thresholding, IEEE T. Inform. Theory, 41, 613–627, 1995. 1285

Donoho, D.: Compressed sensing, IEEE T. Inform. Theory, 52, 1289–1306, 2006. 1285

GMDD

7, 1277–1315, 2014

Donoho, D. L., Tsaig, Y., Drori, I., and Starck, J.-L.: Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit, IEEE T. Inform. Theory, 58, 1094–1121, 2012. 1285, 1294

Duarte, M. F., Wakin, M. B., and Baraniuk, R. G.: Wavelet-domain compressive signal reconstruction using a hidden Markov tree mode, in: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, NV, 2008. 1284

Friedlingstein, P.: Climate–carbon cycle feedback analysis: results from the C4MIP model intercomparison, J. Climate, 19, 3337–3353, 2006. 1278

Gholami, A. and Siahkoohi, H. R.: Regularization of linear and nonlinear geophysical ill-posed problems with joint sparsity priors, Geophys. J. Int., 180, 871–882, 2010. 1285

Goldstein, T. and Osher, S.: The Split Bregman Method for L1-Regularized Problems, SIAM Journal on Imaging Sciences, 2, 323–343, 2009. 1285

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 CO_2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, doi:10.5194/bg-9-457-2012, 2012. 1291, 1307

Gregg, J. S., Andres, R. J., and Marland, G.: China: emission pattern of the world leader in CO_2 emissions from fossil fuel consumption and cement production, Geophys. Res. Lett., 35, L08806, doi:10.1029/2007GL032887, 2008. 1279

Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., and de la Rue de Can, S.: High resolution fossil fuel combustion CO_2 emission fluxes in the United States, Environ. Sci. Technol., 43, 5535–5541, 2009. 1279

Gurney, K. R., Razlivanov, I., Song, Y., Zhou, Y., Benes, B., and Abdul-Massih, M.: Quantification of fossil fuel CO_2 emissions on the building/street scale for a large city, Environ. Sci. Technol., 46, 12194–12202, 2012. 1280

He, L. and Carin, L.: Exploiting structure in wavelet-based Bayesian compressive sensing, IEEE T. Signal Proc., 57, 3488–3497, 2009. 1284

Jafarpour, B.: Wavelet reconstruction of geologic facies from nonlinear dynamic flow measurements, IEEE T. Geosci. Remote, 49, 1520–1535, 2011. 1284

Jafarpour, B.: Sparsity-promoting solution of subsurface flow model calibration inverse problems, in: Advances in Hydrogeology, edited by: Mishra, P. K. and Kuhlman, K. L., Springer, 2013. 1285, 1301

GMDD
7, 1277–1315, 2014

Li, L. and Jafarpour, B.: A sparse Bayesian framework for conditioning uncertain geologic models to nonlinear flow measurements, Adv. Water Resour., 33, 1024–1042, 2010. 1285

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, 4493, doi:10.1029/2002JD003161, 2003. 1291

Loris, I., Nolet, G., Daubechies, I., and Dahlen, F. A.: Tomographic inversion using ℓ_1-norm regularization of wavelet coefficients, Geophys. J. Int., 170, 359–370, 2007. 1285

Mallat, S. and Zhang, Z.: Matching pursuit with time-frequency dictionaries, IEEE T. Signal Proc., 41, 3397–3415, 1993. 1285

McKain, K., Wofsy, S., Nehrkorn, T., Eluszkiewicz, J., Ehrlinger, J. R., and Stephens, B. B.: Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region, Proc. Natl. Acad. Sci. USA, 109, 8423–8428, 2012. 1280

Nassar, R., Jones, D. B. A., Kulawik, S. S., Worden, J. R., Bowman, K. W., Andres, R. J., Suntharalingam, P., Chen, J. M., Brenninkmeijer, C. A. M., Schuck, T. J., Conway, T. J., and Worthy, D. E.: Inverse modeling of CO_2 sources and sinks using satellite observations of CO_2 from TES and surface flask measurements, Atmos. Chem. Phys., 11, 6029–6047, doi:10.5194/acp-11-6029-2011, 2011. 1291

