AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2013JD021296

Link to FFDAS data retrieval and visualization: http://hpcg.purdue.edu/FFDAS/index.php

Key Points:
- Gridded high-resolution time series of global fossil fuel CO_2 emissions is generated
- In situ remote sensing data integrated into fossil fuel data assimilation system
- Spatial patterns of recession are analyzed using CO_2 emissions time series

Supporting Information:
- Readme
- Texts S1 and S2
- Figure S1

Correspondence to:
S. Asefi-Najafabady,
salvi.asefi@asu.edu

Citation:
Asefi-Najafabady, S., P. J. Rayner, K. R. Gurney, A. McRobert, Y. Song, K. Coltin, J. Huang, C. Elvidge, and K. Baugh (2014), A multiyear, global gridded fossil fuel CO_2 emission data product: Evaluation and analysis of results, *J. Geophys. Res. Atmos.*, *119*, doi:10.1002/2013JD021296.

Received 3 DEC 2013
Accepted 6 AUG 2014

A multiyear, global gridded fossil fuel CO_2 emission data product: Evaluation and analysis of results

S. Asefi-Najafabady[1], P. J. Rayner[2], K. R. Gurney[1,3], A. McRobert[2], Y. Song[1], K. Coltin[1], J. Huang[1], C. Elvidge[4], and K. Baugh[4]

[1]School of Life Sciences, Arizona State University, Tempe, Arizona, USA, [2]School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia, [3]Global Institute of Sustainability, Arizona State University, Tempe, Arizona, USA, [4]National Geophysical Data Center, National Oceanic and Atmospheric Administration (NOAA), Boulder, Colorado, USA

Abstract High-resolution, global quantification of fossil fuel CO_2 emissions is emerging as a critical need in carbon cycle science and climate policy. We build upon a previously developed fossil fuel data assimilation system (FFDAS) for estimating global high-resolution fossil fuel CO_2 emissions. We have improved the underlying observationally based data sources, expanded the approach through treatment of separate emitting sectors including a new pointwise database of global power plants, and extended the results to cover a 1997 to 2010 time series at a spatial resolution of 0.1°. Long-term trend analysis of the resulting global emissions shows subnational spatial structure in large active economies such as the United States, China, and India. These three countries, in particular, show different long-term trends and exploration of the trends in nighttime lights, and population reveal a decoupling of population and emissions at the subnational level. Analysis of shorter-term variations reveals the impact of the 2008–2009 global financial crisis with widespread negative emission anomalies across the U.S. and Europe. We have used a center of mass (CM) calculation as a compact metric to express the time evolution of spatial patterns in fossil fuel CO_2 emissions. The global emission CM has moved toward the east and somewhat south between 1997 and 2010, driven by the increase in emissions in China and South Asia over this time period. Analysis at the level of individual countries reveals per capita CO_2 emission migration in both Russia and India. The per capita emission CM holds potential as a way to succinctly analyze subnational shifts in carbon intensity over time. Uncertainties are generally lower than the previous version of FFDAS due mainly to an improved nightlight data set.

1. Introduction

The ability to accurately and reliably project atmospheric carbon dioxide (CO_2) concentrations, the primary driver of anthropogenic climate change, rests on understanding many aspects of carbon exchange between the atmosphere, land, and oceans [*Le Quéré et al*., 2009]. How the terrestrial biosphere and oceans respond to increasing CO_2 concentrations and a changing climate will determine, in large part, future CO_2 levels since over one half of current anthropogenic CO_2 emissions are removed by these two reservoirs [*Friedlingstein and Prentice*, 2010].

Among the global annual net carbon fluxes between the land, ocean, and atmosphere, the CO_2 flux resulting from the combustion of fossil fuels is the largest. Indeed, it is the geologically recent perturbation to the global carbon cycle by fossil fuel combustion and large-scale deforestation emissions that has given rise to increasing greenhouse gas forcing and the transient adjustments in flux between the land, oceans, and the atmosphere.

Atmospheric CO_2 inversion/assimilation and carbon budget studies rely on an accurate quantification of fossil fuel CO_2 emissions to understand where, when, and how carbon is exchanged with the terrestrial biosphere and ocean surface [*Enting*, 2002; *Gurney et al*., 2002]. It has been shown that errors in either the location or timing of fossil fuel CO_2 fluxes can be aliased into the remaining flux components of carbon inversion studies [*Gurney et al*., 2005]. Accurate quantification of the fossil fuel CO_2 flux is not only critical for improved understanding of the global carbon cycle but is equally important to climate policy. Cities, provinces, and countries use fossil fuel CO_2 flux quantification as a means to baseline their emissions for reporting purposes, better understand their options for emissions reductions, and assess progress on reductions [*Gurney*, 2013]. With the rise of carbon markets, fossil fuel CO_2 emissions quantification may be

called on to act as verification of traded emissions or policy measures [*National Research Council*, 2010]. Indeed, monitoring, reporting, and verification have emerged as a distinct need within the international climate change policy community for which better independent quantification of fossil fuel CO_2 emissions figure prominently [*Niederberger and Kimble*, 2011].

Fossil fuel CO_2 emission data sets with comprehensive global coverage are available at two different spatial representations and constructed through very different means: whole-nation estimates and subnational, gridded estimates. The latter rely, to varying degrees, on the former. Country-scale emissions are tabulated by five organizations. The International Energy Agency (IEA) generates national fossil fuel CO_2 with detail on economic sector and derives the information primarily from national energy surveys and emission factors based on Intergovernmental Panel on Climate Change (IPCC) guidelines [*International Energy Agency* (*IEA*), 2011; *Intergovernmental Panel on Climate Change* (*IPCC*), 2006]. The United States Energy Information Administration produces national fossil fuel CO_2 emissions with detail on fuel type (coal, petroleum, and natural gas) derived from a large list of primary energy consumption sources [*IEA*, 2011]. The Carbon Dioxide Analysis Information Center (CDIAC) provides national fossil fuel CO_2 emissions by fuel type and utilizes energy statistics published by the United Nations [*Marland and Rotty*, 1984; *Boden et al*., 1995]. The Joint Research Centre of the European Commission and the Planbureau voor de Leefomgeving Netherlands Environmental Assessment Agency combine efforts to produce the Emission Database for Global Atmospheric Research (EDGAR), a data product that includes many emitted species beyond fossil fuel CO_2 and provides sector, subsector, and fuel type detail [*Olivier et al*., 2005]. The energy data are derived primarily from the IEA national surveys. Finally, the United Nations Framework Convention on Climate Change (UNFCCC) collects CO_2 emission estimates at the national level with detail on sector, subsector, and fuel, although the level of detail in reporting varies according to international agreement [*IPCC*, 1996]. UNFCCC emissions are produced by nations utilizing a variety of internal and external data sources but follow methods outlined by the IPCC [*IPCC*, 2006]. Although the five national data sets are independently produced, they share some aspects of underlying data [*Macknick*, 2011; *Andres et al*., 2012].

The second spatial resolution for which global fossil fuel CO_2 emissions are available is in gridded form and relies primarily on downscaling techniques using spatial proxies. Regularized gridding is particularly useful for use in atmospheric transport models. The CDIAC produces a monthly fossil fuel CO_2 emission data product spanning the time period 1950 to 2010 that distributes the national totals into $1° \times 1°$ grid cells according to population density. For country-year pairs with no monthly data, monthly fractions are determined using proxy assignments and Monte Carlo techniques [*Andres et al*., 1996, 2011, 2012]. The EDGAR data product provides annual estimates on a $0.1° \times 0.1°$ grid spanning the time period 1990 to 2010 and distributes national emissions with a variety of spatial proxies ranging from population density to specific point source location maps. The Open Source Data Inventory of Anthropogenic CO_2 Emission (ODIAC) combines downscaling of national emissions using satellite observations of nighttime lights and a geocoded estimation of power plant CO_2 emissions to generate fossil fuel CO_2 emissions on a 1 km grid from 1980 to 2007 [*Oda and Maksyutov*, 2011]. Finally, a recent effort by *Wang et al*. [2013] combined elements of downscaling with bottom-up data sources to create fossil fuel CO_2 emissions at $0.1° \times 0.1°$ resolution for the year 2007. For example, province/state level fuel consumption was used where available and downscaled using a combination of CO emission inventories or population as spatial proxies, depending upon availability of data.

Over the last decade, there has been an increasing emphasis on building globally gridded fossil fuel CO_2 emission data products at finer space and time scales. This has been driven by the advent of CO_2 measurement from remote sensing platforms, able to quantify column concentration at spatial scales of a few kilometers to tens of kilometers and hourly to weekly time scales [*Yokota et al*., 2009; *Crisp et al*., 2012; *Schneising et al*., 2013; *Bovensmann et al*., 2010]. These observational platforms combined with the growing surface network of greenhouse gas monitoring by ground stations and aircraft have allowed for carbon cycle studies to analyze fluxes at these space and time scales, for which fossil fuel CO_2 emissions are a critical ingredient.

A recent effort by *Rayner et al*. [2010] approached the problem of globally gridded fossil fuel CO_2 emission quantification through a technique that combined some elements of downscaling, bottom-up information, and data assimilation within a model of fossil fuel CO_2 emissions to optimally disaggregate national emissions to a 0.25° global grid. Called the fossil fuel data assimilation system (FFDAS), *Rayner et al*. [2010]

demonstrated that the assimilation approach has a few critical advantages over spatial downscaling efforts. First, at the 0.25° resolution they used, the assimilation approach combining nighttime lights and population produced a better match to the Vulcan data product, the most intensive bottom-up fossil fuel CO2 emission effort to date [*Gurney et al*., 2009] than using either data source alone. Second, FFDAS provided a posterior uncertainty at each grid cell as desired for inverse assimilation efforts on the entire carbon cycle. Finally, the approach is flexible to expansion by economic sector or spatial domain, allowing for new observational operators to condition subsets of the emissions.

In this paper, we build upon the foundation provided by *Rayner et al*. [2010] and update and expand the FFDAS results (FFDAS v2). Instead of a single year, we provide complete global 0.1° coverage from 1997 to 2010. This is constructed from a new multiyear, saturation-corrected Defense Meteorological Satellite Program (DMSP) nightlight data product. We also utilize the recently released multiyear LandScan population data product available for the time period 2000 to 2010. Finally, we utilize an updated and improved global power plant emission data product that now includes improved location information and individual power plant uncertainties. We compare these results to the Vulcan data product in the United States [*Gurney et al*., 2009]. Finally, we analyze trends across the globe, highlighting where emissions growth has been large and sustained over the time period examined.

