

## A new global gridded data set of CO<sub>2</sub> emissions from fossil fuel combustion: Methodology and evaluation

P. J. Rayner,<sup>1,2</sup> M. R. Raupach,<sup>3</sup> M. Paget,<sup>3</sup> P. Peylin,<sup>1</sup> and E. Koffi<sup>1</sup>

Received 27 October 2009; revised 20 April 2010; accepted 27 April 2010; published 7 October 2010.

[1] We describe a system for constraining the spatial distribution of fossil fuel emissions of CO<sub>2</sub>. The system is based on a modified Kaya identity which expresses emissions as a product of areal population density, per capita economic activity, energy intensity of the economy, and carbon intensity of energy. We apply the methodology of data assimilation to constrain such a model with various observations, notably, the statistics of national emissions and data on the distribution of nightlights and population. We hence produce a global, annual emission field at 0.25° resolution. Our distribution of emissions is smoother than that of the population downscaling traditionally used to describe emissions.

Comparison with the Vulcan inventory suggests that the assimilated product performs better than downscaling for distributions of either population or nightlights alone for describing the spatial structure of emissions over the United States. We describe the complex structure of uncertainty that arises from combining pointwise and area-integrated constraints. Uncertainties can be as high as 50% at the pixel level and are not spatially independent. We describe the use of <sup>14</sup>CO<sub>2</sub> measurements to further constrain national emissions. Their value is greatest over large countries with heterogeneous emissions. Generated fields may be found online (<http://ffdas.org/>).

**Citation:** Rayner, P. J., M. R. Raupach, M. Paget, P. Peylin, and E. Koffi (2010), A new global gridded data set of CO<sub>2</sub> emissions from fossil fuel combustion: Methodology and evaluation, *J. Geophys. Res.*, 115, D19306, doi:10.1029/2009JD013439

### 1. Introduction

[2] Motivated by concern about rising atmospheric greenhouse gas concentrations, national data on CO<sub>2</sub> emissions from fossil fuels are widely available from data sets such as those of the International Energy Agency (IEA; <http://www.iea.org/>) and the Carbon Dioxide Information and Analysis Center (CDIAC [Marland *et al.*, 2006]). The broad regional and temporal patterns of emissions drivers are also known [e.g., Raupach, 2007]. However, the finer-scale structure of emissions is not nearly as well known.

[3] Improved knowledge of the distribution of emissions at fine space and time scales is needed to understand the contemporary, strongly anthropogenically disturbed carbon cycle. In the “atmospheric inverse” approach [e.g., Enting, 2002], CO<sub>2</sub> sources and sinks are estimated from atmospheric measurements of CO<sub>2</sub> and other gas concentrations together with transport and process models in inverse mode. This approach is providing increasing insight on the large-scale structure of terrestrial and ocean CO<sub>2</sub> sinks [e.g., Gurney *et al.*, 2002; Baker *et al.*, 2006; Stephens *et al.*,

2007; Rayner *et al.*, 2008]. Inversions use estimates of fossil fuel emissions either as part of the prior estimate of flux or to interpret the posterior flux.

[4] There are several existing maps of the fine structure of emissions. For atmospheric-transport estimates of CO<sub>2</sub> sources and sinks, Tans *et al.* [1990] downscaled spatially coarse, nationally aggregated emissions data [Marland *et al.*, 1989] with gridded global population data to provide finer spatial structure. Disaggregation by population is also used in more recent data sets such as the Emissions Database for Global Atmospheric Research (EDGAR) [Olivier *et al.*, 2005]. More recently, global observations of nightlights [Doll *et al.*, 2000] have likewise been used to provide fine-scale spatial structure to nationally aggregated emissions data. In contrast with global-scale approaches, several studies have used intensive data to produce highly detailed maps for specific regions, notably the “Vulcan” study [Gurney *et al.*, 2009] for the United States. Such regional studies combine data on power plant locations and loads, transport patterns, locations and patterns of industrial and residential energy use, and other sectoral data from which emissions are derived.

[5] All of the above approaches face difficulties. Downscaling national emissions data with population or nightlights data faces the challenge that these two maps do not agree perfectly (there are some bright areas with few people, and some dark areas with many people). More fundamentally, spatial structure of emissions at a sufficiently fine scale is not proportional to either population or nightlights because emis-

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement–Institut Pierre-Simon Laplace, Commissariat B l'Energie Atomique–Université de Versailles Saint-Quentin-en-Yvelines, CNRS, Gif-sur-Yvette, France.

<sup>2</sup>Now at School of Earth Sciences, University of Melbourne, Parkville, Victoria, Australia.

<sup>3</sup>CSIRO Marine and Atmospheric Research, Canberra, ACT, Australia.

sions (for instance from coal-fired power plants) are not exactly coincident with where people live or lights are on. This problem is overcome in regional studies such as Vulcan or Edgar by using large bases of geographically and sectorally explicit data which are available regionally (for the United States and Europe, respectively) but not globally, so that production of global maps by these methods is not yet possible.

[6] The purpose of this work is to develop a new methodology for estimating emissions at fine space-time scales, together with uncertainties on those estimates, by combining all available sources of information. The approach is to assimilate multiple classes of data into a simple model for emissions, thereby constraining a set of parameters in the model to produce best overall agreement with all data. This approach, generically known as “data assimilation” or “model-data fusion,” is described in numerous textbooks [e.g., *Enting*, 2002; *Tarantola*, 2004]. The approach is commonly used in numerical weather prediction [e.g., *Daley*, 1991; *Kalnay*, 2003] where observations are combined with the underlying dynamical equations to determine the atmospheric state before beginning a forecast. The approach has also been used to determine parameters in models of the terrestrial biosphere [e.g., *Rayner et al.*, 2005; *Trudinger et al.*, 2007].

[7] Our motivation and approach are somewhat different than the preceding studies in this area. First we want an approach which is, as far as possible, algorithmic. We would like, ultimately, to generate the structure of emissions over the longest possible time series and the effort of obtaining the pointwise and sectoral information in other studies would be prohibitive. Secondly our starting point is the need for a globally homogeneous product. Many of the intensive data sources are only available locally such as the database of power plant emissions used by *Pétron et al.* [2008]. The approach we describe is sufficiently general to ingest such detailed information later but our first goal is a global baseline product. Thirdly, we need error estimates on the outputs of our system, because an important use of the generated fossil fluxes is as an input to atmospheric inversion calculations. The inversion methodology is inherently statistical so that all inputs must come with associated error statistics. None of the pointwise fossil emission products available today include such errors.

[8] The structure of this paper is as follows: In section 2 we describe the overall method and its components. In section 3 we show the impact of assimilating nightlights into a population-based downscaling, both on fluxes and on their uncertainties. Finally, in section 4 we point out some of the weaknesses of our approach and sketch some of the future enhancements to address them.

[9] This paper is supplemented by *Raupach et al.* [2010], which provides an informal introduction to the concept and a phenomenological investigation of the main data sets used here. In a subsequent paper we will produce a 12 year history of emissions (1992–2003) and investigate its space-time structure. Results can be seen online (<http://ffdas.org/>).

## 2. Methodology

### 2.1. Model-Data Fusion Architecture

[10] All model-data fusion applications involve six essential components, described in detail for the Fossil Fuel Data Assimilation System (FFDAS) in following subsections:

[11] 1. First is a predictive model. The model includes some parameters (things prescribed for the model) and some state variables (things calculated by the model). We will try to improve our knowledge of either state variables or parameters, by adjusting certain quantities in the model to obtain best agreement with some set of measurements. These adjustable quantities are the “control variables.” The choice of control variables is a fundamental one in any assimilation system.

[12] 2. Some prior information on the control variables is usually expressed as probability density functions (pdfs).

[13] 3. Also required is a set of observations, with “observation operators” which transform the predictions of the model into predictions of the observed quantities.

[14] 4. Fourth is a cost function which measures the disagreement between actual observations and their predicted values from the model via the observation operators and, potentially, the disagreement between control variables and their prior estimates. This cost function embodies our knowledge of the underlying probability density functions.

[15] 5. Fifth are uncertainty specifications for all observations, used to weight their contributions to the cost function.

[16] 6. Finally, required is a strategy to estimate various parameters of the posterior pdf of the control variables. This usually includes a search strategy to find the most likely estimate. It should also include a technique for estimating the posterior uncertainty.

