Uncertainty associated with the CDIAC fossil fuel carbon dioxide emissions time series

Bob Andres

Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory

andresrj@ornl.gov

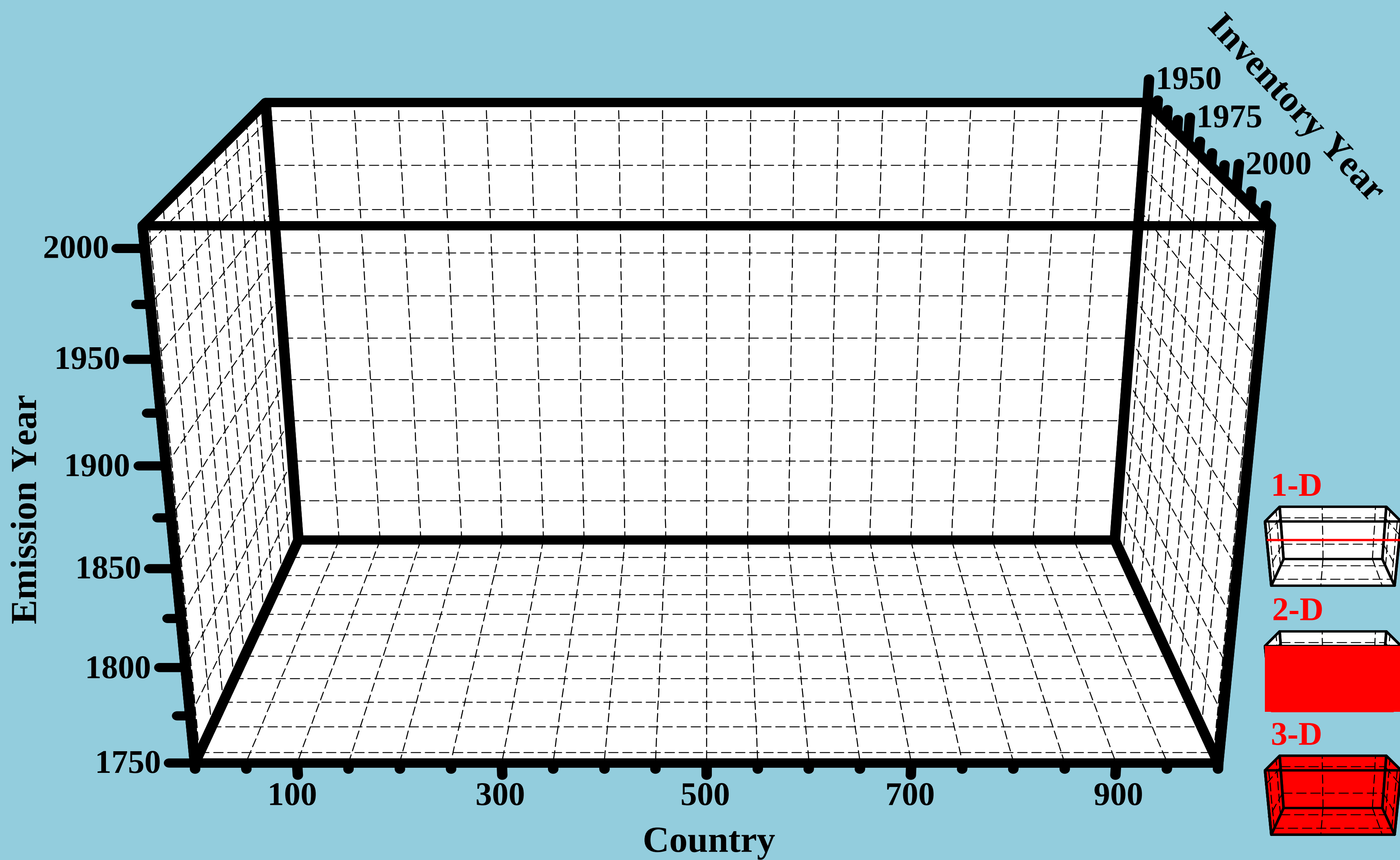
New Global Total

Published last year

Andres RJ, Boden TA, Higdon D (2014) A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission. Tellus B, 66, 23616. doi:10.3402/tellusb.v66.23616.