2. Methodology

2.1. FFDAS

The underlying methodology of data assimilation applied to the problem of estimating fossil fuel CO_2 emissions is described by *Rayner et al*. [2010]. The main difference between this study and *Rayner et al*. [2010] is the treatment of multiple emission sectors and improvement to the quality of the observational constraints used.

Although *Rayner et al*. [2010] only used one sector, they foreshadowed an approach in which the Kaya identity was used for each sector with different factors relating population density to emissions. This implicitly uses population as the basis for each sector. For several important sectors, population is not a natural basis (e.g., power generation). Thus, we now permit each sector to have an independent model of emissions. In this version, we define two sectors: a power generation sector and all other emissions (denoted "other"). The procedure used to estimate emissions in the other sector is *identical* to FFDAS v1, except in FFDAS v2, we utilize updated nightlights and population data sets (discussed in section 2.2).

The emissions at a point, x in country, c are given by

$$F(x,c) = pP(x)g(x)ef(c) \tag{1}$$

where lower case roman variables are the unknowns for which we solve, and we use parentheses to make their dependence explicit. P is a prescribed population density. The unknowns are a scalar multiplier for the population density: a spatially dependent intensity, g, relating economic activity to population; a constant intensity (hence scalar), e, relating energy use to economic activity; and a country-dependent carbon intensity, f, relating carbon emissions to energy use.

There are two observables for the other sector. The first is the national emissions given by

$$E(c) = \sum_{x}^{c} A(x)pP(x)g(x)ef(x) \tag{2}$$

where $A(x)$ is the area of grid cell, x. The second is the nighttime lights given by

$$N(x) = npP(x)g(x) \tag{3}$$

where n is a normalization constant for which we must solve.

The normalization, n, accounts for a conversion of units from gross domestic product (GDP) to nightlight digital numbers. We solve for n each year so its value also absorbs any problem with intercalibration of the nightlights. This does not affect emission estimates, which are normalized by the national totals. It does mean that we cannot comment on trends in g and n separately, since they are not separately observed.

In a general optimization procedure, a series of control variables in a predictive model of emissions is optimized relative to a set of observations and prior estimates. The optimization consists of the minimization of a quadratic

cost function, a choice that implicitly assumes Gaussian probability densities for prior estimates and observations. In the application here, we solve the optimization problem by minimizing the cost function

$$\chi^2 = \sum_{s \in S+D} \frac{[V(s) - V_0(s)]^2}{{\sigma_s}^2} + \sum_{d \in D} \frac{[M_d(V) - O_d]^2}{{\sigma_d}^2} \tag{4}$$

where S is the set of emission sectors, D is the set of observables, V is the control variable for a sector, M is the model for a particular observable, O is the measured value for that observable, σ is the uncertainty, and the zero subscript represents the prior estimate. The minimization in equation (4) indicates that model outcomes are adjusted to match the observations as closely as is reasonable while not straying too far from prior estimates. The weightings given to both terms in the cost function are determined by the uncertainty, σ. Note that the prior term includes contributions from both the model variables and observables, since there may be control variables attached to particular observations but not used in calculating fluxes. An example is n, the normalization constant for nightlights.

Combining equations (1)–(4) yields

$$\chi^2_{\text{other}} = \frac{(n-n_0)^2}{{\sigma_n}^2} + \frac{(p-p_0)^2}{{\sigma_p}^2} + \sum_x \frac{[g(x)-g_0(x)]^2}{{\sigma_g(x)}^2} + \frac{(e-e_0)^2}{{\sigma_e}^2} + \sum_c \frac{[f(c)-f_0(c)]^2}{{\sigma_f}^2} + \sum_x \frac{[npP(x)g(x) - N(x)]^2}{{\sigma_N}^2} + \sum_c \frac{\left[\sum_{x \in c} A(x)\, pP\,(x)\, g\,(x)\, ef\,(c) - F(c)\right]^2}{{\sigma_F}^2} \tag{5}$$

where the index c ranges over countries and x ranges over grid cells.

The variables in the power generation sector are emissions from a list of power plants in each country. These emissions were estimated for the year 2009. For all other years, power plant emission estimates from 2009 were optimized by FFDAS based on the prior uncertainties (see supporting information for more information on calculation of prior uncertainties) such that the sum of all power plant emissions match the IEA national electricity generation total for that year and country. Emissions associated with each power plant were mapped onto the FFDAS grid based on the power plant location. If power plants were located inside a grid cell, the grid cell was given the emission value with no subgrid cell spatial distribution. When location information was not available for a power plant, the power plant emissions were proportionally distributed to the rest of the grid cells containing power plants in that country.

The cost function for this sector is hence

$$\chi^2_{\text{elec}} = \sum_i \frac{[u(i) - u_0(i)]^2}{{\sigma_u(i)}^2} + \sum_c \frac{\left[\sum_{i \in c} u(i) - U(c)\right]^2}{{\sigma_u(c)}^2} \tag{6}$$

where u (utility) is the emission from a particular power plant and U is the national emission from the utility sector. The index, i represents individual power plants. The individual power plant emissions $u(i)$ are the control variables optimized in this sector's assimilation procedure. In this instance, the location information for the power plants is irrelevant beyond the country in which they reside. Note that no variables or observables are present in both equation (5) and equation (6), so that the two sectors can be optimized independently.

In summary, the control variables are a scalar multiple for the global population, a pointwise estimate of the per capita GDP, a global constant for the energy intensity of the economy, a countrywise estimate for the carbon intensity of energy production, a normalization constant for nightlights, and the emissions from each listed power station.

2.2. Input Data

Although FFDAS v2 incorporates a variety of data sets at varying resolutions, the current working resolution of the resulting data product and the analysis carried out in this paper is $0.1° \times 0.1°$ and spans the years 1997–2010. The following data products are used in FFDAS.

2.2.1. National Emissions

FFDAS utilizes national total fossil fuel CO_2 emissions as a key constraint to the application of spatial operators. The International Energy Agency (IEA) estimates CO_2 emissions from fuel combustion using energy

surveys. The IEA [*IEA*, 2011, http://www.iea.org/] estimated national emissions for 137 individual countries and 3 country aggregates (a collection of smaller countries in Asia, Africa, and Latin America), and we utilize annual data spanning the time period 1997 to 2009. At the time of analysis, IEA emissions data were not available for 2010. Therefore, we applied the 2009 to 2010 country-level, annual growth ratio derived from the Carbon Dioxide Information and Analysis Center's (CDIAC) national emissions data [*Carbon Dioxide Information Analysis Center*, 2010, http://cdiac.ornl.gov]. CDIAC estimates 2009 and 2010 emissions for 67 countries. IEA countries for which there was no available CDIAC ratio (50 countries, 3 aggregates) were assigned regional mean ratios. The regions used are defined by *Raupach et al.* [2010], in which nations are clustered into nine regions according to their emissions and economic profiles. The regions are the United States, Europe, Japan, D1 (other developed countries), Former Soviet Union countries, China, India, D2 (other developing countries), and D3 (least developed countries as defined by the United Nations).

In addition to total national emissions, the IEA subdivides emissions into economic sectors and subsectors according to IPCC definitions. We aggregate these subdivisions into two for the purposes here. The first encompasses all power-producing facilities including both grid-connected and "autoproducer" facilities (generation of electricity for internal manufacturing or similar uses). All other emitting categories are aggregated into a single, second category that includes industrial, residential, commercial, and transportation emissions. Domestic and international aviation and waterborne vehicle emissions were removed from the IEA transportation estimates, as these are included in independent data for global aviation and commercial shipping, which we append as an off-line gridded data product to the FFDAS result [*Olivier et al.*, 2005].

IEA does not provide national emissions for cement production and gas flaring; therefore, emissions from these categories are excluded from FFDAS v2 calculations.

The IEA national emissions do not include an estimate of uncertainty. Indeed, little research has been performed on national emissions uncertainty beyond the use of expert judgment [*Marland*, 2008; *Marland et al.*, 2009]. As an attempt at an objective measure, we utilize recent analysis on national/global energy consumption and CO_2 emissions [*Macknick*, 2011] in which five data sets (British Petroleum, CDIAC, U.S. Energy Information Administration, and two variants of the IEA data), representing different efforts at quantifying national CO_2 emissions, were harmonized in an attempt to remove known differences such as categorical definitions and parameter assumptions. By harmonizing the data sets, one can assess their differences on a more equal footing. We used this harmonized database to estimate national fossil fuel CO_2 emissions uncertainty. We define the span of the five estimates for each country as the difference between the maximum member and the minimum member for each year. This span is expressed as a percent through division by the mean of the five-member set. This is done for the 26 countries analyzed by *Macknick* [2011], and these 26 countries represent the top CO_2-emitting nations in the world. Because the analysis performed by *Macknick* [2011] ended in the year 2007, we use the same approach used to estimate the 2010 CO_2 national emissions. Namely, we have used country-level, annual growth ratios derived from the CDIAC national emissions data and applied these ratios to estimate the national span values for 2008 to 2010. On average, the mean percentage span value for the world is about 16%. The smallest 1997 to 2010 mean percentage span is 7.5% for Mexico, while the largest is 50% for South Korea. The percentage span values are: 12% for the United States, 10% for China, 18% for India, 13% for Brazil, 10% for Germany, and 13% for France. For countries beyond the 26 analyzed by *Macknick* [2011] (http://www.iiasa.ac.at/web/home/research/modelsData/EnergyCarbonDatabase.en.html), we assign percentage span values according to the average of country values within their global region, following the regional definitions of *Raupach et al.* [2010], described previously in this text. In the case of the *Raupach et al.*'s [2010] D3 region, there was no representative country in the top 26 as defined by *Macknick* [2011], so the value used for the D2 region was adopted. The percentage span was divided by 4 to represent a 1 sigma uncertainty value and is independently estimated for every year. This value (4) is a rough conversion factor used here to convert range (percentage span) to standard deviation. Such conversion factors depend on the sample size, but for large samples, a factor of 4 is generally used [*Hozo et al.*, 2005].