[17] Formally, these components fit together as follows [*Enting*, 2002; *Tarantola*, 2004; *Raupach et al.*, 2005]. Let  $\tilde{\mathbf{z}}$  be a vector of observations and  $\mathbf{z} = \mathbf{h}(\mathbf{y})$  the corresponding predictions of these observed quantities from the predictive model. Here  $\mathbf{h}$  is the observation operator, a function of a vector of control variables  $\mathbf{y}$  which is a subset of the model state variables and parameters. Note that  $\mathbf{z}$  may include parts of the predictive model but may also include many other observables. The task is to describe the pdf for  $\mathbf{y}$  given those for  $\mathbf{y}_{\text{prior}}$ ,  $\tilde{\mathbf{z}} - \mathbf{z}$  and the operator  $\mathbf{h}$ . Under the Gaussian assumption, the most likely value of  $\mathbf{y}$  occurs at the minimum of the Bayesian cost function:

$$J(\mathbf{y}) = [\tilde{\mathbf{z}} - \mathbf{h}(\mathbf{y})]^T \mathbf{C}_z^{-1} [\tilde{\mathbf{z}} - \mathbf{h}(\mathbf{y})] + [\mathbf{y} - \mathbf{y}_{\text{prior}}]^T \mathbf{C}_p^{-1} [\mathbf{y} - \mathbf{y}_{\text{prior}}], \quad (1)$$

where  $\mathbf{y}_{\text{prior}}$  is the prior estimate of  $\mathbf{y}$ , and the covariance matrices  $\mathbf{C}_z$  and  $\mathbf{C}_p$  define the uncertainties in  $\tilde{\mathbf{z}}$  and  $\mathbf{y}_{\text{prior}}$ , respectively. The first term in equation (1) pulls the posterior estimate ( $\mathbf{y}_{\text{post}}$ ) toward consistency with the observations, while the second (Bayesian prior) term pulls  $\mathbf{y}_{\text{post}}$  toward the prior estimate  $\mathbf{y}_{\text{prior}}$ . The relative influences of the two terms are determined by our levels of confidence in the observations and the priors, which are quantified by the covariance matrices  $\mathbf{C}_z$  and  $\mathbf{C}_p$ , respectively. When the observation operator  $\mathbf{h}(\mathbf{y})$  is linear ( $\mathbf{h}(\mathbf{y}) = \mathbf{H}\mathbf{y}$ ), then the posterior solution for  $\mathbf{y}$  can be obtained analytically. When  $\mathbf{h}(\mathbf{y})$  is nonlinear, as in the present case, then a number of methods can be used to find the minimum solution; Here we use the quasi-Newton algorithm M1QN3 version 3.2 [*Gilbert and Lemaréchal*, 1989; *Liu and Nocedal*, 1989].

## 2.2. Predictive Model

### 2.2.1. Sectorally Aggregated Model

[18] The well-known Kaya identity [Nakicenovic, 2004] expresses the total, sectorally aggregated fossil fuel emission from a large region as

$$F = Pgef, \quad (2)$$

where  $F$  is the regional fossil fuel CO<sub>2</sub> emission flux (kgC yr<sup>-1</sup>),  $P$  is population (persons),  $g$  is per capita GDP (yr<sup>-1</sup> person<sup>-1</sup>) based on Purchasing Power Parity (PPP),  $e$  represents the energy intensity of the economy (MJ<sup>-1</sup>) and  $f$  is the fossil carbon intensity of energy (kgC MJ<sup>-1</sup>).

[19] The Kaya identity applies only for a region large enough that flows across the boundary of the region can be neglected relative to internal production and consumption. We are concerned with emissions from grid cells which may be very small, so the Kaya identity has to be extended. To do this we consider energy production and consumption separately. Let  $E_i^{\text{Prod}}$  be a vector of energy production across a set of grid cells  $i$ ; likewise, let  $E_i^{\text{Cons}}$  be a vector of energy consumption.  $E_i^{\text{Cons}}$  and  $E_i^{\text{Prod}}$  are related by

$$E_i^{\text{Prod}} = \sum_j S_{ij} E_j^{\text{Cons}}, \quad (3)$$

where  $S_{ij}$  is the fraction of energy consumed in cell  $j$  that is produced in cell  $i$ . Thus,  $S_{ij}$  is an energy transfer matrix from production to consumption cells. Row  $i$  of  $S_{ij}$  gives the fractions of the energy produced in cell  $i$  that are consumed in all cells; likewise, column  $j$  gives the fractions of energy consumed in cell  $j$  that are produced in all cells. In the absence of transmission losses and ignoring storage effects, conservation of energy requires that sums over rows and columns of  $S_{ij}$  be 1, but  $S_{ij}$  can also incorporate transmission losses.

[20] Local energy production (in cell  $i$ ) is related to emissions by

$$F_i = f_i E_i^{\text{Prod}}, \quad (4)$$

where  $f_i$  is the carbon intensity of energy in production cell  $i$ . Local energy consumption (in cell  $j$ ) is related to local population, income and energy intensity of GDP, as in the Kaya identity:

$$E_j^{\text{Cons}} = P_j g_j e_j. \quad (5)$$

Combining equations (2)–(5), we obtain

$$F_i = \sum_j f_i S_{ij} P_j g_j e_j. \quad (6)$$

Equation (6) is a local form of the Kaya identity, in which the transfer matrix  $S_{ij}$  is used to handle the fact that energy produced in one cell is consumed in another. If the cells are large enough to assume that energy is both produced and consumed in the same cell, then  $S_{ij}$  reduces to the identity matrix.

[21] It is useful to describe emissions at cell level by the emissions flux density vector  $\phi_i = F_i/A_i$  (with units kgC m<sup>-2</sup> yr<sup>-1</sup>), where  $A_i$  is the vector of cell areas. Likewise the population density (person m<sup>-2</sup>) is  $\rho_i = P_i/A_i$ . (Greek sym-

bols denote areal densities). In terms of these densities, the local Kaya identity, equation (6), becomes

$$\phi_i = \frac{f_i}{A_i} \sum_j S_{ij} A_j \rho_j g_j e_j. \quad (7)$$

### 2.2.2. Sectoral Disaggregation

[22] It is sometimes useful to split total CO<sub>2</sub> emissions from fossil fuels into contributions from several sectors which are potentially observable in different ways. The following sectoral breakdown is based on the six-sector primary classification of the International Energy Agency (IEA; <http://www.iea.org/>): (1) energy ( $E$ ), which includes emissions from the generation of commercial electricity and heat by autoproducers or public utilities (IEA sector 1), cogeneration in other manufacture (IEA sector 2), and ancillary emissions in other energy production (IEA sector 3); (2) manufacturing ( $M$ ), which includes all emissions associated with industrial manufacture (IEA sector 4), other than those accounted above; (3) transport ( $T$ ), which includes all land, sea and air transport, both domestic and international (IEA sector 5, together with emissions from international marine and aviation bunkers); it is also useful to break transport emissions into contributions from land ( $TL$ ), sea ( $TS$ ) and air ( $TA$ ); and (4) Other ( $X$ ), which includes emissions from other activities not accounted above, such as residential, agriculture and fishing. The total flux of emissions in a grid cell  $i$ ,  $F_i$ , is the sum of the emission fluxes from each sector:

$$F_i = F_i^{(E)} + F_i^{(M)} + F_i^{(TL)} + F_i^{(TS)} + F_i^{(TA)} + F_i^{(X)}, = \sum_s F_i^{(s)} \quad (8)$$

where  $s$  is a sectoral index.

## 2.3. Data

[23] The analysis in this paper is carried out at a grid resolution of 0.25°. We use observations from the following three primary data sources.

[24] 1. First are nationally aggregated data on emissions ( $F$ ) from the International Energy Agency (<http://www.iea.org/>). Countries are defined as follows: we aggregate the 206 countries of the International Energy Agency (IEA) data onto a 135 country subset consistent with both the country shape file we use and the 0.25° resolution. For some purposes we assign these 135 countries to the large regions used by Raupach [2007] and Raupach et al. [2010]: United States, Europe, Japan, D1 (other developed countries), FSU (Former Soviet Union countries), China, India, D2 (other developing countries), D3 (least developed countries as defined by the United Nations).