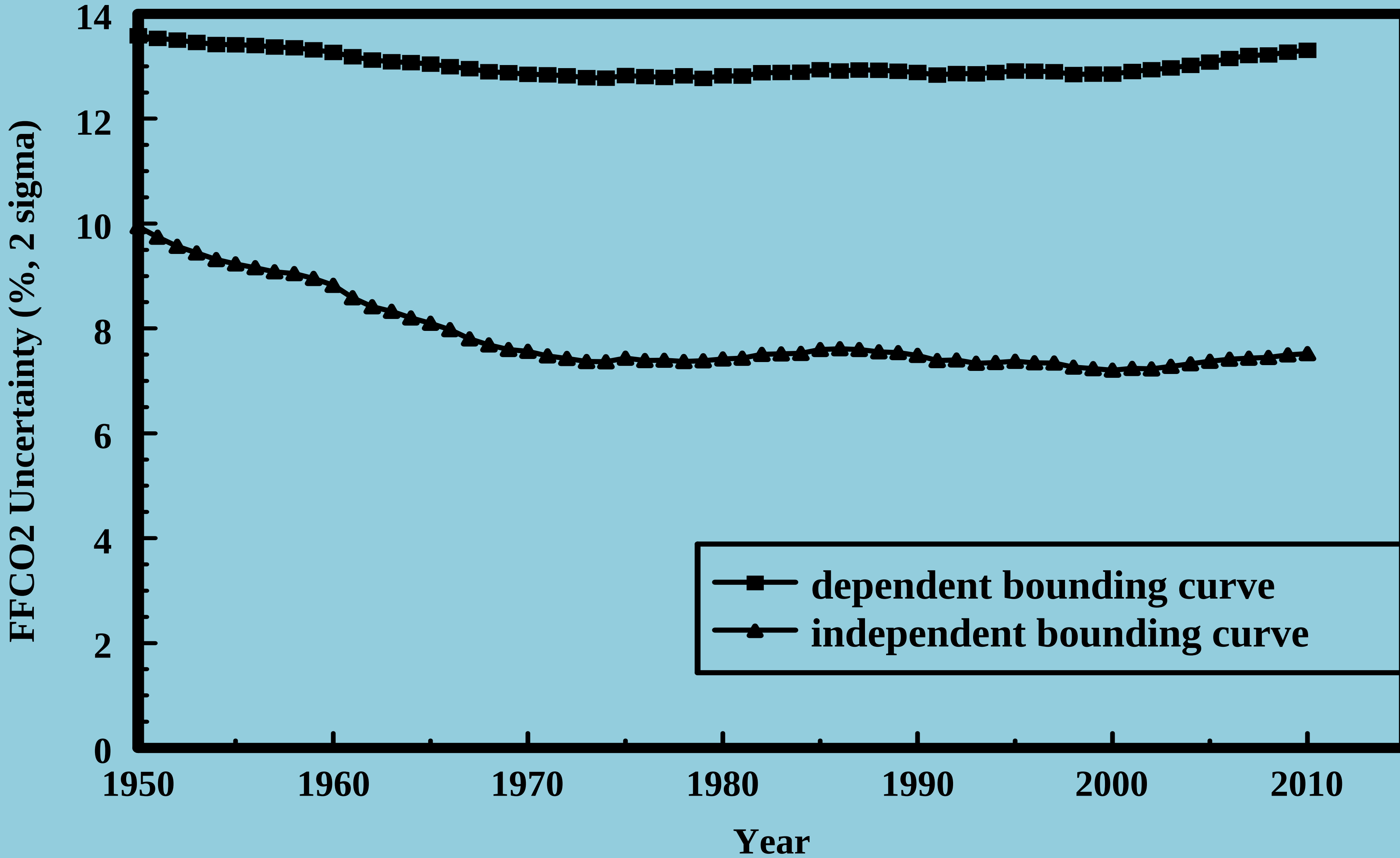
Due to the lack of physical samples at the appropriate temporal and spatial scales, uncertainty was quantified using three different, but complementary assessments