2.2.2. Gridded Population Density

We use the gridded population density data set (2000–2010) provided by LandScan Global Population Database via East View Geospatial (www.geospatial.com). LandScan represents ambient population (the average over 24 h that takes into account the movement of people for work, travel, and sleep) at 0.00833° × 0.00833° (30″) grid resolution [*Bhaduri et al.*, 2007]. In order to extend the population information back to the year 1997, the first year in our emissions estimation time series, we use 1995 and 2000 data from the Gridded Population of the World

(GPW v3) data set provided at 0.04° × 0.04° grid resolution created by the Center for International Earth Science Information Network (http://sedac.ciesin.columbia.edu/gpw/index.jsp). Linear interpolation was performed at the grid cell level to create an annual population time series for the time period 1995–2000 at the GPW 0.04° resolution. Then for the years 1997, 1998, and 1999, the GPW data product was used but downscaled to 0.00833° × 0.00833° resolution using the LandScan spatial distribution from the year 2000.

2.2.3. Power Plant Emissions

As a pointwise prior estimate of fossil fuel CO_2 emissions from the electricity production sector, two different approaches to populate a database of the world's power plants are utilized (see supporting information). Where CO_2 information on power plant emissions and location is publicly disclosed, the information is used directly. Publicly disclosed data sets are available for U.S., India, Canada, South Africa, and the European Union (links and references are presented in the supporting information). For all remaining power plants, a multivariate regression model (see supporting information) is constructed based on the power plants which are present in both the publicly disclosed data and the World Electric Power Plants (WEPP) database. The WEPP database is a comprehensive global inventory of electric power-generating units. Containing no information on electricity generation or CO_2 emissions, the database does contain information on ownership, location, and engineering design data for power plants of all sizes and technologies operated by regulated utilities, private entities, and industrial or commercial self or auto producers across the globe. These emissions were estimated for the year 2009. More details associated with the regression model, calculation of prior power plant emission uncertainties, and results can be found in the supporting information. In the current version of FFDAS, we have not estimated errors associated with power plant locations nor have we accounted for commissioning and decommissioning of power plants. However, the locations of the top ~800 emitting power plants were corrected through a combination of visual inspection in Google Earth and additional information provided on individual facility webpages.

2.2.4. DMSP Radiance Lights

The Operational Linescan System (OLS) flown on the Defense Meteorological Satellite Program (DMSP) has a unique capability to record low-light imaging data at night worldwide. Data are archived at the National Oceanic and Atmospheric Administration National Geophysical Data Center (NGDC) (http://ngdc.noaa.gov/eog/index.html). The digital data record stretches back to 1992 and is ongoing. The sensor is typically operated in a high gain setting to enable the detection of moonlit clouds. However, with 6 bit quantization and limited dynamic range, the recorded data are saturated in the bright cores of urban centers and there is no in-flight calibration. In the FFDAS v1, a correction was applied to DMSP nightlight products based on an analytical model that was developed by *Raupach et al.* [2010]. In the FFDAS v2 presented here, a new nightlight data set (radiance light) was developed by NGDC that is no longer subject to severe saturation errors. *Elvidge et al.* [1999] developed procedures to combine the data collected at multiple gain settings to extend the dynamic range of OLS nighttime light products. The initial procedures were later refined to better handle subpixel saturation and blending with stable lights to extend the detection of dim lighting into suburban and populated rural areas [*Ziskin et al.*, 2010]. With a stable set of procedures in place, NGDC reprocessed the fixed gain collections and blended them with the stable light products from the same years. The products were then intercalibrated, and gas flares were masked using the gas flare mask developed by NGDC (based on techniques explained by *Elvidge et al.* [2009]). A total of six products were produced for the years 1997, 1999, 2000, 2003, 2004, 2006, and 2010. Using this product, we applied a linear interpolation and created a time series of nighttime lights for 1997 to 2010.

Like any other observable in a data assimilation system, the nightlights uncertainty has contributions from the observation itself and the model used to simulate it. In this case, the model error that can not be captured by the uncertainty in the prior parameter is relatively small, since there is an independent degree of freedom, $g(x)$, attached to each observation. Observation errors arise from the quantization of the 6 bit detector plus the task of converting the digital counts to radiances. There is considerable averaging as the native 30″ data are aggregated to the 0.1° grid needed for use in FFDAS. We settled on an uncertainty specification of

$$\sigma_{NL} = 0.5 + 0.1 x_{NL} \tag{7}$$

where X_{NL} is the observed nightlight value. We inflate this uncertainty by 25% in years with interpolated nightlights to allow for interpolation error. These values are checked in the variance tuning procedure outlined in the next section.

2.3. Variance Tuning

Consistency places conditions on the value of the minimum cost function and contributions to it. Put briefly, the statistics of the differences between model outputs and observations must agree with the distributions assumed in the statistical formulation, e.g., normally distributed and independent. Similar conditions apply to the prior estimates and also to any subpopulation of the data. This has been used by *Michalak et al*. [2004] and *Wu et al*. [2013] to estimate parameters of the distribution. Here we limit ourselves to the overall cost function and a rough check on the major data types. In a consistent optimization, the final cost function should approximately equal the number of observations [*Michalak et al*., 2004]. Higher values suggest that the fit to data and prior values is worse than we would expect from the uncertainties. The usual solution is to increase uncertainties to restore consistency. Lower values suggest that input uncertainties (on data or prior estimates) are larger than necessary. Large input uncertainties imply large (hence, conservative) posterior uncertainties.

Cost functions in the 14 years of FFDAS optimizations range from about 0.66 times the number of observations up to about 1.1. There is no clear difference between years with directly observed and interpolated nightlights. Overall, the values suggest that our uncertainty choices are conservative.

2.4. Calculation of Posterior Uncertainties

We follow *Chevallier et al*. [2007] and use a Monte Carlo technique for estimating the posterior uncertainties of the assimilated variables and related fluxes. *Rayner et al*. [2010] estimated the uncertainty of the model variables and provided an algorithm for propagating these through the Kaya identity to generate uncertainties on posterior fluxes. This allowed the generation of complete covariance matrices with the appropriate error correlations. The addition of the electricity generation sector means we can no longer use the Kaya identity for this propagation. Thus, we utilize Monte Carlo realizations of the posterior fluxes at the native 0.1° resolution. We calculate 10 Monte Carlo realizations for years with observed and interpolated nightlights. This is sufficient to give reasonable guidance for downstream users of FFDAS, especially since most will aggregate uncertainties and thereby smooth sampling noise.

2.5. Center of Mass Calculations

A center of mass (CM) is a useful and compact metric to understand and illustrate the spatial changes in fossil fuel CO_2 emissions over time [*Gregg et al*., 2009]. The CM summarizes the distribution of emissions in the same way as the mean summarizes a probability distribution. In order to calculate the CM for each year of the global FFDAS emissions, one must establish an origin point within which the CM is defined. We established this origin point as the equator at the date line. Then longitudinal distance to the center of mass (relative to date line) was calculated as a weighted average of the magnitude of emissions—weighted by distance from the origin. The two dimensions of the CM (X_{CM} and Y_{CM}) can be represented as

$$X_{CM} = \frac{\sum_i E_i \Delta \mathrm{lon}_i}{\sum_i E_i} \tag{8}$$

$$Y_{CM} = \frac{\sum_i E_i \Delta \mathrm{lat}_i}{\sum_i E_i} \tag{9}$$

where E represents the CO_2 emission value for each pixel, Δlat is the latitudinal distance, Δlon is the longitudinal distance from origin to grid cell i (index i ranges over global grid cells), X_{CM} is the longitude and Y_{CM} is the latitude of CM.

3. Results

3.1. Spatial Flux Patterns

Figure 1 presents a global map of FFDAS v2 fossil fuel CO_2 emissions (kg C m^{-2} yr^{-1}) at 0.1° resolution in 2010. One of the improvements in FFDAS v2 is the spatial resolution (0.1° versus 0.25°). As evident in Figure 2, this increased spatial resolution reveals more emission details.

For example, Figure 2 shows the emergence of large roadways in the emissions over China, which were not obvious in the previous 0.25° resolution data product. The explicit inclusion of the power plant emissions as a

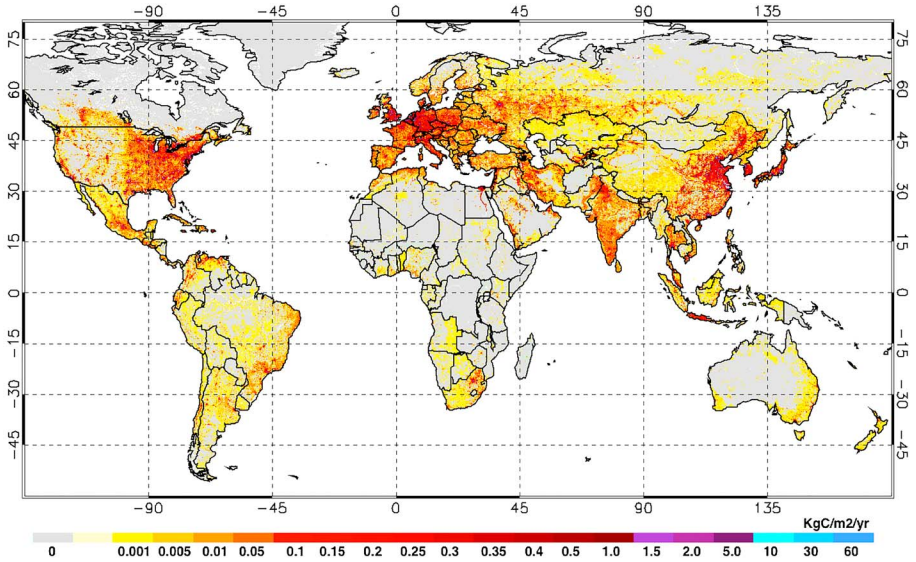

Figure 1. Global map of FFDAS v2 total fossil fuel CO_2 emissions at 0.1° resolution in 2010. Units: kg C m^{-2} yr^{-1}.

pointwise data set also has an impact on the spatial patterns of the final estimate of emissions (Figure 3). Separate estimation of the electricity generation at geocoded points reduces the amount of emissions spread into the populated and lighted areas within a given country. This creates a much more physically realistic spatial representation of actual emissions. For example, 41% of the total emissions in Germany are produced by the power plants. In the previous FFDAS v1, all German emissions would be distributed across the populated/lighted landscape instead of just 59%. Figures 3a and 3b show the effect of this change.

We also calculated uncertainties following section 2.4 using 10 Monte Carlo realizations. Uncertainties in the other sector combine those from nightlights and national emissions. In fractional terms, the errors will roughly sum as with the analysis of fractional trends below. The form for the nightlights uncertainty means that fractional uncertainties will be highest at low nightlight levels, unlike the case of *Rayner et al*. [2010], where uncertainties were driven by the saturation correction that was necessary at the time of FFDAS v1. Once we remove very small mean values, the fractional uncertainty is about 14%. We note again that uncertainties in the other sector are positively correlated within one country.