[25] 2. Second are gridded population data from <http://sedac.ciesin.columbia.edu/gpw/>. We aggregate this from the 2.5 arc minute resolution of the original to the 0.25° grid cells.

[26] 3. Third are satellite observations of nightlights from human settlements [Elvidge et al., 1997, 2001]. These data are obtained from the broadband visible-near-infrared (0.4–1.1 μm) channel of the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS),

**Table 1.** List of Control Variables for the Optimization With Their Dimensions, Prior Values, and Uncertainties<sup>a</sup>

Description	Symbol	Units	Resolution	Prior	Uncertainty
Per capita gross domestic product	$\hat{g}_j$	dollars yr <sup>-1</sup> person <sup>-1</sup>	cell	1.0	0.5
Energy intensity	$\hat{e}$	MJ dollar <sup>-1</sup>	global	1.0	10 <sup>-6</sup>
Carbon intensity	$\hat{f}^r$	kgC MJ <sup>-1</sup>	country	calculated	max(10 $\hat{f}^r$ , 10 <sup>-6</sup> )
Nightlight scaling	$\hat{k}$	counts MJ yr <sup>-1</sup> area	global	calculated	100 $\hat{k}$

<sup>a</sup>The “calculated” values are generated by back-propagating the observations with the observation operator, which is calculating prior values that are consistent with the observations.

usable at night and up to 4 orders of magnitude more sensitive than sensors such as AVHRR. Data are processed by the National Oceanic and Atmospheric Administration (NOAA) National Geophysical Data Center, from DMSP data collected by the US Air Force Weather Agency. Processing yields composite images (18 for the period 1992–2003) at about 1 km spatial resolution, aggregated from finer native resolution. Nightlights data has been shown to correlate well (when further aggregated) with national and regional GDP, and also to provide an initial means of downscaling fossil fuel emissions [Doll *et al.*, 2000]. In Europe, nightlights data correlate well at small (1 km) scales with population data [Briggs *et al.*, 2007]. A global analysis of the relationships between nightlights and population density is given by Raupach *et al.* [2010].

[27] Many other data sources can provide additional constraints on the distribution of emissions. Potential additional data sources, not used here but available for further exploration, include (1) atmospheric measurements of fossil fuel combustion products, such as NO<sub>x</sub>, δ<sup>14</sup>C and CO<sub>2</sub> itself; (2) additional socioeconomic data at national or local scales; and (3) data on the locations and emissions properties of major point sources such as power stations and heavy industry. All these additional data sources can be used either as global fields or for particular regions as available.

## 2.4. Control Variables

### 2.4.1. Spatial Patterns and Scaling Factors

[28] The traditional approach to downscaling emissions (see references in the Introduction) is to assume some spatial pattern of emissions within a country which must sum to an integral constraint such as a national emission. The control variables in the emissions model are the scaling factors for these patterns and are determined by the available observations, in this case, consisting only of integral constraints.

[29] Here we wish to generalize this approach by making use of both spatially aggregated and pointwise observations. To do this, each term in the Kaya identity can be expressed in one of two ways: (1) as a linear combination of disjoint spatial patterns (the “basis functions”), each multiplied by a scaling factor which is a control variable in the model-data fusion problem, or (2) as a set of independent values for each cell, each of which is a control variable. The choice depends on the available observations.

### 2.4.2. Choice of Control Variables

[30] To this point our description has been highly generic. We now make several simplifications for a first demonstration of the approach: (1) the predictive model is sectorally aggregated; (2) we ignore spatial separation between energy production and consumption, so that the transfer matrix  $S_{ij}$  in equations (6) and (7) is the identity matrix;

and (3) we use only one set of spatially aggregated observations (national emissions) and one set of pointwise observations (nightlights); the observation operators are described below.

[31] We also use the following notation: (1) subscripts denote quantities which vary spatially across grid cells; (2) superscripts denote quantities which are spatially aggregated across a country or region; (3) the assignment of grid cells to countries or regions is done with a matrix  $I_j^r$ , where  $I_j^r = 1$  if grid cell  $j$  is in country or region  $r$ , and 0 otherwise (thus, a total regional emission is  $F^r = \sum_j F_j I_j^r$ ); (4) control variables (section 2.1) are denoted by carets; and (5) observations are denoted by tildes. The above simplifications mean that our predictive model for emissions, equation (7), reduces to a conventional Kaya identity at cell level,  $\phi_j = \rho_j g_j e_j f_j$ . Terms in this expression are associated with control variables as follows.

[32] 1. The population density  $\rho_j$  is assumed to be fully specified, without error, from data. Therefore this is a set of parameters in the sense of section 2.1.

[33] 2. The per capita GDP in cell  $j$  is a set of pointwise control variables,  $\hat{g}_j$ .

[34] 3. The energy intensity of GDP in cell  $j$  is  $e_j = \hat{e}$ , a global constant.

[35] 4. The carbon intensity of energy in cell  $j$  is  $f_j = \hat{f}^r I_j^r$  which enforces a constant  $f_j$  within a country. Note that with observations available at the level of GDP and national emissions it is arbitrary whether  $e$  or  $f$  carries information at national resolution.

[36] 5. There is one control variable ( $\hat{k}$ ) in the observation operator for nightlights (Section 2.5.2) which describes a global proportionality between the density of nightlights and that of energy consumption.

### 2.4.3. Uncertainties and Prior Estimates for Control Variables

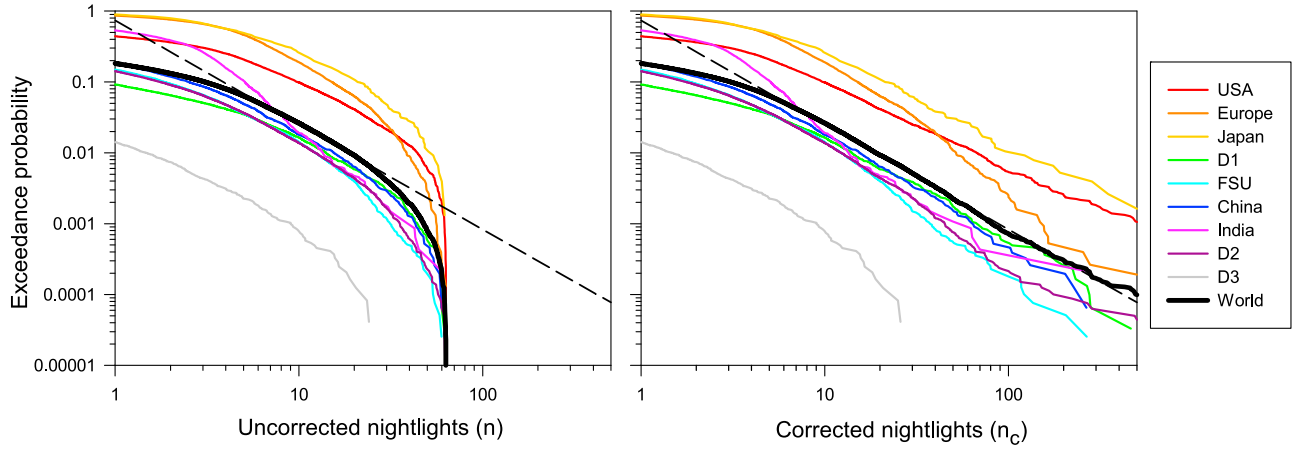
[37] Table 1 lists the control variables for the problem with their dimensions, prior values and uncertainties.

## 2.5. Observation Operators and Uncertainties

[38] In general, the observation operators are expressed as sets of scaling factors which multiply spatial fields. These can be matrix multiplications, often expressing a spatial integral, or pointwise multiplications mapping one spatial field, point by point, onto another. The operators also access different terms in the underlying Kaya identity. In this paper we use two examples of observables: national emissions and nightlights.