FFCO2 emissions estimate data cube



1-D Case: Fuel-based Assessment

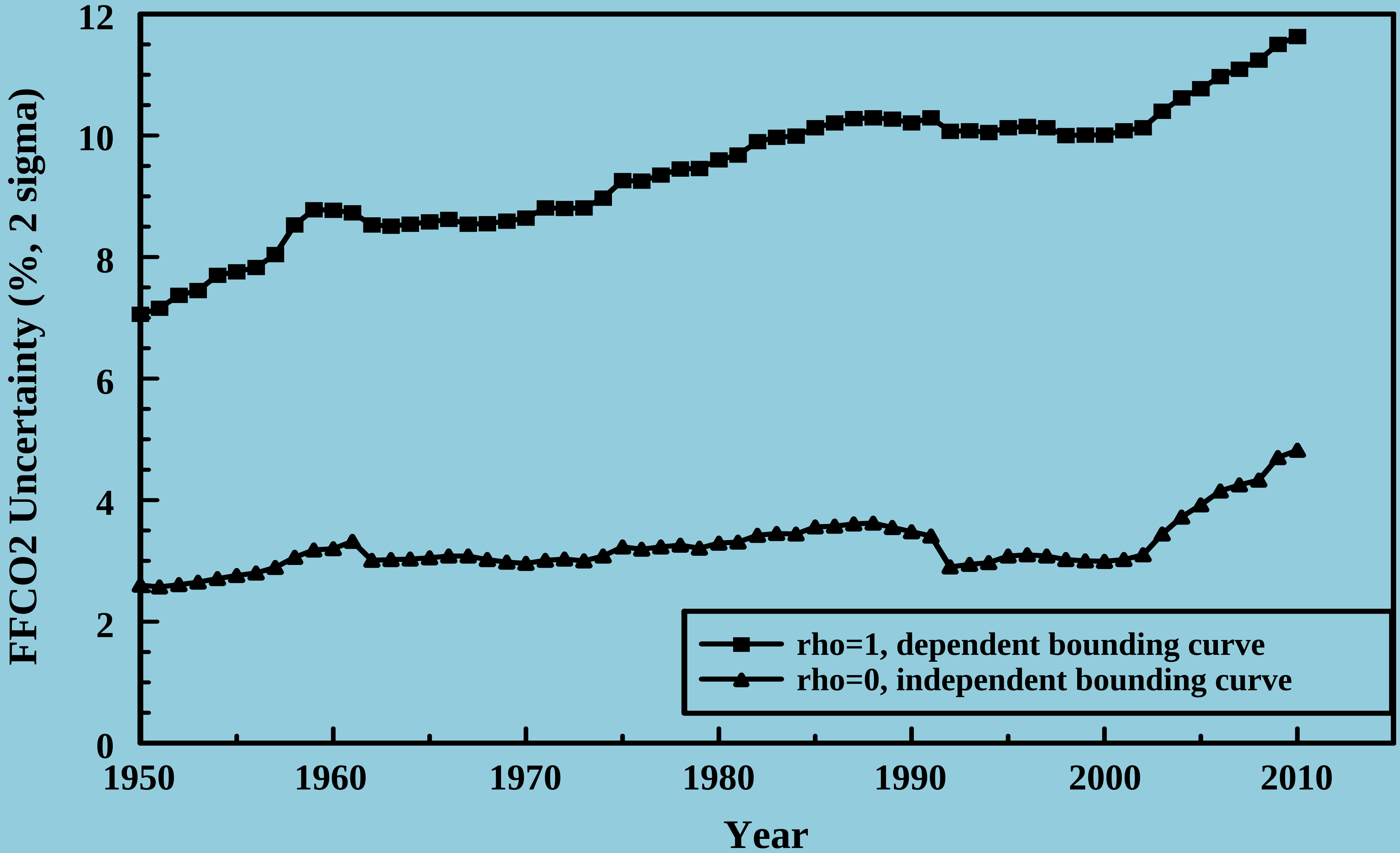
FFCO2 = Fuel Consumed * Fraction Oxidized * Carbon Content