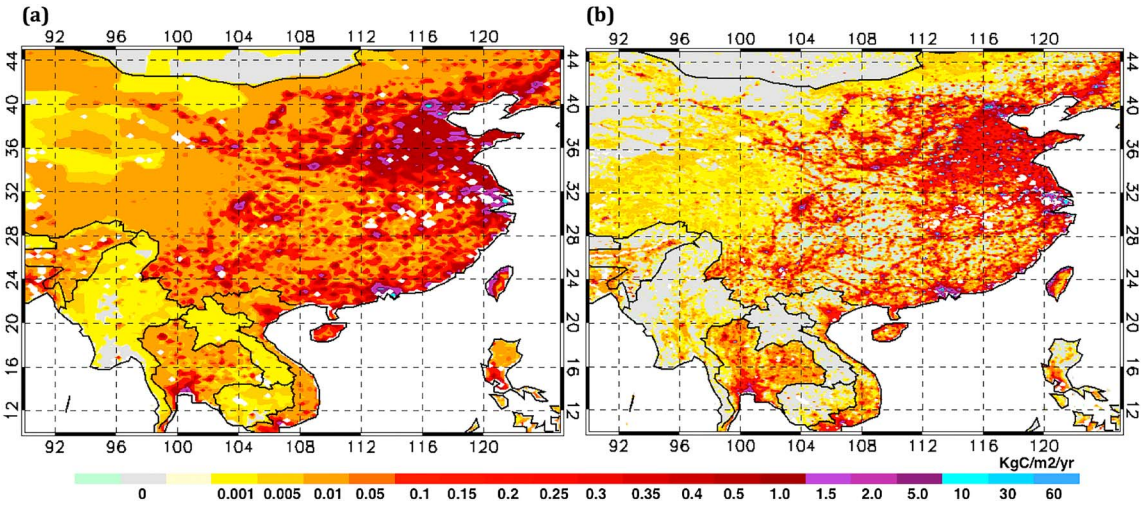

Figure 2. Comparison of total fossil fuel CO_2 emissions for Asia in 2002. (a) FFDAS v1 at 0.25° resolution. (b) FFDAS v2 at 0.1° resolution. Emissions include both power plant and other sector emissions. Units: kg C m^{-2} yr^{-1}.

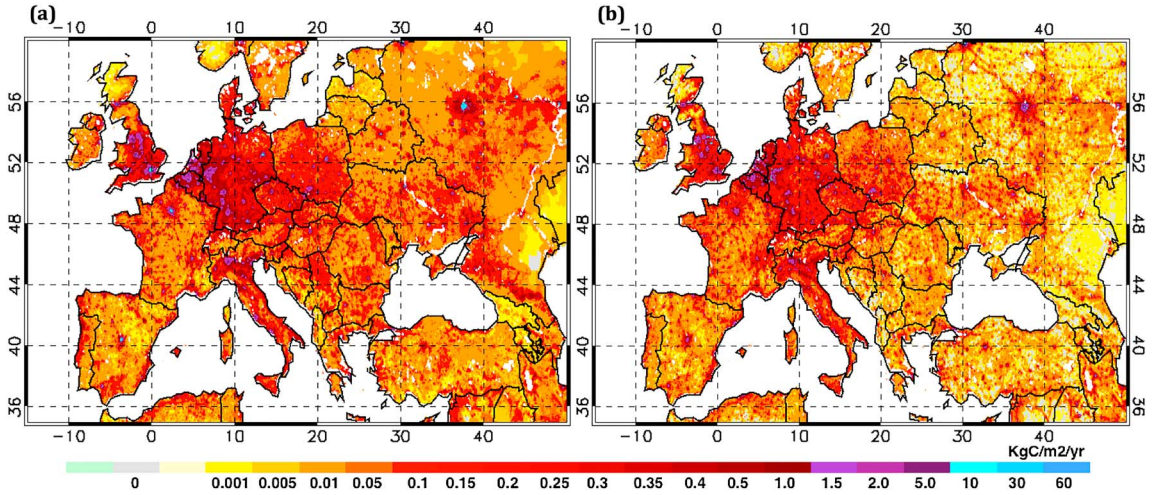

Figure 3. Comparison of FFDAS v2 total fossil fuel CO_2 emissions in 2002 when (a) power plant emissions are not allocated to power plant locations but distributed by nightlights and population and (b) power plant emissions are allocated to power plant locations. Units: kg C m^{-2} yr^{-1}.

The case for the utility sector is different. Pointwise uncertainties for individual plants are not much reduced by the assimilation system, but the constraint of a national total introduces a pooled negative correlation among the posterior estimates.

In order to evaluate the model performance in optimizing posterior power plant emission estimates, we calculated the root-mean-square error (RMSE) of the difference between the prior and posterior values as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(\left(X_{\mathrm{posterior},i} - X_{\mathrm{prior},i} \right) / \sigma_{\mathrm{priorunc},i} \right)^2}{n}} \tag{10}$$

where the index, i represents the individual power plants, $X_{\mathrm{posterior},i}$ is the optimized estimated CO_2 emission for each power plant, $X_{\mathrm{prior},i}$ represents the prior estimated CO_2 emission for each power plant (see supporting information), $\sigma_{\mathrm{priorunc},i}$ represents the estimated prior uncertainty for each power plant (see supporting information), and n is the total number of power plants. We obtain RMSE values between 0.4 and 0.7 for the years 1997 to 2010, suggesting that the deviation of the posterior from prior estimates is within an acceptable range, and therefore, the choice of power plant uncertainties are reasonable.

3.2. Comparison to Bottom-Up Emissions

Due to limited emissions quantification at similar space/time scales, evaluation of FFDAS remains challenging. Although not global in coverage, regional efforts have begun that quantify fossil fuel CO_2 emissions from a bottom-up perspective and offer a useful point of comparison. The Vulcan fossil fuel emissions data product produced by *Gurney et al*. [2009] for the U.S. is considered a state of the art bottom-up approach to fossil fuel emissions quantification. Although the original information in the Vulcan data product is estimated as points, lines, and polygons, the emissions have been regularized onto a 0.1° grid with a temporal resolution of 1 h. Vulcan analyzes fossil fuel CO_2 emissions by economic sector and includes process-level detail such as combustion technology, fuel type, and vehicle class. We compare the FFDAS v2 with the 2002 Vulcan emissions data product for the U.S. domain. Following *Rayner et al*. [2010], we use two metrics: (1) summed absolute difference (SAD), which is the sum of the absolute difference of the field over the domain and (2) spatial correlation, which quantifies the magnitude-independent correspondence of the spatial patterns. The resolution of both data products is varied from 0.1° × 0.1° to 4.0° × 4.0° through aggregation to test how resolution influences the level of agreement. Furthermore, we run FFDAS v2 in which only population acts to distribute the "other" national emissions -"FFDAS total (POP)," only nightlights act to distribute national emissions "FFDAS total (NL)," and the optimal combination of the two "FFDAS total (POP and NL)." Finally, we isolate and compare emissions other than those associated with electricity production- "FFDAS other (POP and NL).

Table 1. Comparison of Both the 2002 FFDAS v2 and the FFDAS v1 [*Rayner et al.*, 2010] Fossil Fuel CO_2 Emissions to the Vulcan Fossil Fuel CO_2 Emissions Data Product From *Gurney et al.* [2009][a]

	Other Sector		FFDAS Total (NL and POP)		FFDAS Total (NL)		FFDAS Total (POP)		FFDAS v1 (Total)	
Resolution	SAD	Correl	SAD	Correl	SAD	Correl	SAD	Correl	SAD	Correl
0.1	551	0.48	714	0.86	731	0.86	749	0.85	-	-
0.2	446	0.67	581	0.88	598	0.88	599	0.87	-	-
0.3	403	0.76	515	0.90	535	0.90	530	0.89	-	-
0.4	371	0.81	471	0.91	490	0.92	487	0.90	-	-
0.5	358	0.84	449	0.92	468	0.92	459	0.91	1143	0.74
1	301	0.91	355	0.95	368	0.95	378	0.95	900	0.85
2	243	0.94	260	0.97	274	0.97	304	0.96	651	0.91
3	211	0.96	230	0.98	233	0.98	254	0.98	545	0.92
4	186	0.97	200	0.99	205	0.99	255	0.97	479	0.93

[a]The FFDAS v2 total results are constructed as only population-based (FFDAS total (POP)), only nightlight-based (FFDAS total (NL)), and combined population- and nightlight-based (FFDAS total (NL and POP)). Results for the FFDAS v2 are also provided for just the other sector. The summed absolute difference (SAD) and spatial correlation (correl) are used as the comparison metric.

The results of these comparisons are presented in Table 1. The spatial correlation between the FFDAS results and Vulcan exceeds 0.8 for the FFDAS total (POP), FFDAS total (NL), and FFDAS total (POP and NL) cases at the 0.1° resolution. As the spatial resolution is decreased, the spatial correlation increases surpassing 0.9 at 0.5° and greater. At 4.0° resolution, the spatial correlation reaches 0.99. These correlation values are substantially higher than was found in the FFDAS v1, where a spatial correlation value of 0.74 was achieved at the 0.5° resolution [*Rayner et al.*, 2010]. The result in which only nightlights are used as the spatial proxy exhibits the highest correlation value regardless of resolution, although the differences between the three sensitivity cases are very small. We will return to this phenomenon later. Much of the high spatial correlation of the total emissions is due to the presence of the pointwise power plant emissions. Vulcan and FFDAS have nearly identical power plant emissions data, and hence, the total correlation is driven by the near-perfect agreement for this sector. This is shown quantitatively by the comparison of FFDAS other (POP and NL) to the Vulcan data product (for which we sum all sectors other than those attributed to electricity generation). This provides a more exacting test of how well the FFDAS distributes emissions within the United States. At 0.1° resolution, the spatial correlation value is 0.48 but increases to 0.97 when the resolution is aggregated to 4.0°. These values also represent an improvement over FFDAS v1, where a spatial correlation value of 0.74 was achieved at the 0.5° resolution (as compared to 0.84 here). The performance of the FFDAS other result when compared to Vulcan is explored further in Table 2. Here the FFDAS other result is compared to individual sectoral results from the Vulcan data product. These results demonstrate two aspects of the FFDAS. First, the overall correlation is highest in the residential and onroad transportation sectors, regardless of the spatial proxy applied. This suggests that population and nightlights are spatially correlated with fossil fuel combustion in homes and vehicles but have poor spatial correlation with emissions from pointwise manufacturing or the fuel consumption at airports. The other notable pattern from Table 2 is that the optimal spatial proxy