Oda, T. and Maksyutov, S.: A very high-resolution (1 km × 1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys., 11, 543–556, doi:10.5194/acp-11-543-2011, 2011. 1279

Olivier, J. G. J., Aardenne, J. A. V., Dentener, F. J., Pagliari, V., Ganzeveld, L. N., and Peters, J. A. H. W.: Recent trends in global greenhouse gas emissions: regional trends 1970–2000 and spatial distribution of key sources in 2000, Journal of Integrative Environmental Science, 2, 81–99, 2005. 1296

Pacala, S. W., Breidenich, C., Brewer, P. G., Fung, I. Y., Gunson, M.R., Heddle, G., Law, B., Marland, G., Paustian, K., Prather, M., Randerson, J. T., Tans, 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, The National Academies Press, available at: http://www.nap.edu/openbook.php?record_id=12883 (last access: 3 February 2014), 2010. 1279

Pregger, T., Scholtz, Y., and Friedrich, R.: Documentation of the anthropogenic GHG emission data for Europe provided in the frame of CarboEurope GHG and CarboEurope IP, Final

GMDD
7, 1277–1315, 2014

Report, Tech. rep., Institute of Energy Economics and the Rational Use of Energy, University of Stuttgart, 2007. 1279

Raupach, M. R., Rayner, P. J., and Page, M.: Regional variations in spatial structure of nightlights, population density and fossil-fuel CO_2 emissions, Energ. Policy, 38, 4756–4764, 2009. 1280, 1289

Ray, J.: Estimating ffCO_2 using a MsRF and sparse reconstruction, available at: http://www.sandia.gov/~jairay/software.html (last access: 3 February 2014), 2013. 1301

Ray, J., Lee, J., Lefantzi, S., Yadav, V., Michalak, A. M., Bloemen-Waanders, B., and McKenna, S. A.: A multiresolution spatial parametrization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions, SAND Report SAND2013-2919, Sandia National Laboratories, Livermore, CA 94551-0969, available at: http://www.sandia.gov/~jairay (last access: 3 February 2014), 2013. 1290, 1291, 1293, 1294, 1297, 1300

Rayner, P. J., Raupach, M. R., Paget, M., Peylin, P., and Koffi, E.: A new global gridded data set of CO_2 emissions from fossil fuel combustion: methodology and evaluation, J. Geophys. Res., 115, D19306, doi:10.1029/2009JD013439, 2010. 1279, 1280

Romberg, J. K., Choi, H., and Baraniuk, R. G.: Bayesian tree-structured image modeling using wavelet-domain Hidden Markov Models, IEEE T. Image Process., 10, 1056–1068, 2001. 1284

Simons, F. J., Loris, I., Nolet, G., Daubechies, I. C., Voronin, S., Judd, J. S., Vetter, P. A., Vetter, P. A., Charlety, J., and Vonesch, C.: Solving or resolving global tomographic models with spherical wavelets and the scale and sparsity of seismic heterogeneity, Geophys. J. Int., 187, 969–988, 2011. 1285

Skamarock, W. C. and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, J. Comput. Phys., 227, 3465–3485, 2008. 1291

Strang, G. and Nguyen, T.: Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley, MA, 1997. 1281

Streets, D. G., Zhang, Q., Wang, L., He, K., Hao, J., Wu, Y., Tang, Y., and Carmichael, G. R.: Revisting China's CO emissions after transport and chemical evolution over the Pacific: synthesis of inventories, atmospheric modeling and observations, J. Geophys. Res., 111, D14306, doi:10.1029/2006JD007118, 2006. 1280

Taubman, D. and Marcellin, M.: JPEG2000: Image Compression Fundamentals, Standards and Practice, Springer, 2002. 1284