CMS Telecon, 2015, 4/13

2-D Case: Country-based Assessment

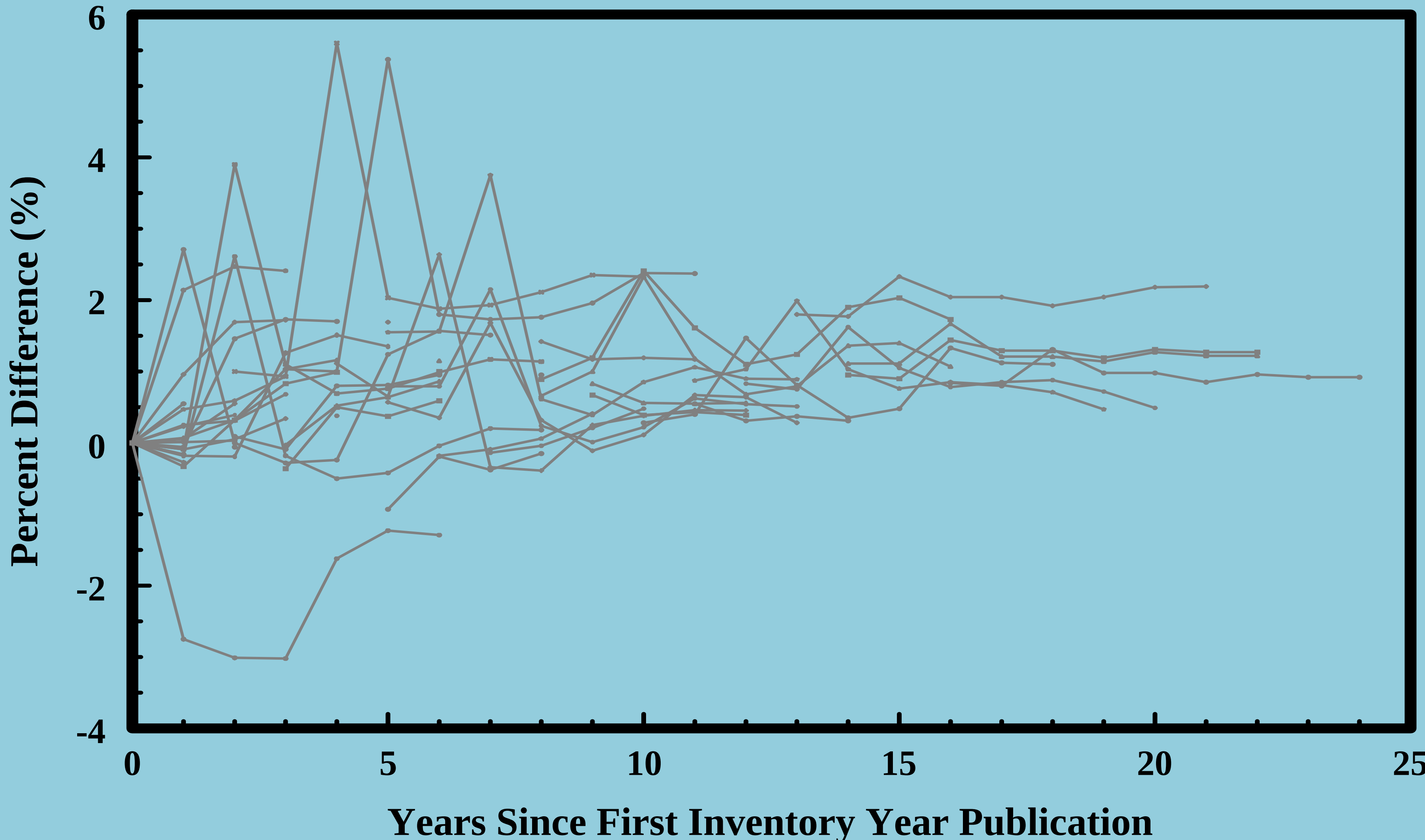
global total = sum country totals



CMS Telecon, 2015, 5/13

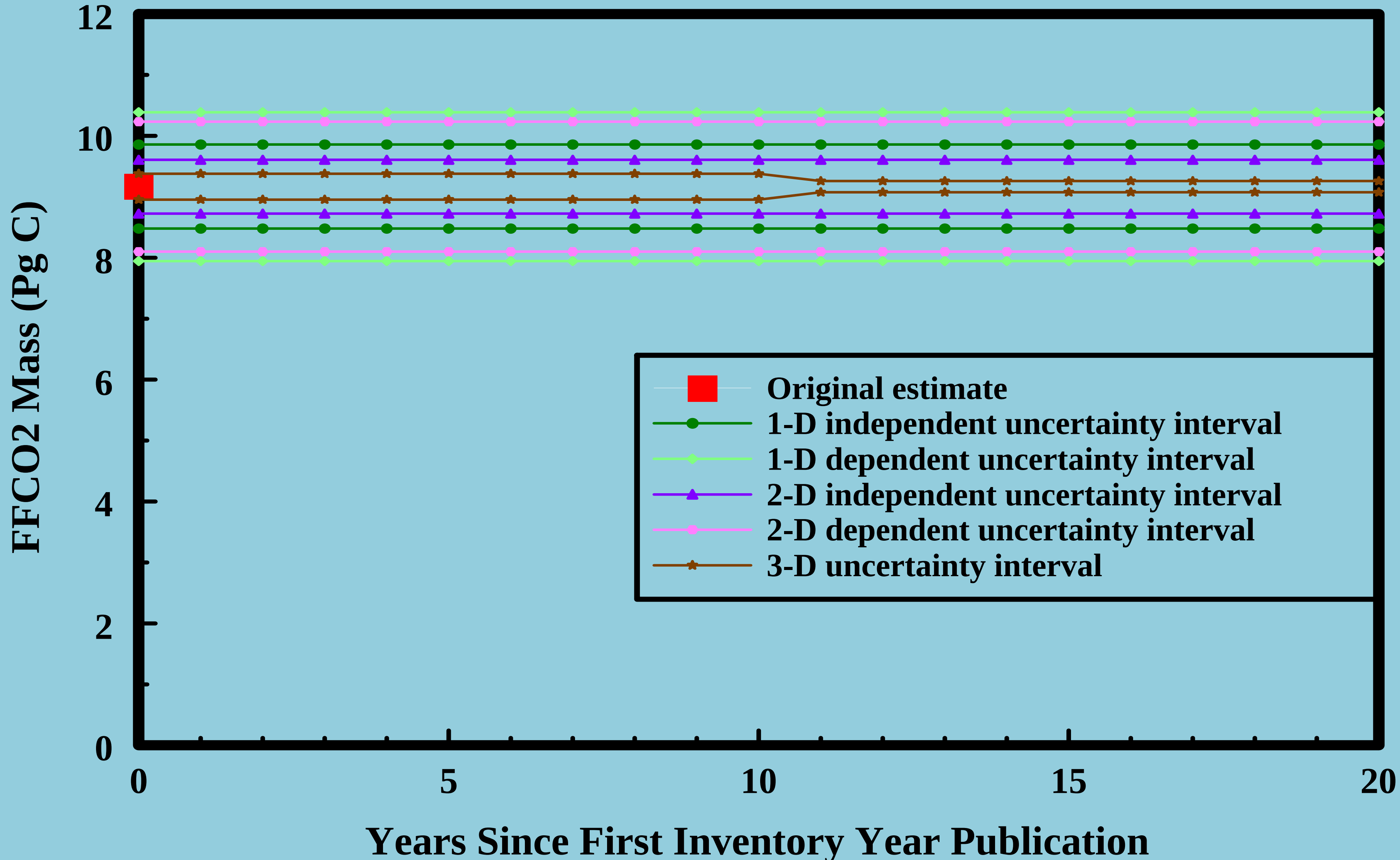