Table 2. Comparison of Sector-Specific 2002 FFDAS v2 Fossil Fuel CO_2 Emissions for the United States With Vulcan Fossil Fuel CO_2 Emissions Data Product From *Gurney et al.* [2009][a]

	FFDAS (NL and POP)		FFDAS (NL)		FFDAS (POP)	
Sectoral Comparison at 0.1°	SAD	Correl	SAD	Correl	SAD	Correl
FFDAS versus Vulcan all sectors	714	0.86	731	0.86	749	0.85
FFDAS other versus vulcan other	551	0.48	567	0.49	587	0.45
FFDAS other versus vulcan residential	740	0.69	791	0.69	733	0.84
FFDAS other versus vulcan onroad	552	0.82	587	0.83	577	0.76
FFDAS other versus vulcan nonroad	802	0.64	854	0.64	795	0.67
FFDAS other versus vulcan commercial	781	0.50	832	0.50	775	0.54
FFDAS other versus vulcan industrial	858	0.16	904	0.17	863	0.12
FFDAS other versus vulcan airports	820	0.25	871	0.24	815	0.18

[a]The FFDAS v2 results are constructed as only population-based (FFDAS (POP)), only nightlight-based (FFDAS (NL)), and combined population- and nightlight-based (FFDAS (NL and POP)). The summed absolute difference (SAD) and spatial correlation (correl) are used as the comparison metric.

depends upon the sector considered. For example, the spatial correlation value when FFDAS is compared to the Vulcan residential sector is highest with the use of population as the spatial proxy. This is not surprising given the fact that emissions from the residential sector are strongly tied to population density. By contrast, the correlation value for the onroad transportation sector is highest when using nightlights only. This suggests that the intensity of lighting is a better predictor of traffic density than population density. The combination of these two results demonstrates why further sectoral division will be needed to improve the FFDAS and more accurately estimate emissions. Indeed, the aggregation of sectoral activity into one category makes the combination of the two spatial proxies a poorer fit than either individually. With additional sectoral separation, a superior spatial proxy will be applied where it shows the best calibration to the Vulcan data product. That work is ongoing and will be included in a future FFDAS release. Evaluation of FFDAS versus the Vulcan data product in the U.S. has shortcomings. For example, many regions of the world have different industrial development trajectories with potentially different spatial patterns in terms of lighting distribution and their relationship to fossil fuel combustion. Furthermore, the accuracy of the population density likely varies with higher accuracy in locations such as the U.S., where detailed census systems are in place.

Although there are no products comparable to Vulcan in other parts of the world, for China, we can compare state/provincial emission totals for China, the U.S., and Australia with state/province-integrated FFDAS emissions [*Liu et al*., 2013; www.EIA.gov; http://ageis.climatechange.gov.au]. The resulting correlations from this analysis were 0.60 for the U.S., 0.60 for Australia, and 0.57 for China. This suggests that FFDAS has skill in distributing emissions at the subnational level. Both the representation of lighting and population accuracy will vary within countries. For example, China and India have large urban centers with pervasive electricity delivery and dense lighting. However, there are areas in which, although considerable population is present, limited lighting exists (dark emission areas). Indeed, in the U.S., lighting and population density are highly correlated ($r^2 = 0.85$), explaining the small spatial correlation differences between the FFDAS and Vulcan in the U.S. (Table 1) regardless of the spatial proxy utilized. Therefore, if only nightlight is utilized as the principal spatial proxy for emission distribution, emissions will be overallocated to the more developed, lighted portions of countries for which there are larger emissions than might be found in the United States. The problem associated with dark emission areas was noted by *Oda and Maksyutov* [2011] and addressed by *Ghosh et al*. [2010], in which the authors used nightlights and population to separately distribute emissions globally. In dark emission areas, they used population only and assumed that per capita emissions were 50% of the per capita emissions of lighted areas, which relied on nightlights only. FFDAS, by contrast, uses an assimilation approach which incorporates observational uncertainties, and therefore, it is not required to follow observations exactly. Thus, even dark areas can produce emissions. Refining the treatment of dark areas is a topic for future work.

3.3. Long-Term Trends

The spatial and sectoral breakdown of emissions in FFDAS over a multiyear time series allows for analysis of fossil fuel CO_2 emission trends at subnational scales. A linear least squares regression was used to calculate CO_2 emission trends for each grid cell (Figure 4). We also calculated uncertainties in the trend by propagating uncertainties from the FFDAS annual posterior uncertainty values. *Raupach et al*. [2007] used the derivative of the Kaya identity to estimate contributions to the growth in emissions [*Nakicenov*, 2004]. Our use of the Kaya identity for the other sector means we can apply the same analysis at the grid cell scale. This requires care since there is more noise in the various input data sets at the 0.1° scale, and this noise is amplified when calculating fractional trends for low mean values. Figure 4 shows fractional trends in total emissions, population, and nighttime lights, respectively. A minimum mean threshold has been used when calculating the percentage trends for the Kaya variables to remove some of this noise. A minimum of 10^{-5} people/m^2/yr is used for population and 1 DN/m^2/yr for nightlights ("DN" refers to digital number). For emissions, we eliminate all the points for which the absolute value of the trend is smaller than its standard deviation, and a minimum value of 0.01 kg C m^{-2} yr^{-1} has been applied to mean fossil fuel CO_2 emissions. Large positive emission trend values occur in China, portions of India, and South Asia, consistent with national statistics on both energy use and CO_2 emissions. Nearly 25% of grid cells in India and China have emission trends of 5%/yr or greater, while 15% of the grid cells in eastern China have emission trends of 10%/yr or greater. The large subnational trends in the southern and northwestern portions of India encompass the cities of Bangalore, Chennai, and New Delhi. The large growth seen in the eastern half of China encompasses both the large eastern urban centers and industrial

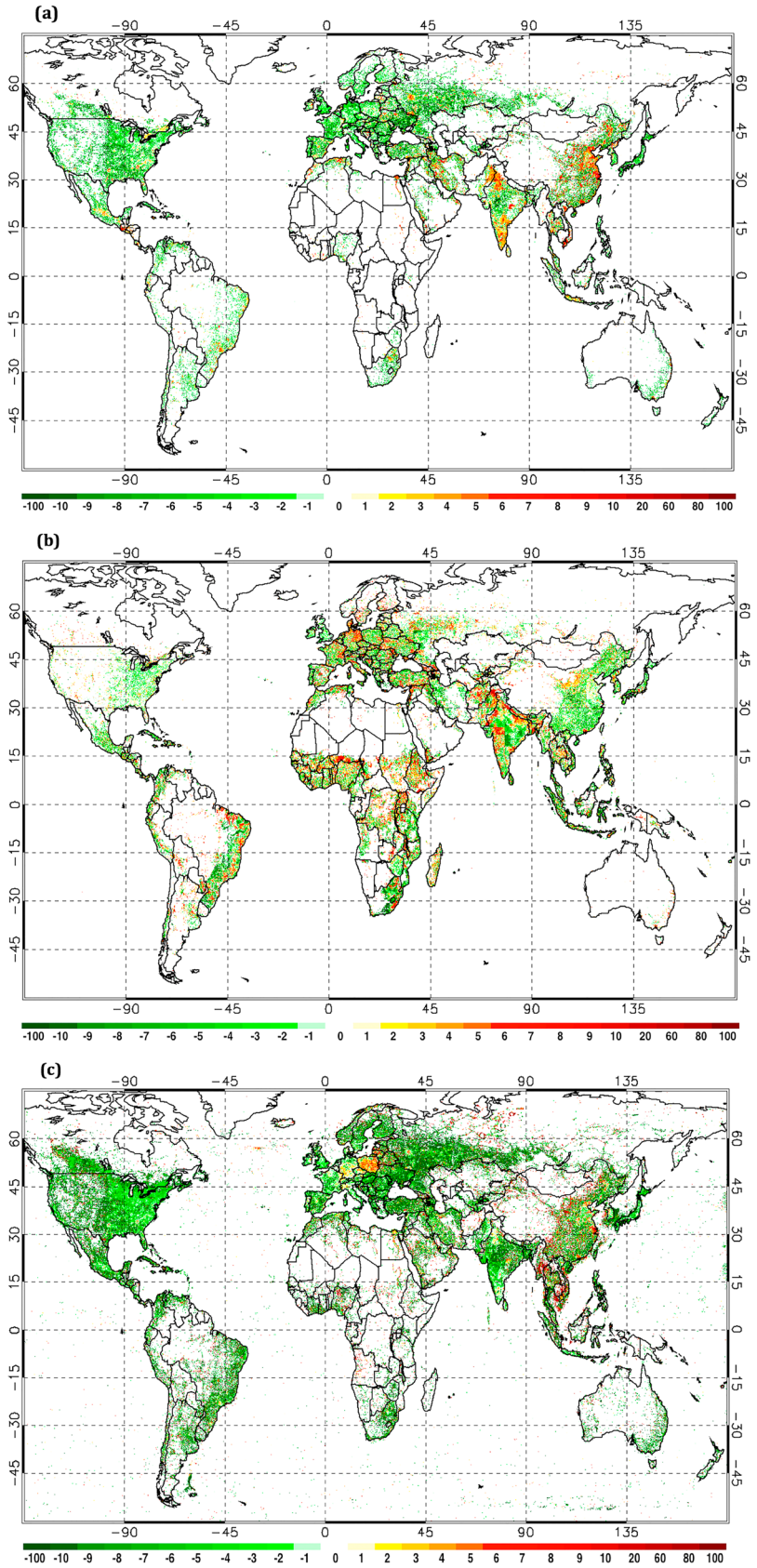

Figure 4

zones. Other notable high-growth areas include northwestern Iran (Tehran), Pakistan (Lahore, Islamabad), and Algeria (Algiers). There is also a distinct latitudinal structure to the trends causing a southward shift of emissions from the northern hemisphere. This is a likely contributor to the changes in the atmospheric CO_2 interhemispheric gradient noted by *Francey et al.* [2013].

The United States offers a particular case in which the nighttime lights, population, and emissions show a negative percentage trend for the eastern half of the country. This suggests that both in-country population migration and economic shifts are directly linked and covary spatially. In contrast, China exhibits increases in population in the Beijing to Shanghai corridor and in the rapidly growing southern coastal cities such as Guangzhou. However, CO_2 emissions show a decline in the region of Beijing but increases in many of the other industrial centers, suggesting that as Beijing becomes the focus of administrative activities, carbon-intensive industrial activity is centered elsewhere. India shows increases in population in and around most of the large urban centers of Mumbai, Dehli, Chennai, Bangalore, and Calcutta, but emissions rise in all but Mumbai, which show declines across the west central portion of the country. In this case, the trends in nighttime lights and population are quite different.