GMDD
7, 1277–1315, 2014

Tropp, J. and Gilbert, A. C.: Signal recovery from partial information via orthogonal matching pursuit, IEEE T. Inform. Theory, 53, 4655–4666, 2007. 1285

Tsaig, Y. and Donoho, D.: Extensions of compressed sensing, Signal Process., 86, 533–548, 2006. 1293

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, Atmos. Chem. Phys., 11, 705–721, doi:10.5194/acp-11-705-2011, 2011. 1280

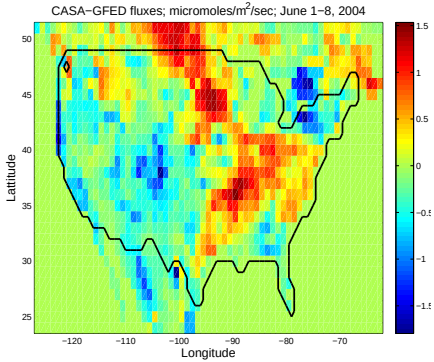
Walker, J. S.: A Primer on Wavelets and Their Scientific Applications, Chapman and Hall/CRC, 2008. 1283

Welstead, S.: Fractal and wavelet image compression techniques (SPIE tutorial texts in optical engineering Vol. TT40), SPIE Press, Bellingham, WA USA, 1999. 1284

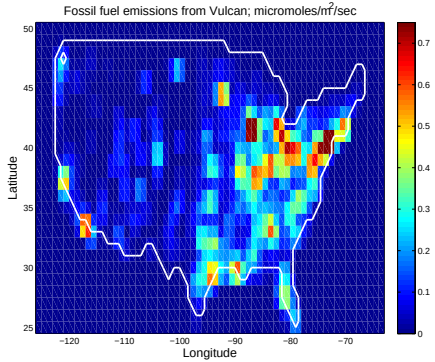
Williams, J. R. and Amaratunga, K.: Introduction to wavelet engineering, Int. J. Numer. Meth. Eng., 37, 2365–2388, 1994. 1283

GMDD

7, 1277–1315, 2014



(a) Biospheric fluxes



(b) $ffCO_2$ emissions

Fig. 1. Differences in the spatial distribution of biospheric (left) and fossil-fuel (right) CO_2 fluxes. The biospheric fluxes are stationary, whereas $ffCO_2$ emissions are non-stationary and correlated with human habitation. The biospheric fluxes are for 1–8 June 2004, obtained from CASA-GFED (http://www.globalfiredata.org/index.html). The post-processing steps to obtain the fluxes as plotted are described in Gourdji et al. (2012). The units of fluxes/emissions are $\mu mol\,s^{-1}\,m^{-2}$ of C. The $ffCO_2$ emissions are calculated by spatiotemporal averaging of the Vulcan inventory. Note the different colormaps; $ffCO_2$ emissions can assume only non-negative values.