3-D Case: Historical Assessment

data revisions, missing data filled, methodology refined

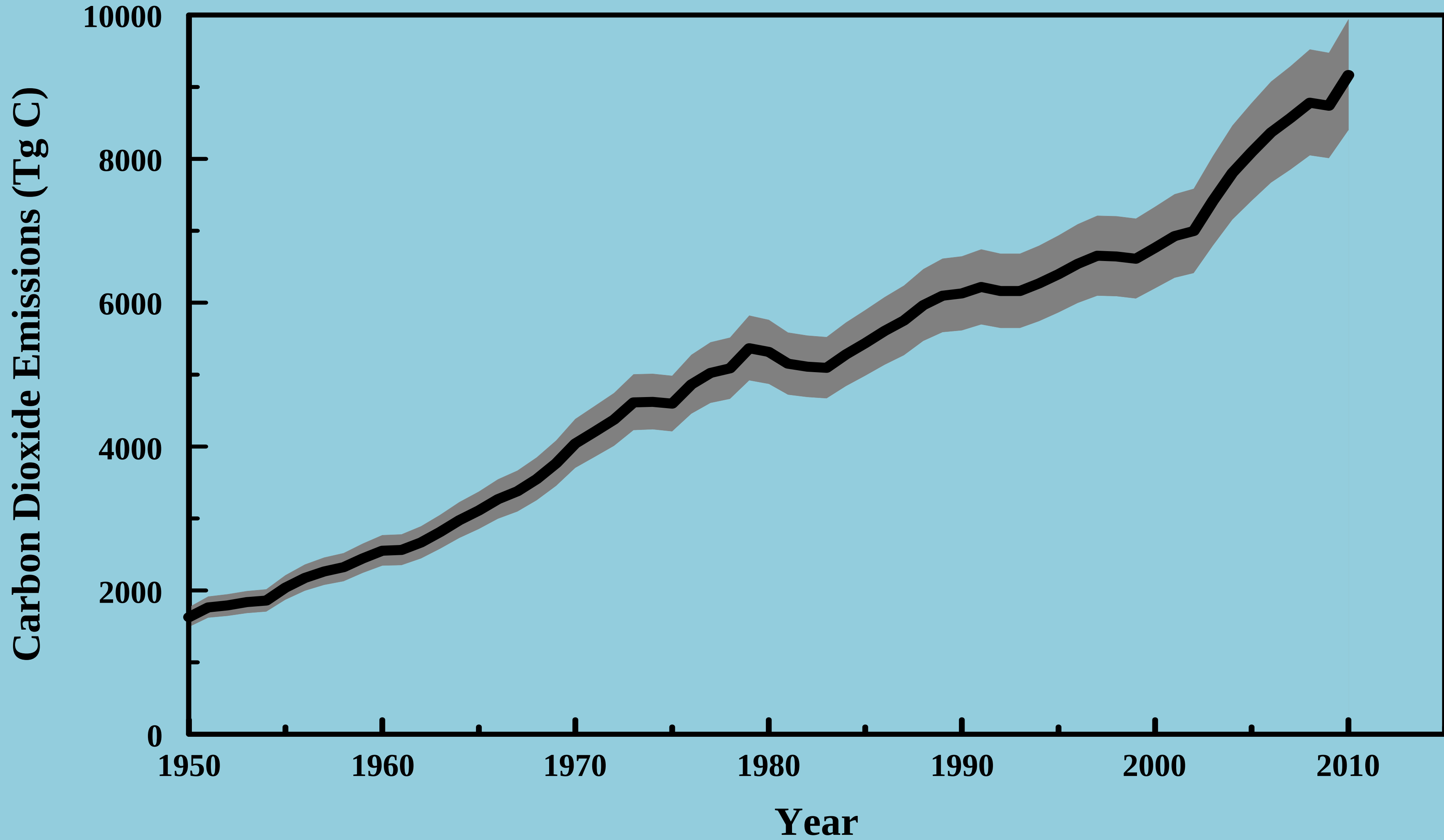


CMS Telecon, 2015, 6/13

3 Assessments Combined: Example