### 2.5.1. National Emissions

[39] The observed national emission for country  $r$  is  $\tilde{F}^r$  (the tilde denoting an observation). Modeled national



**Figure 1.** (left) Probability  $P(n)$  that the measured nightlight intensity in a  $0.25^\circ$  cell exceeds the value  $n$ . (right) Probability  $P(n_c)$  that the corrected nightlight intensity exceeds the value  $n_c$ . Probability distributions are aggregated over regions and over the world as a whole. The fit of equation (11) for the world is shown in both panels.

emissions are given by the following observation model (an operator on the control variables):

$$F^r = \sum_j A_j \rho_j \hat{g}_j \hat{e}_j^r I_j^r. \quad (9)$$

The uncertainty in the observation  $\tilde{F}^r$  is given by a variance  $\sigma^2(\tilde{F}^r)$ , which forms part of the diagonal of the observation covariance matrix in equation (1). For countries in the four most developed regions of *Raupach* [2007] (United States, Europe, Japan, D1) we attach a relative error of 5% to  $\tilde{F}^r$ , while for other countries we use 15%.

### 2.5.2. Nightlights

[40] The observed nightlight intensity in cell  $j$  is  $\tilde{n}_j$ , in units of sensor counts from the 8-bit detector (0–63). The observation model for nightlights starts from the assumption that nightlights are proportional to the areal density of energy consumption,  $E_j/A_j$ , with a single global proportionality constant  $\hat{k}$ . Support for this is provided by *Raupach et al.* [2010]. Note that  $\hat{k}$  is itself a control variable to be found in the assimilation process. This nightlights observation model is therefore

$$n_j = k(E_j/A_j) = \hat{k} \rho_j \hat{g}_j \hat{e}_j. \quad (10)$$

[41] Observations of nightlights from the DMSP-OLS sensor are well known to be subject to significant errors from sensor saturation. We apply a correction to account for this saturation error following *Raupach et al.* [2010]. We hypothesize that the exceedance probability distribution (EPD) of nightlights intensity ( $n$ , in sensor counts from 0 to 63) follows a power law:

$$P(n) = \text{Prob}(n > n) = (n/n_0)^{-\alpha}, \quad (11)$$

where  $P(n)$  is the probability (across the ensemble of cells in a given region) that the nightlights intensity exceeds  $n$ ,  $\alpha$  is a power law exponent, and  $n_0$  is a scaling parameter.

[42] Figure 1 (left) (based on the work of *Raupach et al.* [2010]) shows  $P(n)$  for measured nightlights, aggregated

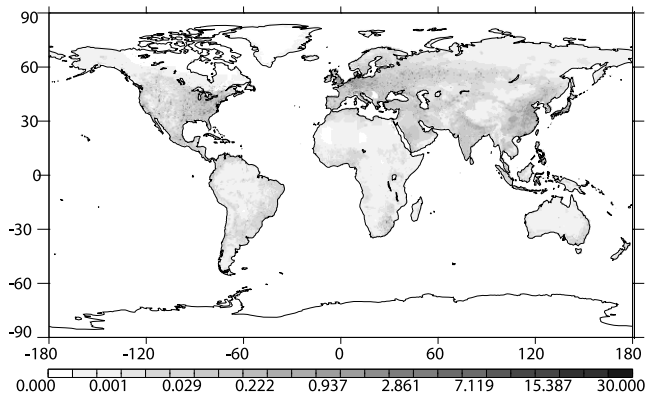
over regions and over the world as a whole, together with fits of equation (11) for each curve. The power law is followed over a range of  $n$  of nearly a decade, with the observed EPD departing from the power law as  $n$  approaches the saturation value of 63 counts. The exponent  $\alpha$  depends on the choice of region:  $\alpha$  is about 2 for the world as a whole, and is lower (higher) than this value for more developed (less developed) regions, respectively. Also,  $\alpha$  depends on the cell size over which observations of  $n$  are aggregated, which for this work is always  $0.25^\circ$ .

[43] The hypothesis that the distribution of nightlights obeys equation (11) is based not only on the observation of power law behavior of  $P(n)$  over a range of  $n$  below saturation, but also on observed power law distributions for other metrics of the spatial arrangement of human settlements. In particular, the EPD  $P(\rho)$  for the population density ( $\rho$ ) is observed to follow a power law over a wide range of  $\rho$  [*Raupach et al.*, 2010]. In this case there is no saturation error in measurements at high values of  $\rho$ , and there is no sign of systematic departures from the power law at high  $\rho$ . This motivates our hypothesis that the true distribution  $P(n)$  for nightlights follows equation (11) out to values of  $n$  which encompass the brightest  $0.25^\circ$  cells on the planet, and that sensor saturation error accounts for the observed roll-off of  $P(n)$  as  $n$  approaches the saturation value of 63.

[44] The mathematical form of the correction is based on the following general fact. We are given an ensemble of measured values  $n_m$  to be corrected to “true” values  $n_c$  with a mapping  $n_c = f(n_m)$ . We restrict  $f$  to be monotonic, which ensures that  $f$  preserves the rank order of the ensemble. It follows that  $P_m(n_m) = P_c(n_c)$ , where  $P_m$  and  $P_c$  are the EPDs of  $n_m$  and  $n_c$ , respectively. The correction mapping is thus  $f(\cdot) = P_c^{-1}(P_m(\cdot))$ , where  $P_c^{-1}$  is the functional inverse of  $P_c$ . Since  $P_m$  is determined by the measurements, an assumed form for  $P_c$  determines  $f$ .

[45] In the present case  $P_c$  is given by equation (11) and the correction mapping is

$$n_c = n_m (r_{\text{EPD}}(n_m))^{-1/\alpha}, \quad (12)$$



**Figure 2.** Fossil fuel emissions for 2002 ( $\text{kgC m}^{-2} \text{yr}^{-1}$ ) at  $0.25^\circ$  resolution. The calculation uses the spatial distribution of population and national emissions data. Values have been transformed according to  $F^{0.2}$  to improve the resolution for low values.

where  $r_{\text{EPD}}(n_m)$  is the ratio of the measured EPD to the true (power law) EPD:

$$r_{\text{EPD}}(n_m) = \frac{P_m(n_m)}{(n_m/n_0)^{-1/\alpha}}. \quad (13)$$

This is the ratio between the solid (measured) to the dashed (power law) lines in Figure 1.

[46] We constructed an empirical fit to  $r_{\text{EPD}}(n_m)$  for the world as a whole, and thence derived a nightlights correction mapping using equation (12). When  $\log_{10}(n_m) < 1.3$  or  $n_m < 19.9526$ , there is no correction ( $n_c = n_m$ ). When  $\log_{10}(n_m) > 1.3$ , the correction mapping is given by

$$\log_{10}(n_c) = \log_{10}(n_m) + 0.0684035 \cdot \left( -\frac{1}{\log_{10}(n_m/66)} - 1.92477 \right)^{0.74}. \quad (14)$$

The correction increases rapidly as  $n_m$  approaches saturation. For  $n_m = (40, 50, 60, 63)$ , equation (14) yields  $n_c = (55.4, 92.9, 286.5, 980.5)$ .

[47] Figure 1 (right) shows the resulting EPD of corrected nightlights ( $n_c$ ) for the world and for each region. Most of the saturation-induced roll-off in EPD is removed. For the whole world (black line), the corrected EPD follows the power law very well except for the few brightest cells. The corrected EPDs for individual regions show systematic departures from the regional power law fits, which arise because power law properties and  $r_{\text{EPD}}(n_m)$  are different for the various regions. However, we have chosen to apply a globally uniform nightlights correction and to regard these departures as contributing to error in  $n_c$ .

[48] Uncertainty in the corrected nightlights arises from two sources: measurement error in the observations ( $n_m$ ) and error in the correction algorithm yielding  $n_c$ . The former is likely to be constant, and the latter related to the magnitude of the difference  $n_c - n_m$ . We therefore express the uncertainty in  $n_c$  as

$$\sigma(n_c) = 1.5 + (n_c - n_m). \quad (15)$$

## 2.6. Calculation of Posterior Uncertainties

[49] An important product of any assimilation system is a posterior uncertainty estimate either for control variables or for diagnosed quantities. We generate these using the Monte Carlo technique described by *Chevallier et al.* [2007]. The approach can be summarized as follows: (1) choose a control vector  $\mathbf{y}_t$  which we assume true. This is usually the value of a previous optimization; (2) simulate all observables using  $\mathbf{y}_t$  to produce  $\mathbf{z}_t$ ; (3) perturb  $\mathbf{z}_t$  with noise consistent with the error statistics described in the observation error covariance matrix  $\mathbf{C}_z$  to produce  $\mathbf{z}$ ; (4) perturb  $\mathbf{y}_t$  with noise consistent with the prior covariance matrix  $\mathbf{C}_p$  to produce a prior estimate  $\mathbf{y}_{\text{prior}}$ ; (5) carry out an inversion using  $\mathbf{y}_{\text{prior}}$ ,  $\mathbf{z}$  and the relevant covariance matrices to produce an updated estimate  $\mathbf{y}$ ; and (6) for the linear Gaussian case, the statistics of  $\mathbf{y} - \mathbf{y}_t$  are consistent with the usual posterior covariance matrix except that we only have a limited number of realizations for calculating the statistics. We can repeat steps 3–5 as often as we can afford to improve these statistics. Here we use 25 realizations. Each grid point is also a realization so the ensemble statistics of many points are more reliable than those for an individual point.