3.4. Short-Term Variations

In addition to the examination of long-term trend behavior at the grid cell level, short-term variations can also be examined. Figure 5 shows variations about a local trend. We calculate this as follows:

1. Remove the long-term trend,
2. Calculate the standard deviation of the residuals, and
3. Normalize the residuals by the standard deviation.

We examine the years 2006 and 2010, years for which noninterpolated nightlight observations are available and that bracket the global financial crisis (GFC). In the year 2006, negative emission anomalies are present in much of the northern half of the eastern U.S. and throughout northern Europe, India, and eastern China (Figure 5a). This indicates values that are greater than the long-term mean trend. Positive anomalies, by contrast, appear in the southern half of the eastern U.S. and parts of southern Europe. In 2010, however, the spatial pattern of these anomalies has reversed, most notably in the eastern U.S. with positive anomalies now in the north and negative anomalies in the south. Similarly, positive anomalies now appear throughout much of Europe, India, and China (Figure 5b). The resulting patterns of emission anomalies closely follow the patterns exhibited by nighttime light anomalies (see supporting information for more information on the short-term trend anomalies of nightlights). The global financial crisis (GFC) occurred during the period spanning 2006 to 2010 [*Peters et al.*, 2012]. The observed subnational spatial patterns of negative and positive emission anomalies may therefore be a reflection of the subnational impacts of the GFC. For example, the observed 2010 positive anomalies in northern U.S., Europe, India, and China may reflect a faster recovery or a lesser impact from the GFC after the peak of the crisis in 2009. In contrast, the 2010 negative anomalies in areas such as the southern half of the U.S. may show the areas that experienced slower recovery or were generally more affected.

3.5. Center of Mass

A useful and compact metric to understand and decompose the space/time signature of globally gridded emissions is via a center of mass (CM) calculation [*Gregg et al.*, 2009]. Figure 6 presents the CM for each year of the global FFDAS emission time series. The geographic location of the CM is dependent upon the global frame used, and we define the globe as having a western edge at the date line moving east toward the U.S., Europe, and Asia. With this definition in place, the emission CM is initially located in the south central Mediterranean Sea and progressively moves toward the east and to the south slightly, ending up in Jordan near the Syrian border in the last year of the time series, 2010. The general placement of the CM in the Mediterranean Sea is reflective of the three large foci of emissions: U.S., Europe, and East Asia. The migration of the CM over the time period 1997 to 2010 is consistent with the relative increase in fossil fuel CO_2 emissions in East and South Asia.

Figure 4. Global map of (a) FFDAS v2 total fossil fuel CO_2 emissions, (b) population, and (c) nightlights expressed as mean linear percentage trends over the 14 year time period. In order to reduce noise in fossil fuel CO_2 emission percentage trends, all the points for which the absolute value of the trend is smaller than its standard deviation are removed and a minimum value of 0.01 kg C m^{-2} yr^{-1} has been applied to the fossil fuel CO_2 emissions, a value of 10^{-7} people m^{-2} yr^{-1} has been applied to population, and a value of 1 DN m^{-2} yr^{-1} has been applied to the nightlights. Units: %/yr^{-1}.

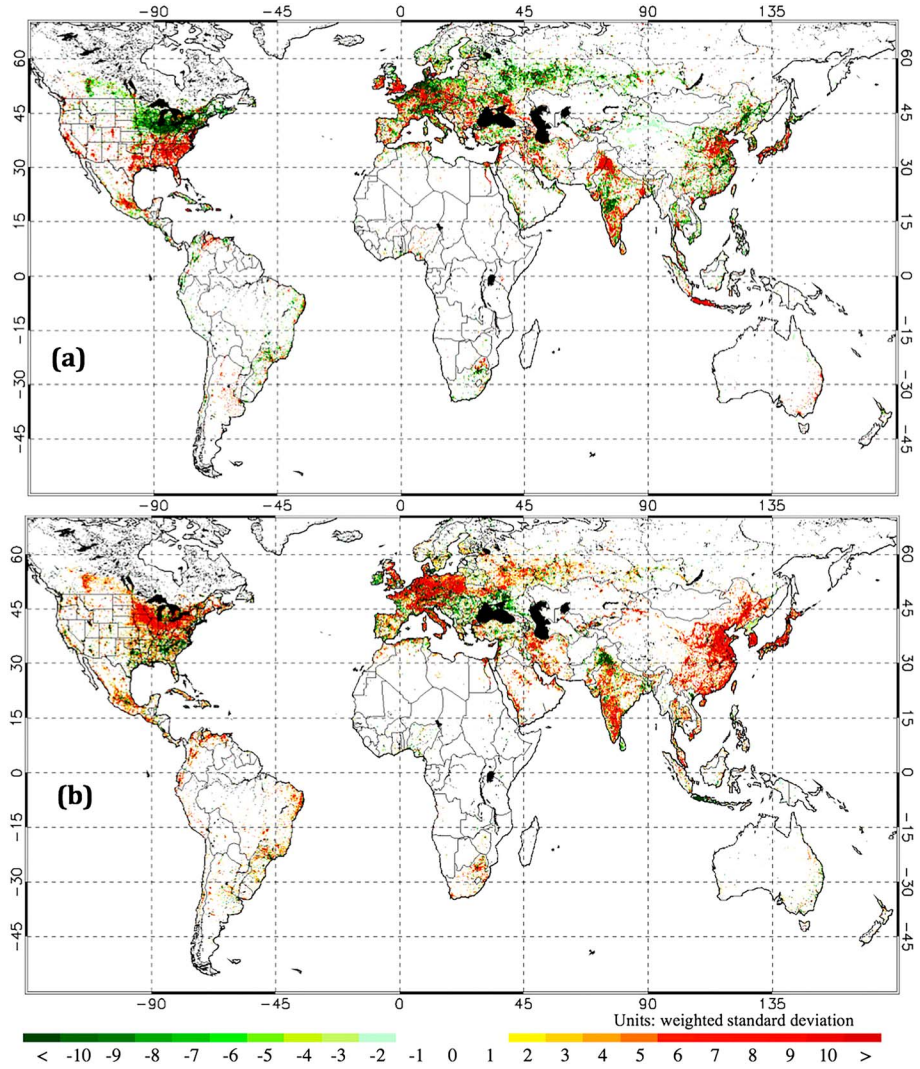

Figure 5. Annual FFDAS v2 fossil fuel CO_2 anomalies for 2 years before and after the Global Financial Crisis. Years (a) 2006 and (b) 2010. Units: weighted standard deviation (see main text for the details on units).

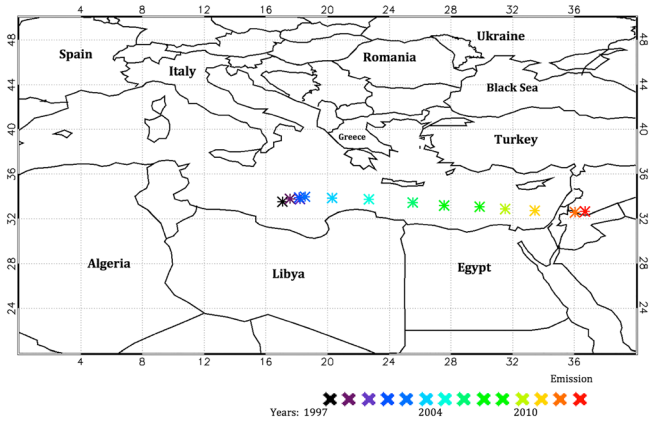

Figure 6. Annual FFDAS v2 total fossil fuel CO_2 emissions global center of mass (CM) over the time period 1997 to 2010.

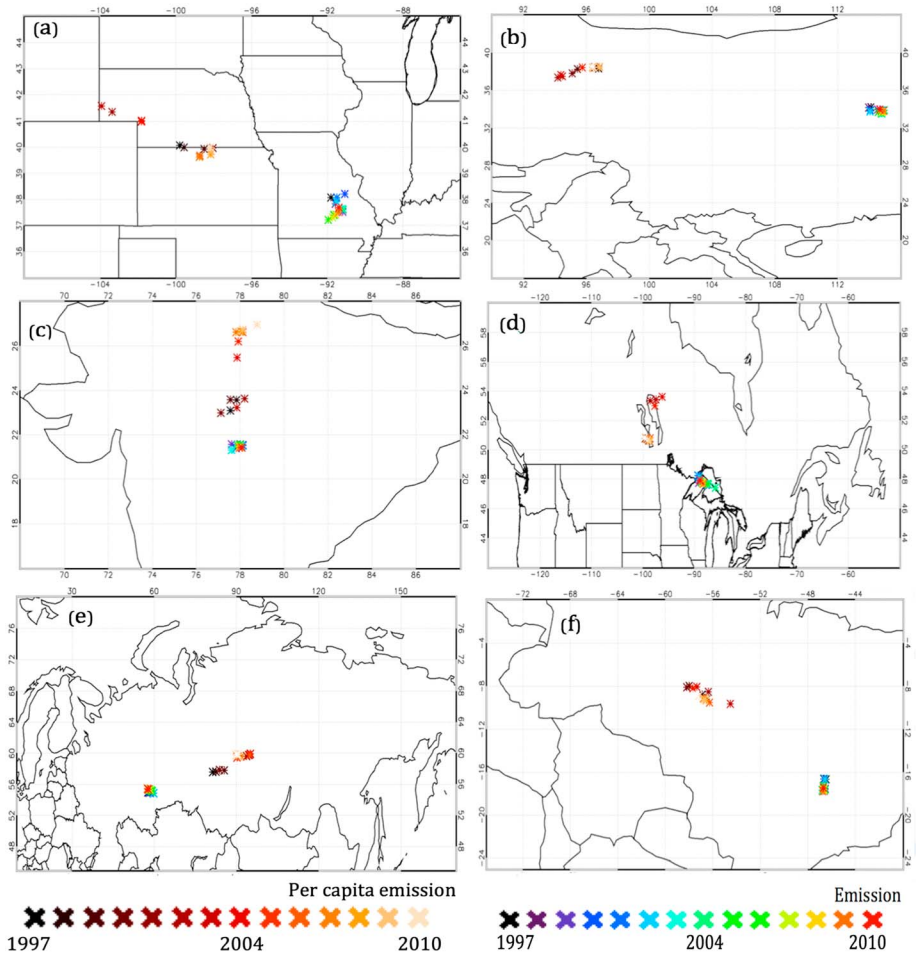

Figure 7. The center of mass (CM) associated with the annual FFDAS v2 other sector fossil fuel CO_2 emissions (blue cross) and per capita fossil fuel CO_2 emissions (red cross) for six countries over the time period 1997 to 2010. (a) U.S., (b) China, (c) India, (d) Canada, (e) Russia, and (f) Brazil.