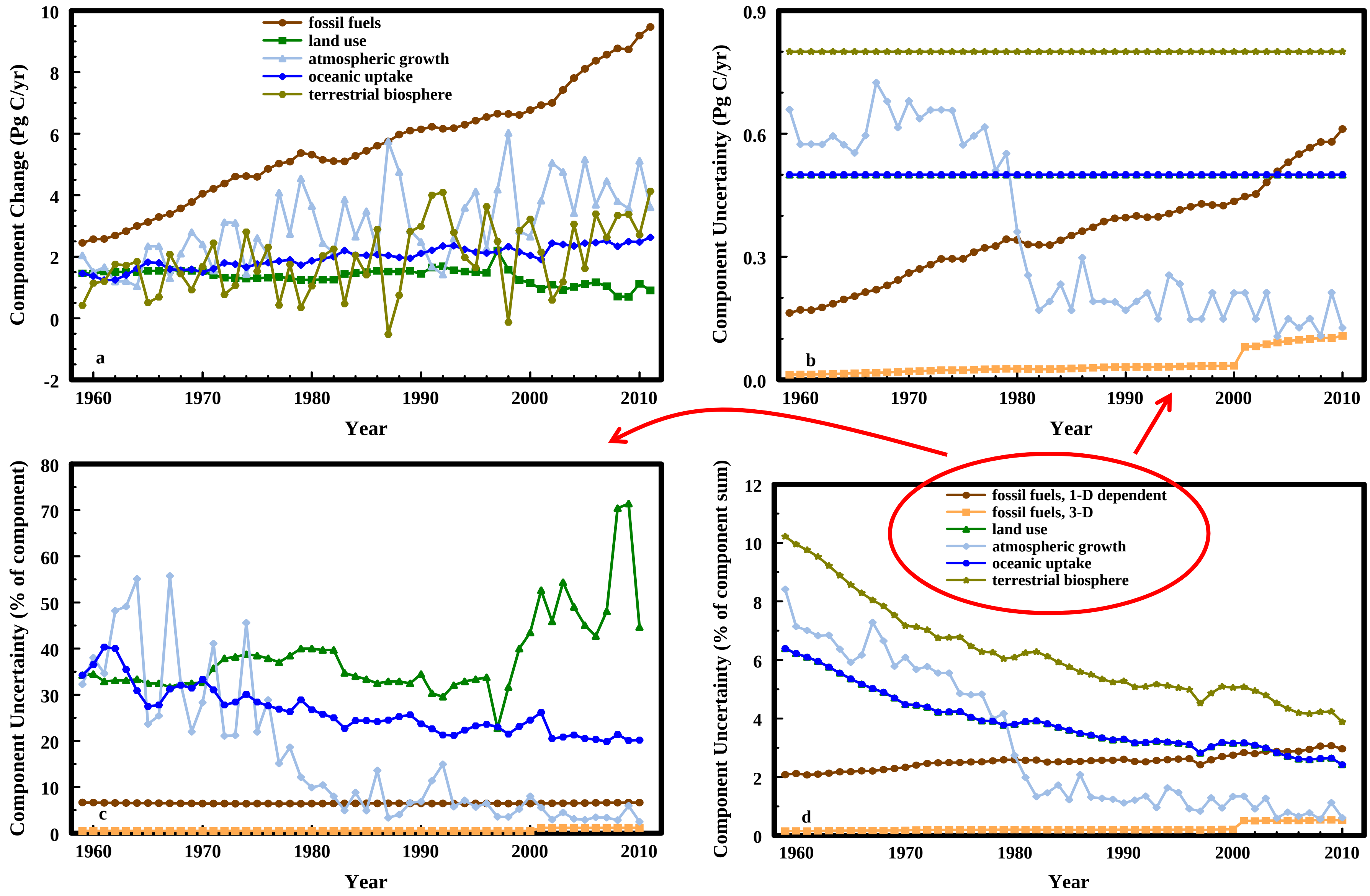


Simplified 8.4% Uncertainty & Time Trend



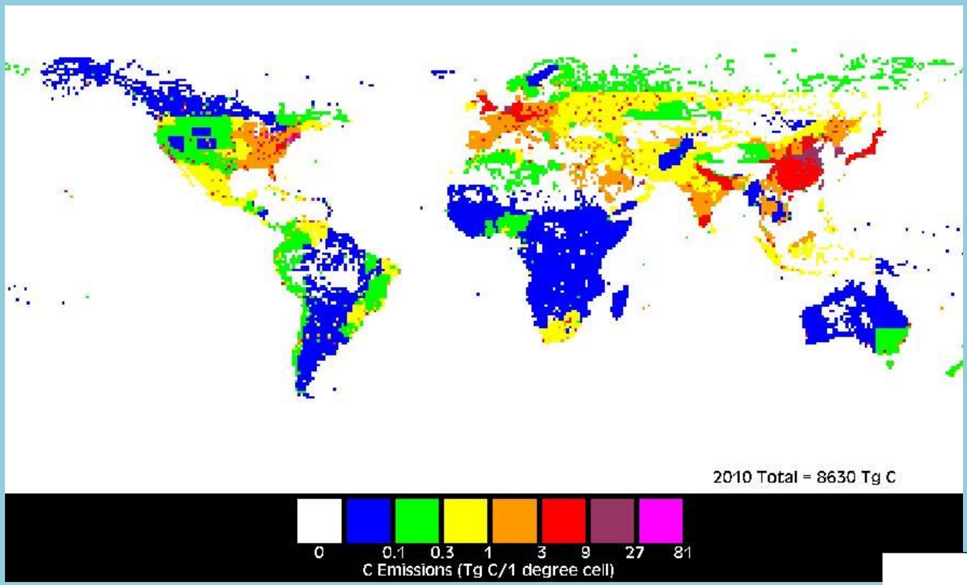
CMS Telecon, 2015, 8/13

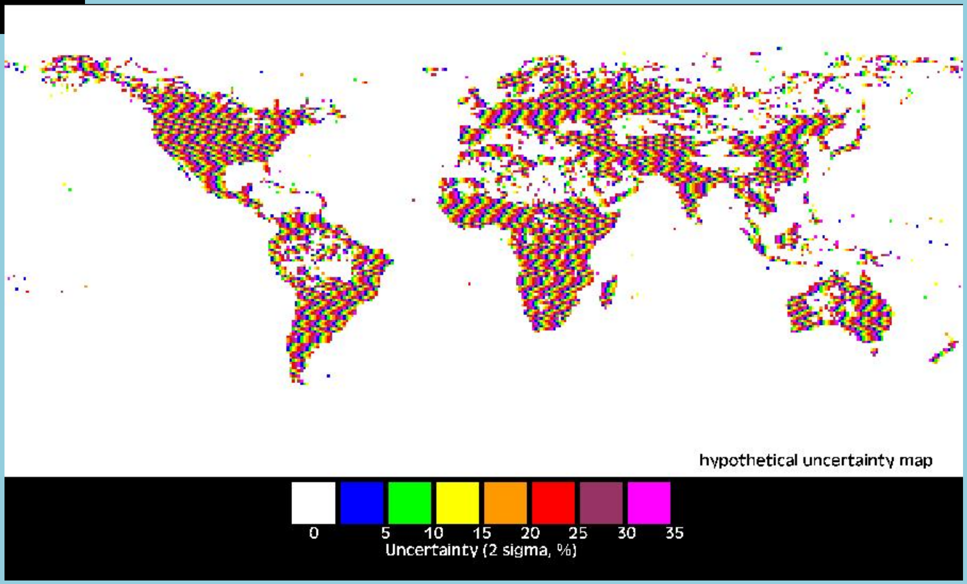