GMDD
7, 1277–1315, 2014

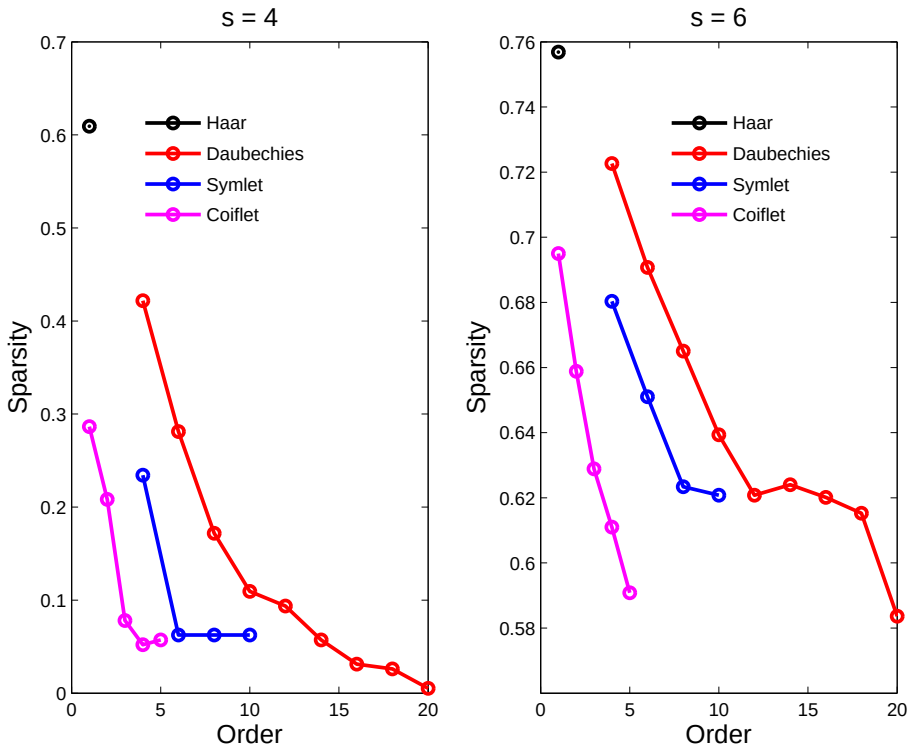


Fig. 2. Sparsity of representation at scale $s = 4$ (left) and $s = 6$ (right) for a combination of wavelet families and orders. We find that Haars provide the sparsest representation.

GMDD
7, 1277–1315, 2014

A spatial parameterization for fossil-fuel carbon dioxide emissions

J. Ray et al.

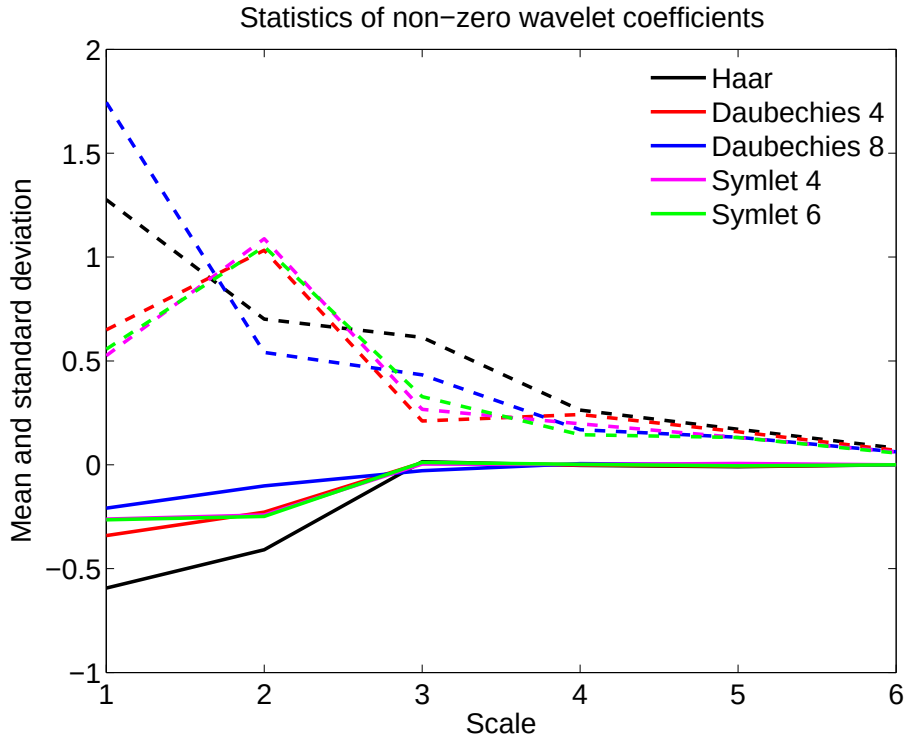


Fig. 3. We plot the average value of the non-zero coefficients (solid lines) and their standard deviation (dashed line), at different scales s, when $\overline{\mathbf{f}_V}$ is subjected to wavelet transforms using Haars, Daubechies 4 and 6 and Symlet 4 and 6 wavelets. We find that while Haars may provide the sparsest representation (Fig. 2), the non-zero values tend to be large and distinct.