Further insight can be gained by calculating the per capita CM. In this case, we utilize only the other sector as a per capita normalization. The electricity production sector holds little geographic meaning as electricity production facilities are not typically associated with population. Figure 7 shows six countries in which we calculate the absolute emissions, population, and per capita CM for each year utilizing the other sector emissions.

In some cases, the per capita emission CM exhibits a quasi-linear migration in space over the time period 1997 to 2010. The per capita emission CM will change position driven by either an emission or population effect (or a combination). For example, the population in a country could remain fixed in space and time but the spatial distribution of emission change. The per capita emissions CM, in this example, would migrate toward those areas for which emissions rose and/or away from those regions where emissions fell. The population effect could be described by imagining a country in which there was no change in the spatial pattern of emissions, but the spatial distribution of population changed. In this case, the per capita CM would migrate away from a population increase and toward the population decline, all else being equal. Both population and emissions can alter their space and time distributions, of course, and hence, CM migration can incorporate both effects. The influence of these factors can be assessed by examining the movement of both other sector emission CM and population CM.

In the case of Russia, the emission CM moves in the northeasterly direction, while the population CM moves in the southwesterly direction. Both reinforce a per capita emission CM that moves in the northeasterly direction, away from the large population centers in the eastern half of the country. This could indicate a decline in the population in the eastern population centers, for which the associated emissions decline is

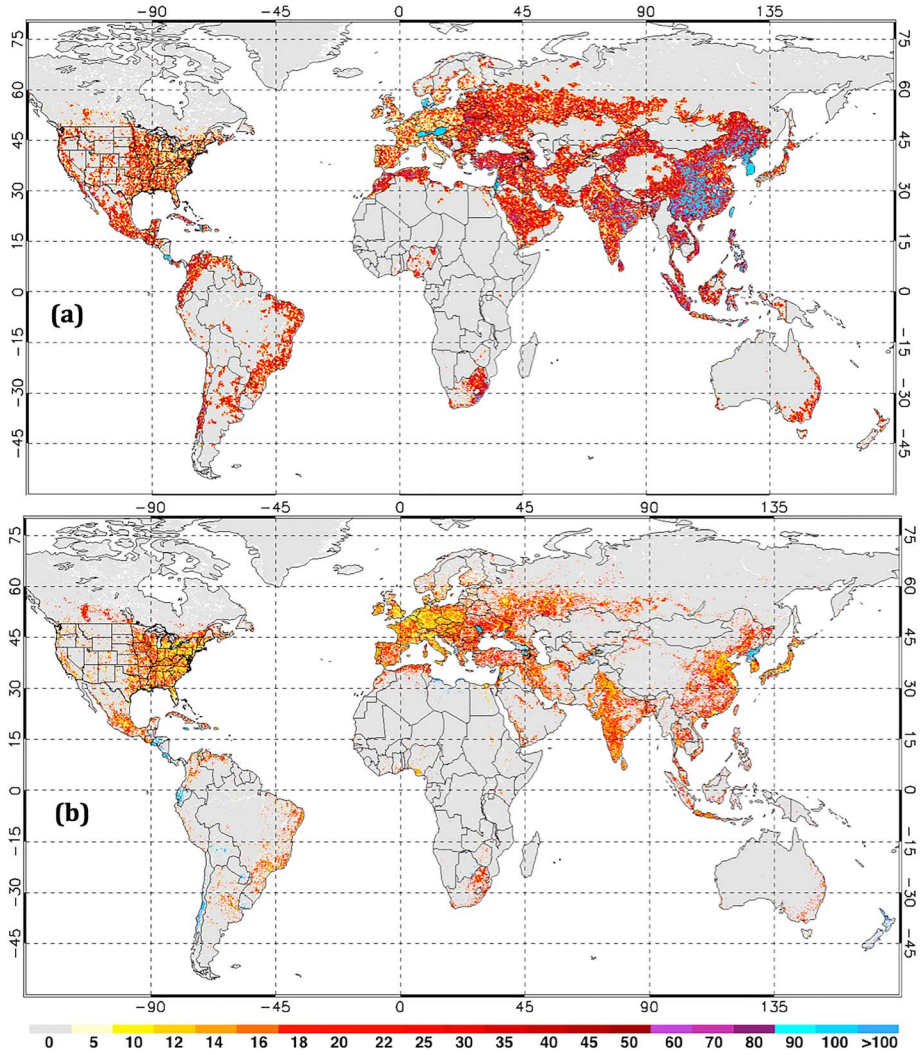

Figure 8. Posterior uncertainties associated with the total fossil fuel CO_2 emissions (expressed as percentage standard deviation). (a) FFDAS v1 at 0.25° for 2002. (b) FFDAS v2 at 0.1° for 2003. Year 2003 was used for FFDAS v2 results due to the fact that 2002 was an interpolated nighttime lights year.

proportionately more rapid causing the per capita emissions to decline in this region and driving the per capita emission CM in the NE direction.

India also exhibits a dramatic, nearly linear, migration of the per capita emission CM. The per capita emission CM moves in a decidedly NNE direction. Both the population and emission CMs move in a NNE direction. The per capita emission CM movement could be due to a rise in both the population and emissions of Calcutta and New Delhi but with emissions rising proportionately quicker. This would cause a rise in per capita emissions in those cities and the migration of the per capita emission CM toward the north.

The remaining four countries show per capita emission CM movement that is more challenging to interpret given the available metrics. Analysis of this metric is left for future research perhaps in conjunction with a thorough analysis of other subnational socioeconomic spatiotemporal patterns.

3.6. Posterior Uncertainties

Figure 8 shows the posterior uncertainties (represented as percentage standard deviations) for FFDAS v1 at 0.25° resolution compared to the FFDAS v2 posterior uncertainties at 0.1° resolution for the year 2002. The average of the grid cell-level FFDAS v2 uncertainties for the other sector are roughly 30%, while the values for FFDAS v1 are greater at approximately 50%. Year 2002 uses interpolated nightlight input data and therefore

has nightlight observational uncertainties increased by 25% (see section 2.2.4). Therefore, the posterior uncertainties on the emissions are also larger by 25% in 2002. The FFDAS v2 other sector uncertainties arise from the combined uncertainty associated with the national emissions and nightlights. For example, the large posterior uncertainties in North Korea are a reflection of larger national emissions and nightlight uncertainties. North Korea is one of the countries with small nightlight values in populated areas, and therefore, large observational uncertainties have been applied to this country.

3.7. Future Work

FFDAS v2 is now a fully configured data assimilation system which ingests a series of observed quantities to act as constraints to the generation of high-resolution fossil fuel CO_2 emissions. However, the current list of observables and sectors does not exploit the full power of this approach. The current sectors are completely separable in the sense that no individual observable is influenced by state variables from other sectors. Cross influence will occur as additional observables are introduced to the system such as satellite-based column CO_2 [*Kort et al.*, 2012], atmospheric $^{14}CO_2$ radiocarbon [*Turnbull et al.*, 2009], or NO_2 [*Berezin et al.*, 2013].

We also intend to subdivide the fossil fuel CO_2 emissions into additional sectors such as onroad transportation, for which we have detailed global road atlases. This will provide a significant shift in the spatial pattern associated with emissions, particularly in those countries for which onroad vehicle travel is a sizable portion of their total emissions. Similarly, we will test the ability of other remotely sensed data products to act as activity observables for individual economic sectors. For example, we will test the use of impervious surface for separating residential/commercial from other sectors. Finally, we will introduce economic data directly into the FFDAS system. This may not improve the gridded emission estimates in the short-term but will allow analysis of the carbon intensity and energy efficiency of national and subnational economies.

4. Conclusions

High-resolution, globally comprehensive quantification of fossil fuel CO_2 emissions has emerged as a critical need in carbon science and climate policy. We have significantly advanced the fossil fuel data assimilation system (FFDAS), first described by *Rayner et al.* [2010], by improving the underlying observationally based data sources, extending the system to a sector-based treatment and generating a complete time series of emissions from 1997 to 2010. The FFDAS v2 comparison to the high-resolution, bottom-up fossil fuel CO_2 emissions data product for the U.S. (Vulcan) shows improved performance over the FFDAS v1 and demonstrates the advantages of approaching this problem by isolating and assembling emitting sectors separately.

Long-term trends show considerable subnational spatial structure in large active economies such as the United States, China, and India. These three countries, in particular, show different long-term trends. Exploration of the trends in nighttime lights and population reveals decoupling of population and emissions at the subnational level. Analysis of shorter-term variations also reveals subnational spatial structure. Of note is the impact of the global financial crisis in 2008–2009 with widespread negative anomalies in emissions across the U.S., Europe, and Japan. The year 2010 shows a recovery of emissions, with the U.S. recovery marked by a north-south division in the eastern half of the country.

Finally, we compute the center of mass (CM), a useful compact metric for expressing the evolution of spatial patterns in fossil fuel CO_2 emissions. The global emission CM was located in the southern Mediterranean Sea in 1997 but had progressed to the Middle East by 2010. The per capita emission CM holds potential as a way to analyze subnational shifts in carbon intensity over time. Some countries analyzed here show provocative CM migration suggesting the importance of the urban/rural carbon intensity gradient, but further research is needed to fully understand these dynamics.

Acknowledgments

This work was made possible by NASA ROSES grant NNX11AH86G and NASA CMS grant NNX12AP52G. Rayner is in receipt of an Australian Professorial Fellowship (DP1096309). We also acknowledge the helpful input from Tomohiro Oda.

References

Andres, R. J., G. Marland, I. Fung, and E. Matthews (1996), A one degree by one degree distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950–1990, *Global Biogeochem. Cycle*, *10*, 419–429, doi:10.1029/96GB01523.

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

Andres, R. J., T. A. Boden, and G. Marland (2012), *Monthly Fossil-Fuel CO_2 Emissions: Mass of Emissions Gridded by One Degree Latitude by One Degree Longitude*, Carbon Dioxide Inf. Anal. Cent. (CDIAC), Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn., doi:10.3334/CDIAC/ffe.MonthlyMass.2012.