## 3. Results

[50] In this preliminary analysis we consider the separate and joint use of nightlights and population data to refine country-level information on emissions. We also explore the relationships between nightlight intensity, population density and aggregated emissions. As well as revealing the separate information content in the nightlights and population data, this also allows us to test the various downscaling approaches open to us, at least in an aggregated sense. Finally we assess the value of various data sources, particularly measurements of  $^{14}\text{CO}_2$ , in constraining emissions.

### 3.1. Flux Patterns

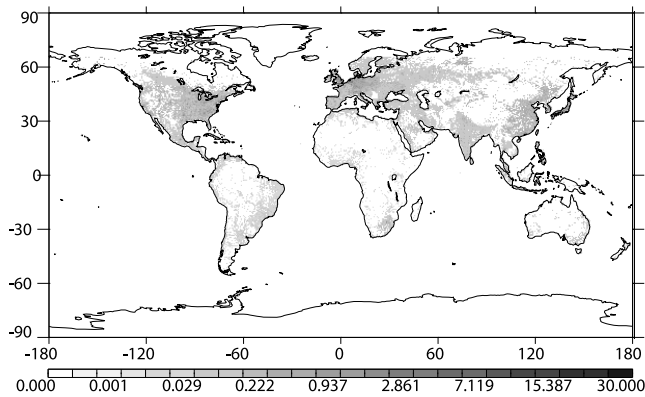
[51] We carry out three different assimilation experiments using different assumed spatial patterns and different data sources. In the first we use the spatial distribution of population and only use the country-level emission data. Thus the disaggregation of national emissions is driven entirely by population, as in the distribution used by *Tans et al.* [1990]. This distribution of emissions is shown in Figure 2.

[52] The second experiment is identical with the first except that we use the uncorrected distribution of nightlights for 2002 to carry out the disaggregation. The distribution of emissions is shown in Figure 3.

[53] Finally, Figure 4 shows the results of an assimilation using both population and corrected nightlights. The  $\chi^2$  value for the assimilation is 246213. We use a total of 245642 observations. Ideally the  $\chi^2$  value is equal to the number of observations [*Tarantola*, 1987, p. 211] and our case is very close to this suggesting that the statistical assumptions underlying the assimilation are valid.

[54] We see that the population based estimate (Figure 2) is the most variable of the three distributions, showing the strongest peaks. Figure 3 is the smoothest and Figure 4 intermediate. We see substantial differences between the nightlights and population-based downscaling over many regions and without a predictable sign. Over Nepal the



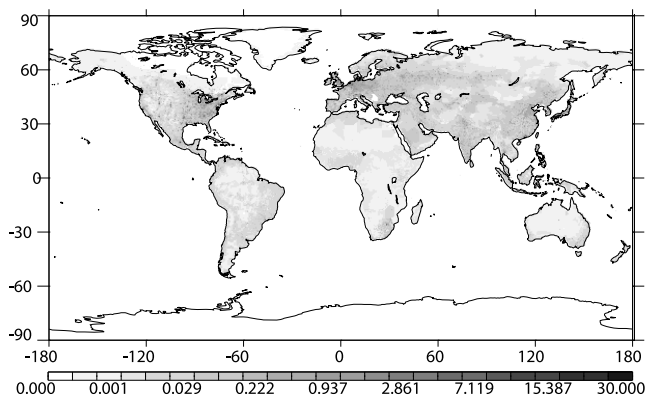


**Figure 3.** Fossil fuel emissions for 2002 ( $\text{kgC m}^{-2} \text{yr}^{-1}$ ) at  $0.25^\circ$  resolution. The calculation uses the spatial distribution of nightlights and national emissions data. Values have been transformed according to  $F^{0.2}$  to improve the resolution for low values.

population-based downscaling concentrates emissions in the Kathmandu valley while nightlights gives a more even distribution. In nearby regions of southwest China, the nightlights-based downscaling produces the more concentrated emissions. In both cases the FFDAS distribution tends to lie between them, but this is not true everywhere. The maximum emissions for the three distributions are different, New York for the population-based estimate, a pixel in Chinese Taipei for the nightlights and Tokyo for the FFDAS. Also, despite the general smoothing of the FFDAS compared to population downscaling, the maximum for FFDAS is larger than for the other two distributions, a result of the dual constraint of nightlights and national emissions.

### 3.2. Comparison With Vulcan Inventory

[55] Although building an assimilation system to combine available information on fossil fuel emissions is an interesting methodological problem, it is of little use to the carbon cycle and emissions communities unless it can produce more reliable products than existing systems. To test



**Figure 4.** Fossil fuel emissions for 2002 ( $\text{kgC m}^{-2} \text{yr}^{-1}$ ) at  $0.25^\circ$  resolution. The calculation uses the spatial distribution of population, nightlights, and national emissions data in an assimilation. Values have been transformed according to  $F^{0.2}$  to improve the resolution for low values.

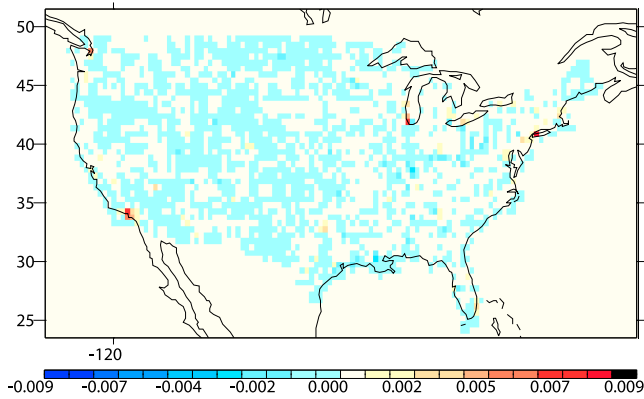
**Table 2.** Comparison of Population-Based, Nightlight-Based, and Assimilation-Based Fossil Inventories for 2002 Over the United States With the Vulcan Inventory of Gurney *et al.* [2009]<sup>a</sup>

Resolution (deg)	Population		Nightlights		FFDAS		Brenkert [1998]	
	Diff.	Correl.	Diff.	Correl.	Diff.	Correl.	Diff.	Correl.
0.5	1213	0.70	1360	0.68	1143	0.74	-	-
1.0	1006	0.80	1087	0.81	900	0.85	1045	0.75
2.0	806	0.84	810	0.88	651	0.91	788	0.84
3.0	670	0.87	691	0.90	545	0.92	654	0.87
4.0	608	0.88	641	0.92	479	0.93	644	0.87

<sup>a</sup>The fourth inventory (also population based) is taken from Brenkert [1998]. All fields have been rescaled to match the Vulcan total. The “diff” columns (in units of  $\text{MtC yr}^{-1}$ ) show the sum of the absolute differences over the Vulcan domain ( $23.5^\circ\text{N}$ – $51.5^\circ\text{N}$  and  $127.5^\circ\text{W}$ – $62.5^\circ\text{W}$ ). Because the Vulcan domain does not include points outside the United States, we have zeroed pixels in the fields that are zero for Vulcan. The Brenkert [1998] inventory is constructed at  $1^\circ \times 1^\circ$ , so we do not compare it at  $0.5^\circ \times 0.5^\circ$ . FFDAS, Fossil Fuel Data Assimilation System.