GMDD
7, 1277–1315, 2014

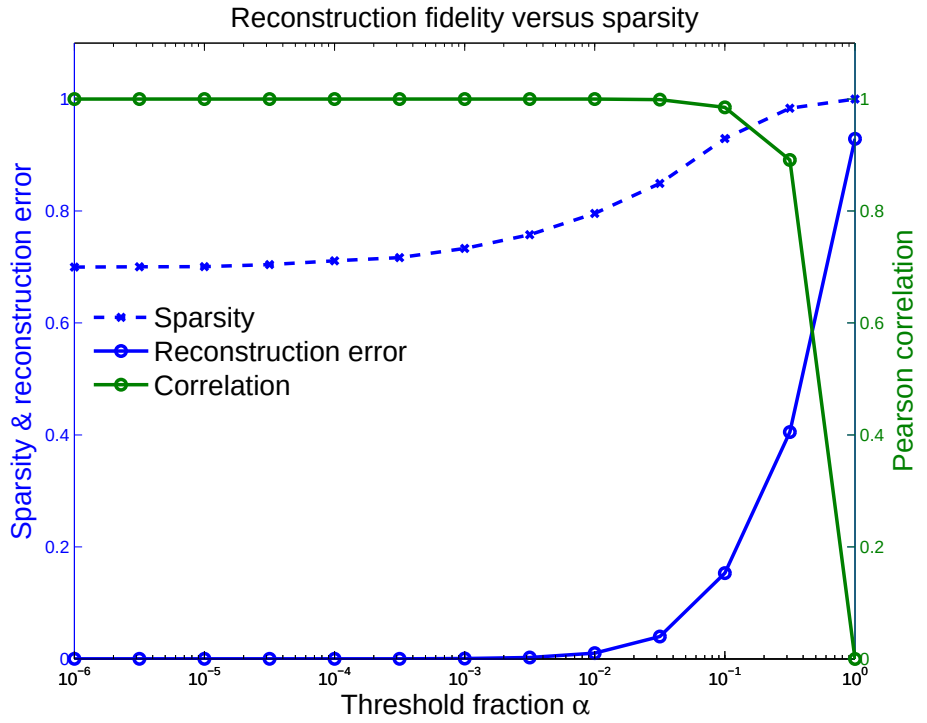


Fig. 4. Variation of sparsity, reconstruction error ϵ and the Pearson correlation between the true and reconstructed $\overline{\mathbf{f}_V}$ i.e. $\rho(\overline{\mathbf{f}_V}, \overline{\mathbf{f}_V}')$ as a function of α.