Berezin, E. V., I. B. Konovalov, P. Ciais, A. Richter, S. Tao, G. Janssens-Maenhout, M. Beekmann, and E.-D. Schulze (2013), Multiannual changes of CO_2 emissions in China: Indirect estimates derived from satellite measurements of tropospheric NO_2 columns, *Atmos. Chem. Phys.*, *13*, 255–309.

Bhaduri, B., E. Bright, P. Coleman, and M. Urban (2007), LandScan USA: A high resolution geospatial and temporal modeling approach for population distribution and dynamics, *GeoJournal*, *69*, 103–117.

Boden, T. A., G. Marland and R. J. Andres (1995), Estimates of global, regional, and national annual CO_2 emissions from fossil-fuel burning, hydraulic cement production, and gas flaring: 1950–1992, ORNL/CDIAC-90, NDP-30/R6, Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn.

Bovensmann, H., M. Buchwitz, J. P. Burrows, M. Reuter, T. Krings, K. Gerilowski, O. Schneising, J. Heymann, A. Tretner, and J. Erzinger (2010), A remote sensing technique for global monitoring of power plant CO_2 emissions from space and related applications, *Atmos. Meas. Tech.*, *3*, 781–811, doi:10.5194/amt-3-781-2010.

Carbon Dioxide Information Analysis Center (2010), Fossil-fuel CO_2 emissions. [Available at http://cdiac.ornl.gov/trends/emis/meth_reg.html.]

Chevallier F., F. M. Bron, and P. J. Rayner (2007), The contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, *J. Geophys. Res.*, *112*, D09307, doi:10.1029/2006JD007375.

Crisp, D., et al. (2012), The ACOS CO_2 retrieval algorithm – Part II: Global XCO_2 data characterization, *Atmos. Meas. Tech.*, *5*, 687–707, doi:10.5194/amt-5-687-2012.

Elvidge, C. D., K. E. Baugh, J. Dietz, T. Bland, P. C. Sutton, and H. W. Kroehl (1999), Radiance calibration of DMSP-OLS low-light imaging data of human settlements, *Remote Sens. Environ.*, *68*(1), 77–88.

Elvidge, C. D., D. Ziskin, K. E. Baugh, B. T. Tuttle, T. Ghosh, D. W. Pack, E. H. Erwin, and M. Zhizhin (2009), A fifteen year record of global natural gas flaring derived from satellite data, *Energies*, *2*(3), 595–622.

Enting, I. (2002), *Inverse Problems in Atmospheric Constituent Transport*, Cambridge Univ. Press, New York.

Francey, R. J., et al. (2013), Atmospheric verification of anthropogenic CO_2 emission trends, *Nat. Clim. Change*, *3*, 520–524, doi:10.1038/nclimate1817.

Friedlingstein, P., and I. C. Prentice (2010), Carbon-climate feedbacks: A review of model and observation based estimates, *Curr. Opin. Environ. Sustainability*, *2*, 251–257.

Ghosh, T., C. D. Elvidge, P. C. Sutton, K. E. Baugh, D. Ziskin, and B. T. Tuttle (2010), Creating a global grid of distributed fossil fuel CO_2 emissions from nighttime satellite imagery, *Energies*, *3*, 1895–1913.

Gregg, J. S., L. M. Losey, R. J. Andres, T. J. Blasing, and G. Marland (2009), The temporal and spatial distribution of carbon dioxide emissions from fossil-fuel use in North America, *J. Appl. Meteorol. Climatol.*, *48*, 2528–2542, doi:10.1175/2009JAMC2115.1.

Gurney, K. R. (2013), Beyond hammers and nails: Mitigating and verifying greenhouse gas emissions, *Eos Trans. AGU*, *94*(22), 199, doi:10.1002/2013EO220003.

Gurney, K. R., et al. (2002), Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, *Nature*, *415*, 626–630.

Gurney, K. R., Y. H. Chen, T. Maki, S. R. Kawa, A. Andrews, and Z. Zhu (2005), Sensitivity of atmospheric CO_2 inversion to seasonal and interannual variations in fossil fuel emissions, *J. Geophys. Res.*, *110*, D10308, doi:10.1029/2004JD005373.

Gurney, K. R., D. L. Mendoza, Y. Zhou, M. L. Fischcer, C. C. Miller, S. Geethkumar, and S. de la Rue du Can (2009), High resolution fossil fuel combustion CO_2 emission fluxes for the United States, *Environ. Sci. Technol.*, *43*(14), 5535–5541.

Hozo, S. P., B. Djulbegovic, and I. Hozo (2005), Estimating the mean and variance from the median, range, and the size of a sample, *BMC Med. Res. Method.*, *5*, 13, doi:10.1186/1471-2288-5-13.

International Energy Agency (2011), CO_2 emissions from fuel combustion, 2010 ed., OECD/IEA, Paris, 1997. [Available at http://www.iea.org/.]

Intergovernmental Panel on Climate Change (IPCC) (1996), Revised 1996 IPCC Guidelines for National Greenhouse Gas Emission Inventories. Three volumes: Reference manual, Reporting Guidelines and Workbook. IPCC/OECD/IEA. IPCC WG1 Technical Support Unit, Hadley Centre, Meteorological Office, Bracknell, U. K.

Intergovernmental Panel on Climate Change (IPCC) (2006), *IPCC Guidelines for National Greenhouse Gas Inventories. Prepared by the National Greenhouse Gas Inventories Programme (NGGIP)*, edited by H. S. Eggleston et al., IGES, Japan. [Available at www.ipccnggip.iges.or.jp/public/2006gl/index.html.]

Kort, E. A., C. Frankenberg, and C. E. Miller (2012), Space-based observations of megacity carbon dioxide, *Geophys. Res. Lett.*, *39*, L17806, doi:10.1029/2012GL052738.

Le Quéré, C., et al. (2009), Trends in the sources and sinks of carbon dioxide, *Nat. Geosci.*, *2*, 831–836, doi:10.1038/ngeo689.

Liu, M., H. Wang, H. Wang, T. Oda, Y. Zhao, X. Yang, R. Zang, B. Zang, J. Bi, and J. Chen (2013), Refined estimate of China's CO_2 emissions in spatiotemporal distributions, *Atmos. Chem. Phys.*, *13*, 10,873–10,882, doi:10.5194/acp-13-10873-2013.

Macknick J. (2011), Energy and CO_2 emissions data uncertainties, *Carbon Manage.*, *2*(2), 189–205. Internet: Energy and Carbon Emission Harmonization Database. [Available at http://www.iiasa.ac.at/web/home/research/modelsData/EnergyCarbonDatabase.en.html.]

Marland, G. (2008), Uncertainties in accounting for CO_2 from fossil fuels, *J. Ind. Ecol.*, *12*, 136–139.

Marland, G., and R. M. Rotty (1984), Carbon dioxide emissions from fossil fuels: A procedure for estimation and results for 1950–1982, *Tellus B*, *36*, 232–261.

Marland, G., K. Hamal, and M. Jonas (2009), How uncertain are estimates of CO_2 emissions?, *J. Ind. Ecol.*, *13*, 4–7.

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

Nakicenov, N. (2004), *In The Global Carbon Cycle: Integrating Humans, Climate and the Natural World*, *SCOPE Ser.*, vol. 62, edited by B. C. Field and M. R. Raupach, pp. 225–242, Island Press, Washington, D. C.

National Research Council (2010), *Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements*, The National Academies Press, Washington, D. C.

Niederberger, A. A., and M. Kimble (2011), MRV under the UN climate regimes: Paper tiger or catalyst for continual improvement?, *Greenhouse Gas Meas. Manage.*, *1*, 47–54.

Oda, T., and S. Maksyutov (2011), 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.

Olivier, J. G. J., J. A. Van Aardenne, F. 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, *Environ. Sci.*, *2*(2–3), 81–99, doi:10.1080/15693430500400345.

Peters, G. P., G. Marland, C. L. Quéré, T. A. Boden, J. G. Canadell, and M. R. Raupach (2012), Rapid growth in CO_2 emissions after the 2008–2009 global financial crisis, *Nat. Clim. Change*, *2*, 2–4.

Raupach, M. R., G. Marland, P. Ciais, C. LeQuere, J. G. Canadell, G. Klepper, and C. B. Field (2007), Global and regional drivers of accelerating CO_2 emissions, *Proc. Natl. Acad. Sci. U.S.A.*, *104*, 10,288–10,293.

Raupach, M. R., P. J. Rayner, and M. Paget (2010), Regional variations in spatial structure of nightlights, population density and fossil-fuel CO_2 emissions, *Energy Policy*, *38*, 4756–4764, doi:10.1016/j.enpol.2009.08.021.

Rayner, P. J., M. R. Raupach, M. Paget, P. Peylin, and E. Koffi (2010), 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.

Schneising, O., J. Heymann, M. Buchwitz, M. Reuter, H. Bovensmann, and J. P. Burrows (2013), Anthropogenic carbon dioxide source areas observed from space: Assessment of regional enhancements and trends, *Atmos. Chem. Phys.*, *13*, 2445–2454.

Turnbull J. C., P. Rayner, J. Miller, T. Naegler, P. Ciais, and A. Cozic (2009), On the use of $^{14}CO_2$ as a tracer for fossil fuel CO_2: Quantifying uncertainties using an atmospheric transport model, *J. Geophys. Res.*, *114*, D22302, doi:10.1029/2009JD012308.

Wang, R., et al. (2013), High-resolution mapping of combustion processes and implications for CO_2 emissions, *Atmos. Chem. Phys.*, *13*, 5189–5203.

Wu L., M. Bocquet, F. Chevallier, T. Lauvaux, and K. Davis (2013), Hyperparameter estimation for uncertainty quantification in mesoscale carbon dioxide inversions, *Tellus B*, *65*, 20,894, doi:10.3402/tellusb.v65i0.20894

Yokota, T., Y. Yoshida, N. Eguchi, Y. Ota, T. Tanaka, H. Watanabe, and S. Maksyutov (2009), Global concentrations of CO_2 and CH_4 retrieved from GOSAT: First preliminary results, *SOLA*, *5*, 160–163, doi:10.2151/sola.2009-041.

Ziskin, D., K. Baugh, F. C. Hsu, T. Ghosh, and C. Elvidge (2010), Methods used for the 2006 radiance lights, *Proc. Asia-Pacific Adv. Network*, *30*, 131–142.