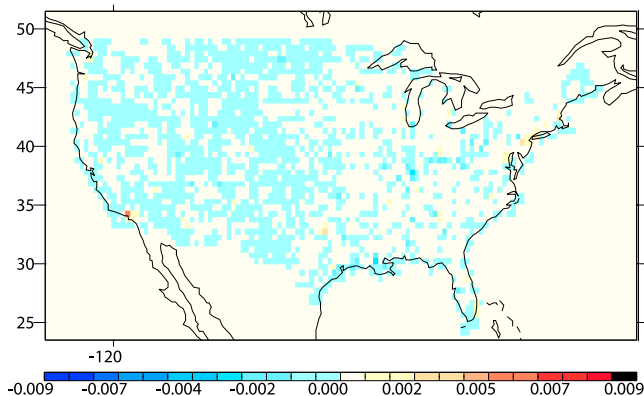
this we need an agreed standard likely to be better than the approaches we are testing. This exists locally for the United States in the form of the Vulcan inventory produced by Gurney *et al.* [2009]. We need to test therefore whether our assimilated product is closer to the Vulcan inventory than a pure population or pure nightlights downscaling. We choose two metrics which reflect likely uses for the inventory. The summed absolute difference (SAD) is the sum of the absolute difference of the field over the domain. This is important if one is going to use the resulting fields directly, e.g., as input for a forward model or for inversions at the pixel level. If one wishes to use the emissions field for an inversion where one solves for large-scale magnitudes such as national emissions, the important point is the agreement in the pattern within the country. Thus we use the spatial correlation which is independent of magnitude. Table 2 shows these measures at resolutions from  $0.5^\circ \times 0.5^\circ$  (the lowest common multiple of the native resolutions of FFDAS and Vulcan) to  $4^\circ \times 4^\circ$ . We show the comparison for the fields distributed by population alone, by uncorrected nightlights alone, the FFDAS assimilated product and finally the field from Brenkert [1998]. This is also a population-based downscaling for the year 1995. We rescale it to 2002 for fairer comparison. The native resolution of the Brenkert [1998] inventory is  $1^\circ \times 1^\circ$  so comparison at finer resolution is meaningless.

[56] We see first that all four products perform fairly well relative to Vulcan. This is surprising given that population and nightlights are only a rough guide to the placement, for example, of fossil fuel burning power stations. All metrics improve as we move to coarser resolution as one would expect. We see also that the FFDAS product is superior to the other three at all resolutions. The improvement in spatial correlation is especially important if one wishes to use the patterns of the fossil emissions directly in an atmospheric inversion. Comparison of (uncorrected) nightlights and population-based estimates is mixed, with the population-based estimates producing lower absolute differences but also lower correlations. A similar analysis with the corrected nightlights produced performance intermediate between the population-based and uncorrected nightlight-based estimates in Table 2.

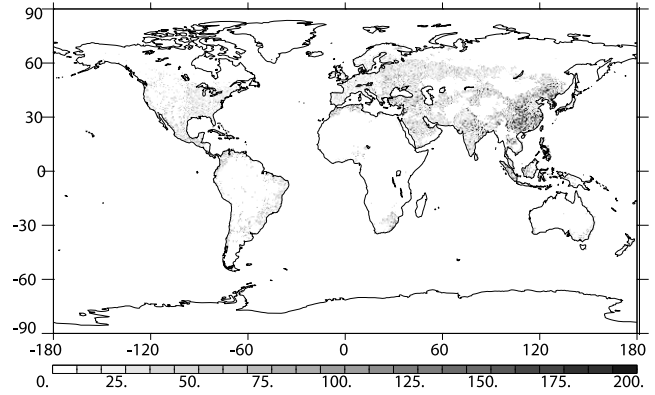


**Figure 5.** Differences ( $\text{kgC m}^{-2} \text{yr}^{-1}$ ) between a population-based downscaling and the Vulcan inventory of *Gurney et al.* [2009] at  $0.5^\circ$  resolution. Values in which the Vulcan inventory is zero have been removed from both fields to limit the differences to the Vulcan domain. The downscaled emissions have been rescaled to the Vulcan total.

[57] To understand the causes of the improvement, Figure 5 and Figure 6 show the differences between the population and FFDAS-based inventories relative to Vulcan. The improvement is most noticeable over high population densities such as the eastern seaboard, the Los Angeles area and around Chicago. Here it appears that the population-based estimate overestimates emissions while assimilation of the nightlights data trims these peaks. This is true despite the large saturation correction in some of these regions. Note that Figure 2 and Figure 4 show differences over most points of extreme population density suggesting that this improvement may be globally important. Another difference is in the regions of significant power generation in the Ohio and Tennessee valleys. Here the low population density forces low emissions in the population-based estimate while the assimilation of



**Figure 6.** Differences ( $\text{kgC m}^{-2} \text{yr}^{-1}$ ) between Fossil Fuel Data Assimilation System (FFDAS) emissions and the Vulcan inventory of *Gurney et al.* [2009] at  $0.5^\circ$  resolution. Values in which the Vulcan inventory is zero have been removed from both fields to limit the differences to the Vulcan domain. The FFDAS emissions have been rescaled to the Vulcan total.



**Figure 7.** Fractional uncertainty (standard deviation divided by percentage emissions) for 2002 at  $0.25^\circ$  resolution. The calculation uses the spatial distribution of population, nightlights, and national emissions data in an assimilation. Values have been clipped at 200% to avoid statistical outliers.

nightlights partially corrects the emissions. Significant errors remain which could be addressed by inclusion of data on power station emissions. We note the recent work of *Oda and Maksyutov* [2010] who included a database of point sources in their estimates and improved the comparison with the work of *Gurney et al.* [2009].

### 3.3. Posterior Uncertainties

[58] Figure 7 shows the fractional uncertainty of the estimated fluxes. Note that with the relatively small number of realizations in the Monte Carlo calculation and the large number of pixels, we expect some outliers hence not all details in this map are significant. We see a reflection of the higher emission uncertainties in some countries than others, e.g., comparing the United States and China. We also see that the relative uncertainties on the pixel-level fluxes are quite high, on average around 50%. This arises from the combined uncertainty of the two data sources we employ. With national emissions uncertain at either the 5% or 15% level we cannot expect pixel-level fluxes to have a smaller uncertainty than this. Second, the uncertainty in the nightlights observation is often large. At low values the fixed uncertainty of 1.5 is often greater than 100% while at high values the uncertainty in the saturation correction approaches 100%. Note that using the uncertainty of this pointwise map alone in an inversion is a serious error since it assumes independence of errors, contradicting the use of national emissions data in the assimilation. It would, for example, yield very small errors on the national emissions of large countries. The spread of national emission estimates calculated from our realizations is consistent with the uncertainty in national emissions we use.

[59] Given the warning about the pointwise uncertainty, a question arises of how to transmit then include these uncertainties in subsequent inversions. The usual method, the posterior covariance, is impractical since our problem has about 245,000 unknowns. Furthermore the limited number of Monte Carlo realizations we have carried out cannot hope to capture the detail of such a matrix. We can, however, take advantage of equation (2) and the particular form of national



and pointwise disaggregation of our unknowns. We can write the flux at grid cell  $i$  as

$$F_i = (\mathbf{I}_i^T A) \times B_i, \quad (16)$$

where  $F$  is the flux,  $A$  is a vector of national emissions with dimension the number of countries for which we solve and  $B$  is a vector of pointwise multipliers disaggregating these national emissions.  $\mathbf{I}_i^T$  describes the assignment of grid points to countries and was introduced in Section 2.4.2.

[60] With the information in use in this version of FFDAS, we can approximate the covariances of  $A$  and  $B$  as independent, both of each other and internally. Note that there may well be uncertainty correlation between the average of  $B$  for a country and  $A$  for that country but the uncertainty variance of the country average of  $B$  will be so small for almost all countries that we can ignore the covariance. Assuming independence of  $A$  and  $B$  we can write the uncertainty covariance of  $F$  as

$$C(F) = \text{DIAG}(\mathbf{I}A)C(B) + \text{DIAG}(B)^2\text{IC}(A)\mathbf{I}^T, \quad (17)$$

where the superscript  $T$  represents the transpose and the operation **DIAG** transforms a vector into a diagonal matrix. Note that the second term on the right of equation (17) introduces off-diagonal terms into  $C(F)$  accounting for the propagation of uncertainty in national emissions to all points within the country. We can estimate the covariances on the right of equation (17) from the Monte Carlo realizations already discussed. All the terms on the right of equation (17) are available on the FFDAS Web site so researchers can construct a version of  $C(F)$  relevant to the resolution of their inversion. Note that the technique used here relies on the fact that only two scales of information are in use and that the spatial projections of the information do not overlap. This will break down in a future version when we begin to use other constraints. In that case only estimates of the posterior covariance taken directly from the Monte Carlo realizations will be available. Probably we will need to increase the number of such realizations to estimate off-diagonal terms in  $C(F)$ .