GMDD
7, 1277–1315, 2014

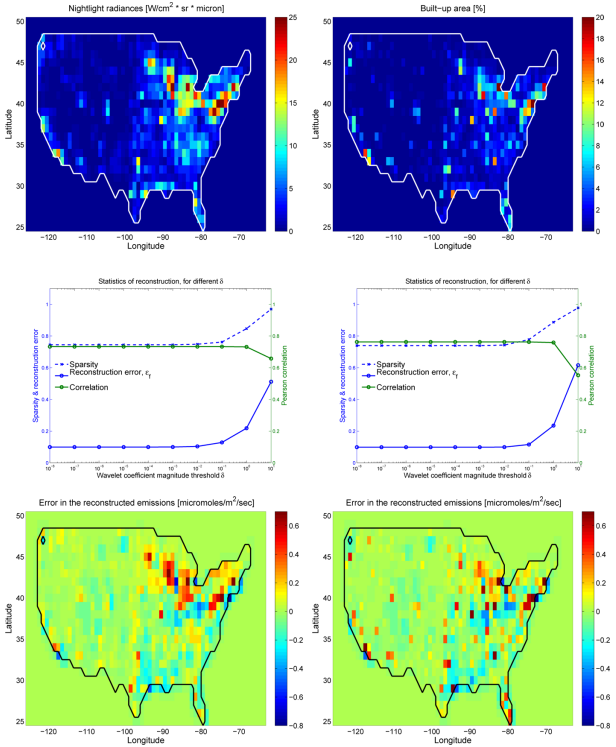


Fig. 5. Top row: Maps of nightlight radiances (left) and BUA percentage (right), for the US. Middle row: The sparsity of representation, the correlation between **X** and $\overline{\mathbf{f}_V}$ and the normalized error ϵ_f between the Vulcan emissions $\overline{\mathbf{f}_V}$ and the sparsified form obtained by projecting it on **X**. These values are plotted for nightlights (left) and the BUA maps (right). Bottom row: Plots of ($\mathbf{f}_{pr} - \overline{\mathbf{f}_V}$) obtained from nightlights (left) and BUA maps (right).

GMDD
7, 1277–1315, 2014

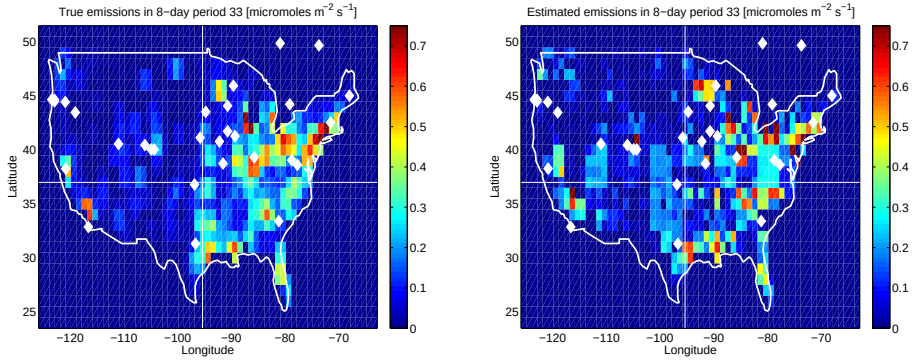


Fig. 6. Reconstruction of the $ffCO_2$ emissions from the 35 towers (plotted as diamonds) over one year. The true emissions are on the left and the reconstructions on the right. We see that the large scale structure of the emissions have been captured. The figures represent emissions for $k = 33$ (end of August). The west coast of the US has few towers near heavily populated regions and thus is not very well estimated. On the other hand, due to the higher density of towers in the Northeast, the true and estimated emissions are qualitatively similar. Emissions have units of $\mu mol\, m^{-2}\, s^{-1}$ of C (not CO_2).

GMDD
7, 1277–1315, 2014

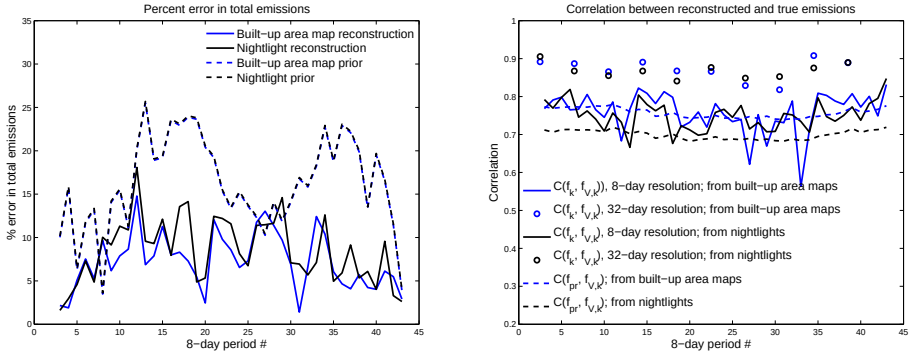


Fig. 7. Comparison of reconstruction error and correlations. Left: We plot the error between the reconstructed and true (Vulcan) emissions in black (using nightlights as priors) and in blue (using BUA priors). We plot the error between $\mathbf{f}_{pr}$ and Vulcan emissions using dashed lines – black for nightlights and blue for BUA. We see that assimilation of $\mathbf{y}^{obs}$ leads to significantly improved accuracy vis-à-vis $\mathbf{f}_{pr}$. Right: We plot the accuracy of the spatial distribution of the reconstructed emissions. The Pearson correlations $\rho(\mathbf{E}_k, \mathbf{f}_{V,k})$ and $\rho(\mathbf{f}_{pr}, \mathbf{f}_{V,k})$ show that incorporating $\mathbf{y}^{obs}$ clearly improves the spatial agreement of estimated emissions vs. the true one when using nightlights, though the results are less clear for BUA priors. If the emissions are averaged over 32 day periods, rather than 8 day periods, the correlation with true (Vulcan) emissions rises to around 0.85, irrespective of the prior used.

GMDD
7, 1277–1315, 2014

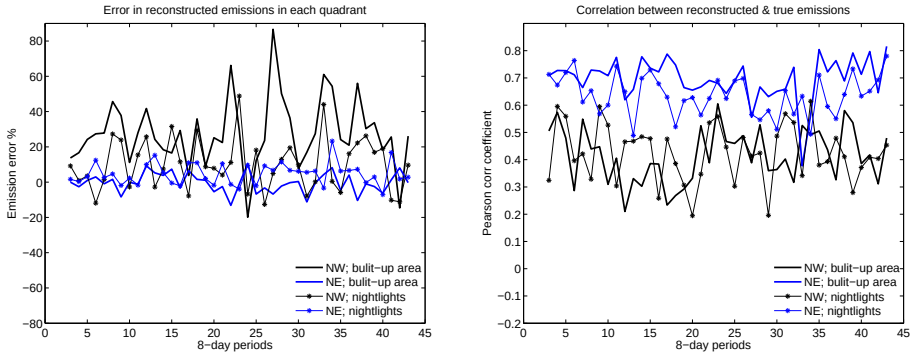


Fig. 8. Left: Emission reconstruction error in the NE (blue) and NW (black) quadrants, when performed with BUA (line) and nightlights (symbols) as proxies. We see that the NW quadrant is very badly constrained and the BUA-based estimates have very large errors. The errors in the NE quadrant are far smaller and very similar when generated using the competing proxies. Right: The comparison of correlations between true and reconstructed emissions shows similar trends.

GMDD
7, 1277–1315, 2014

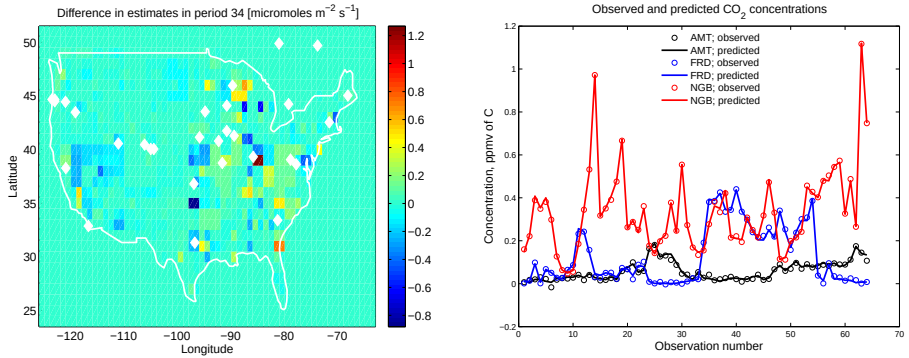


Fig. 9. Left: Comparison of emission estimates developed using $\mathbf{f}_{pr}$ constructed from night-light radiances and BUA maps. We plot the difference between the two estimates. We see that differences are not localized in any one area. Right: Prediction of $ffCO_2$ concentrations at 3 measurement locations, using the true (Vulcan; plotted with symbols) and reconstructed emissions (blue lines) over an 8 day period (Period no. 31). Observations occur every 3 h. We see that the concentrations are accurately reproduced by the estimated emissions.

GMDD
7, 1277–1315, 2014