### 3.4. Use of $^{14}\text{CO}_2$ Measurements

[61] The importance of fossil fuel emissions as a driver of anthropogenic climate change has motivated efforts to measure them as directly as possible. This is especially true in the light of treaty obligations to limit or reduce emissions. One potential tracer of fossil fuel emissions is  $^{14}\text{CO}_2$ . *Turnbull et al.* [2009] showed that, in the contemporary atmosphere, this tracer is only weakly contaminated by other signals, at least in the northern hemisphere. FFDAS allows us to test the added constraint afforded by  $^{14}\text{CO}_2$  measurements relative to and in the presence of other measurements.

[62] Using the same Monte Carlo approach, we simulated annually averaged measurements of  $^{14}\text{CO}_2$  at 194 stations. These are a superset of the stations used by *Piao et al.* [2009]. We use the retroplumes from the LMDZ transport model described by *Peylin et al.* [2005] to calculate the sensitivity of each of these measurements to annually averaged fluxes from each of the  $3.75^\circ$  longitude by  $2.5^\circ$  latitude grid boxes of the LMDZ model. We assign each measurement an uncertainty of 1 ppm following *Turnbull et*

*al.* [2009]. The additional reduction of uncertainty on pixel level fluxes (relative to the standard assimilation) is quite weak, about 15% when globally integrated. This is not surprising since many pixels will fall in the footprint of each measurement and this annually averaged sampling approach will not localize an emission very well. The effect on the uncertainty of national emissions is much stronger. Using the square root of the trace of the uncertainty covariance as a global measure we see a reduction of 70%. Much of this is achieved by added constraint on emissions from the United States and China. Indeed the  $^{14}\text{CO}_2$  measurements are most effective in large countries. Here the measurement footprint sees only one unknown national emission. Small countries suffer the same problem as individual pixels. Of course the use of 194 stations is unrealistic but this is partly counterbalanced by the availability of more frequent measurements in reality.

## 4. Discussion

[63] When assessing the usefulness of our results it is important to recall the limited scope of the work. The weaknesses in the spatial structure are shown by the comparison with the inventory of *Gurney et al.* [2009]. If the detailed inputs used by *Gurney et al.* [2009] were available globally and historically our study would become irrelevant. However the global extension is difficult and a retrospective analysis almost impossible. The Vulcan inventory of *Gurney et al.* [2009] also pertains to 2002, and while it is certain that subsequent versions will be more timely, the gathering of the necessary detailed information will always impose a substantial delay. The DMSP/OLS series of instruments is still flying however and raises the tantalizing possibility of a more rapid update to spatially explicit emissions.

[64] Underlying any plans for continuing this work is the question whether the approach is superior to downscaling by population alone. This appears to be the case for the one region we tested. Some of the improvements have implications for coarse-resolution ( $>2^\circ$ ) inversions. Differences between our approach and population downscaling remain physically significant even at this resolution. There are differences between our estimates and many previous inventories. We do not include emissions from the cement industry in our estimates since these are not included in the IEA estimates we use. This is neither a strength nor a weakness but should be kept in mind when making comparisons.

[65] Sources appear to be more dispersed with our approach than with the population downscaling but it appears we still often overestimate peak emissions in the Vulcan inventory. This will require a reassessment of which sites are regarded as polluted by large fossil signals in the transport model and opens the way for using more of the concentration data in inversions. It does not, however, yet allow the much more important change to using high-frequency data. These data are sure to reflect details of the spatial variability of the fossil sources that we still do not capture but, more importantly, temporal structures in these sources we have not even attempted to estimate. We are unlikely to find proxies for the temporal variation in overall

source so we will be forced to impose temporal structures derived from intensive observations at a few points.

[66] The method for using  $^{14}\text{CO}_2$  measurements as a direct constraint appears promising and realistic. It is clear that an individual measurement will provide less constraint in our system (where fluxes must be estimated at each point) than a system where emission factors multiply fixed spatial patterns. The differences in these spatial patterns demonstrated throughout this paper suggest these large-scale approaches are risky indeed. We have not, however, exhausted the capability of direct measurements of combustion in the atmosphere. *Levin and Karstens* [2007] have demonstrated the potential of  $^{14}\text{CO}_2$  as a calibration for the far less expensive CO measurements. For some seasons and latitudes this may provide a rich source of data which FFDAS could ingest without much difficulty.

[67] Probably the greatest advance in this work, will come when we can separate the downscaling approaches by sector. The IEA, as already mentioned, divide emissions into six sectors. Spatially explicit proxies for all of these do not exist. We do have (and can use) direct estimates for some of them for some regions such as the emissions from U.S. power stations of *Pétron et al.* [2008]. for others there are indirect proxies such as satellite measurements of oxides of nitrogen.

[68] There are emissions not treated within FFDAS either because they do not appear within the IEA national statistics or because they do not fall within the country grids we use. Examples include gas flares and emissions from international transport. Gas flares are visible as nightlights and so could be treated as a separate category. International transport remains problematic and awaits improvement in inventory methodology.

[69] Finally we make some comments on how to use FFDAS estimates in other inversions. FFDAS estimates a pdf for the spatial distribution of fossil fuel emissions. This can form part of the prior pdf for a  $\text{CO}_2$  flux inversion. It must be combined with pdfs for other components following, for example, the calculation of *Chevallier et al.* [2006]. The different spatial structures of the uncertainty correlations of the biospheric and fossil flux estimates must be accounted for, as must the lack of information on temporal variability in FFDAS. The alternative is to include the FFDAS methodology directly within the  $\text{CO}_2$  inversion. We provide forward, tangent linear and adjoint versions of FFDAS to facilitate this.

## 5. Conclusions

[70] This paper describes and demonstrates an assimilation system (FFDAS) for estimating the spatial structure of carbon fluxes arising from fossil fuel combustion. The system currently assimilates data on national emissions and fixed nightlights. The conclusions can be summarized as follows.

[71] 1. FFDAS produces estimates which are spatially smoother than previous estimates downscaled by population. This occurs despite the inflation of the raw nightlights measurements to account for instrumental saturation.

[72] 2. The spatial structure from FFDAS over the United States agrees more closely with the detailed bottom-up

inventory of *Gurney et al.* [2009] than downscaling either by population or using the raw nightlights values.

[73] 3. Relative uncertainties in the emissions are generally around 50% at 0.25° resolution. They show spatial correlations within countries.

[74] 4. The use of  $^{14}\text{CO}_2$  measurements can provide a considerable regional constraint but should target strong emission regions if they are to constrain national totals.

[75] **Acknowledgments.** P.J.R. is the recipient of an Australian Research Council Professorial Fellowship (DP1096309).

## References

- Baker, D. F., et al. (2006), TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional  $\text{CO}_2$  fluxes, 1988–2003, *Global Biogeochem. Cycles*, 20, GB1002, doi:10.1029/2004GB002439.
- Brenkert, A. L. (1998), Carbon Dioxide Emission Estimates From Fossil-Fuel Burning, Hydraulic Cement Production, and Gas Flaring for 1995 on a One Degree Grid Cell Basis, <http://cdiac.esd.ornl.gov/ndps/ndp058a.html>, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Briggs, D. J., J. Gulliver, D. Fecht, and D. M. Vienneau (2007), Dasymeric modelling of small-area population distribution using land cover and light emissions data, *Remote Sens. Environ.*, 108, 451–466.
- Chevallier, F., N. Viovy, M. Reichstein, and P. Ciais (2006), On the assignment of prior errors in Bayesian inversions of  $\text{CO}_2$  surface fluxes, *Geophys. Res. Lett.*, 33, L13802, doi:10.1029/2006GL026496.
- Chevallier, F., F.-M. Bron, and P. J. Rayner (2007), The contribution of the Orbiting Carbon Observatory to the estimation of  $\text{CO}_2$  sources and sinks: Theoretical study in a variational data assimilation framework, *J. Geophys. Res.*, 112, D09307, doi:10.1029/2006JD007375.
- Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge Univ. Press, Cambridge, U. K.
- Doll, C. N. H., J.-P. Muller, and C. D. Elvidge (2000), Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions, *Ambio*, 29, 157–162.
- Elvidge, C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, and E. R. Davis (1997), Mapping city lights with nighttime data from the DSMP operational linescan system, *Photogramm. Eng. Remote Sens.*, 63, 727–734.
- Elvidge, C. D., M. L. Imhoff, K. E. Baugh, V. R. Hobson, I. Nelson, J. Safran, J. B. Dietz, and B. T. Tuttle (2001), Night-time lights of the world: 1994–1995, *J. Photogramm. Remote Sens.*, 56, 81–99.
- Enting, I. G. (2002), *Inverse Problems in Atmospheric Constituent Transport*, 392 pp., Cambridge Univ. Press, New York.
- Gilbert, J. C., and C. Lemaréchal (1989), Some numerical experiments with variable storage quasi-Newton algorithms, *Math. Programm.*, 45, 407–435.
- Gurney, K. R., et al. (2002), Towards robust regional estimates of  $\text{CO}_2$  sources and sinks using atmospheric transport models, *Nature*, 415, 626–630.
- Gurney, K. R., D. L. Mendoza, Y. Zhou, M. L. Fischer, C. C. Miller, S. Geethakumar, and S. de la Rue du Can (2009), High resolution fossil fuel combustion  $\text{CO}_2$  emission fluxes for the United States, *Environ. Sci. Technol.*, 43, 5535–5541, doi:10.1021/es900806c.
- Kalnay, E. (2003), *Atmospheric Modelling, Data Assimilation, and Predictability*, Cambridge Univ. Press, New York.
- Levin, I., and U. Karstens (2007), Inferring high-resolution fossil fuel  $\text{CO}_2$  records at continental sites from combined  $^{14}\text{CO}_2$  and CO observations, *Tellus, Ser. B*, 59, 245–250, doi:10.1111/j.1600-0889.2006.00244.x.
- Liu, D. C., and J. Nocedal (1989), On the limited memory BFGS method for large scale optimization, *Math. Programm.*, 45, 503–528.
- Marland, G., T. A. Boden, R. C. Griffin, S. F. Huang, P. Canciruk, and T. R. Nelson (1989), Estimates of  $\text{CO}_2$  emissions from fossil fuel burning and cement manufacturing, based on the U.S. Bureau of Mines cement manufacturing data, *Rep. ORNL/CDIAC-25, NDP-030*, Carbon Dioxide Inf. Anal. Cent. Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Marland, G., T. A. Boden, and R. J. Andres (2006), Global, regional, and national  $\text{CO}_2$  emissions, in *Trends: A Compendium of Data on Global Change*, Carbon Dioxide Inf. Anal. Cent. Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Nakicenovic, N. (2004), Socioeconomic driving forces of emissions scenarios, in *The Global Carbon Cycle: Integrating Humans, Climate, and the Natural World*, edited by C. B. Field and M. R. Raupach, pp. 225–239, Island Press, Washington, D. C.

- Oda, T., and S. Maksyutov (2010), A very high-resolution global fossil fuel CO<sub>2</sub> emission inventory derived using a point source database and satellite observations of nighttime lights, 1980–2007, *Atmos. Chem. Phys. Discuss.*, *10*, 16,307–16,344.
- Olivier, J. G. J., J. A. V. Aardenne, F. J. Dentener, V. Pagliari, L. N. Ganzeveld, and J. A. H. W. Peters (2005), Recent trends in global greenhouse gas emissions: Regional trends 1970–2000 and spatial distribution of key sources in 2000, *J. Integrative Environ. Sci.*, *2*, 81–99, doi:10.1080/15693430500400345.
- Pétron, G., P. Tans, G. Frost, D. Chao, and M. Trainer (2008), High resolution emissions of CO<sub>2</sub> from power generation in the USA, *J. Geophys. Res.*, *113*, G04008, doi:10.1029/2007JG000602.
- Peylin, P., et al. (2005), Daily CO<sub>2</sub> flux estimate over Europe from continuous atmospheric measurements: Part 1, Inverse methodology, *Atmos. Chem. Phys.*, *5*, 3173–3186.
- Piao, S., J. Fang, P. Ciais, P. Peylin, Y. Huang, S. Sitch, and T. Wang (2009), The carbon balance of terrestrial ecosystems in China, *Nature*, *45*, 1009–1013.
- Raupach, M. R. (2007), Dynamics of resource production and utilisation in two- 29 component biosphere-human and terrestrial carbon systems, *Hydro. Earth Syst. Sci.*, *11*, 875–889.
- Raupach, M. R., P. J. Rayner, D. J. Barrett, R. DeFries, M. Heimann, D. S. Ojima, S. Quegan, and C. Schimmlus (2005), Model-data synthesis in terrestrial carbon observation: Methods, data requirements and data uncertainty specifications, *Global Change Biol.*, *11*, 378–397.
- Raupach, M. R., P. J. Rayner, and M. Paget (2010), Regional variations in spatial structure of nightlights, population density and fossil-fuel CO<sub>2</sub> emissions, *Energy Policy*, *38*, 4756–4764, doi:10.1016/j.enpol.2009.08.021.
- Rayner, P. J., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann (2005), Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), *Global Biogeochem. Cycles*, *19*, GB2026, doi:10.1029/2004GB002254.
- Rayner, P. J., R. M. Law, C. E. Allison, R. J. Francey, and C. Pickett-Heaps (2008), The interannual variability of the global carbon cycle (1992–2005) inferred by inversion of atmospheric CO<sub>2</sub> and δ<sup>13</sup>CO<sub>2</sub> measurements, *Global Biogeochem. Cycles*, *22*, GB3008, doi:10.1029/2007GB003068.
- Stephens, B. B., et al. (2007), Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO<sub>2</sub>, *Science*, *316*, 1732–1735, doi:10.1126/science.1137004.
- Tans, P. P., I. Y. Fung, and T. Takahashi (1990), Observational constraints on the global atmospheric CO<sub>2</sub> budget, *Science*, *247*, 1431–1438.
- Tarantola, A. (1987), *Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation*, Elsevier, Amsterdam.
- Tarantola, A. (2004), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, Pa.
- Trudinger, C. M., et al. (2007), The OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models, *J. Geophys. Res.*, *112*, G02027, doi:10.1029/2006JG000367.
- Tumbull, J., P. Rayner, J. Miller, T. Naegler, P. Ciais, and A. Cozic (2009), On the use of <sup>14</sup>CO<sub>2</sub> as a tracer for fossil fuel CO<sub>2</sub>: Quantifying uncertainties using an atmospheric transport model, *J. Geophys. Res.*, *114*, D22302, doi:10.1029/2009JD012308.

E. Koffi and P. Peylin, Laboratoire des Sciences du Climat et de l'Environnement–Institut Pierre-Simon Laplace, Commissariat B l'Energie Atomique–Université de Versailles Saint-Quentin-en-Yvelines, CNRS, F-91191 Gif-sur-Yvette, France. (philippe.peylin@lscce.ipsl.fr)

M. Paget and M. R. Raupach, CSIRO Marine and Atmospheric Research, GPO Box 3023, Canberra, ACT 2601, Australia. (matt.paget@csiro.au; michael.raupach@csiro.au)

P. J. Rayner, School of Earth Sciences, University of Melbourne, Grattan Street, Parkville, Vic 3010, Australia. (prayner@unimelb.edu.au)



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

Rayner, P. J.; Raupach, M. R.; Paget, M.; Peylin, P.; Koffi, E.

**Title:**

A new global gridded data set of CO<sub>2</sub> emissions from fossil fuel combustion: methodology and evaluation

**Date:**

2010

**Citation:**

Rayner, P. J., Raupach, M. R., Paget, M., Peylin, P. & Koffi, E. (2010). A new global gridded data set of CO<sub>2</sub> emissions from fossil fuel combustion: methodology and evaluation. *Journal of Geophysical Research: Atmospheres*, 115, doi:10.1029/2009JD013439.

**Publication Status:**

Published

**Persistent Link:**

<http://hdl.handle.net/11343/32771>

**File Description:**

A new global gridded data set of CO<sub>2</sub> emissions from fossil fuel combustion: methodology and evaluation