Context



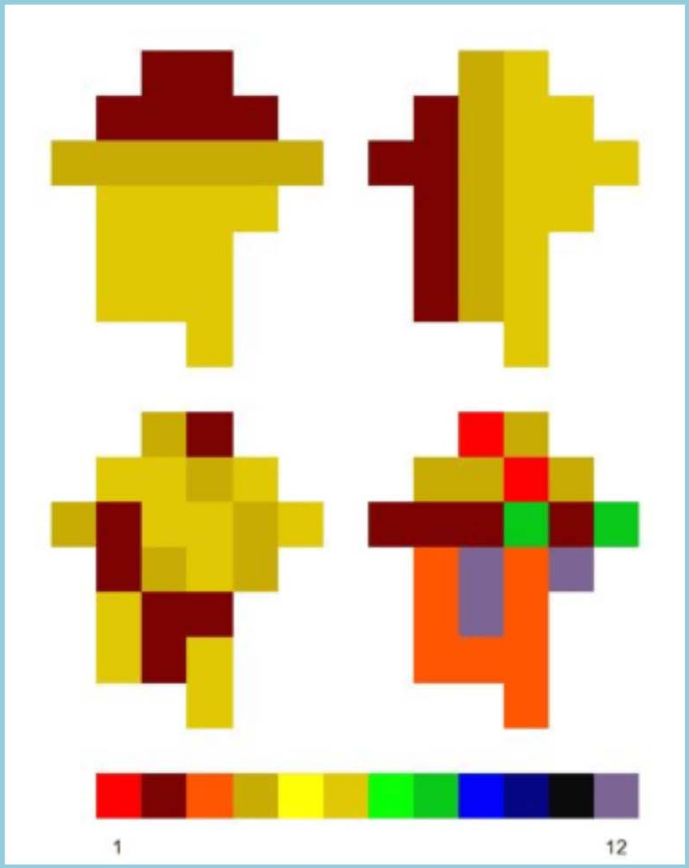
fossil fuels
land use
atmospheric growth
oceanic uptake
terrestrial biosphere
Component Change (Pg C/yr)
Year
a
Component Uncertainty (Pg C/yr)
b
Component Uncertainty (% of component)
c
Component Uncertainty (% of component sum)
d
fossil fuels, 1-D dependent
fossil fuels, 3-D

Uncertainty Paired with Mass Maps

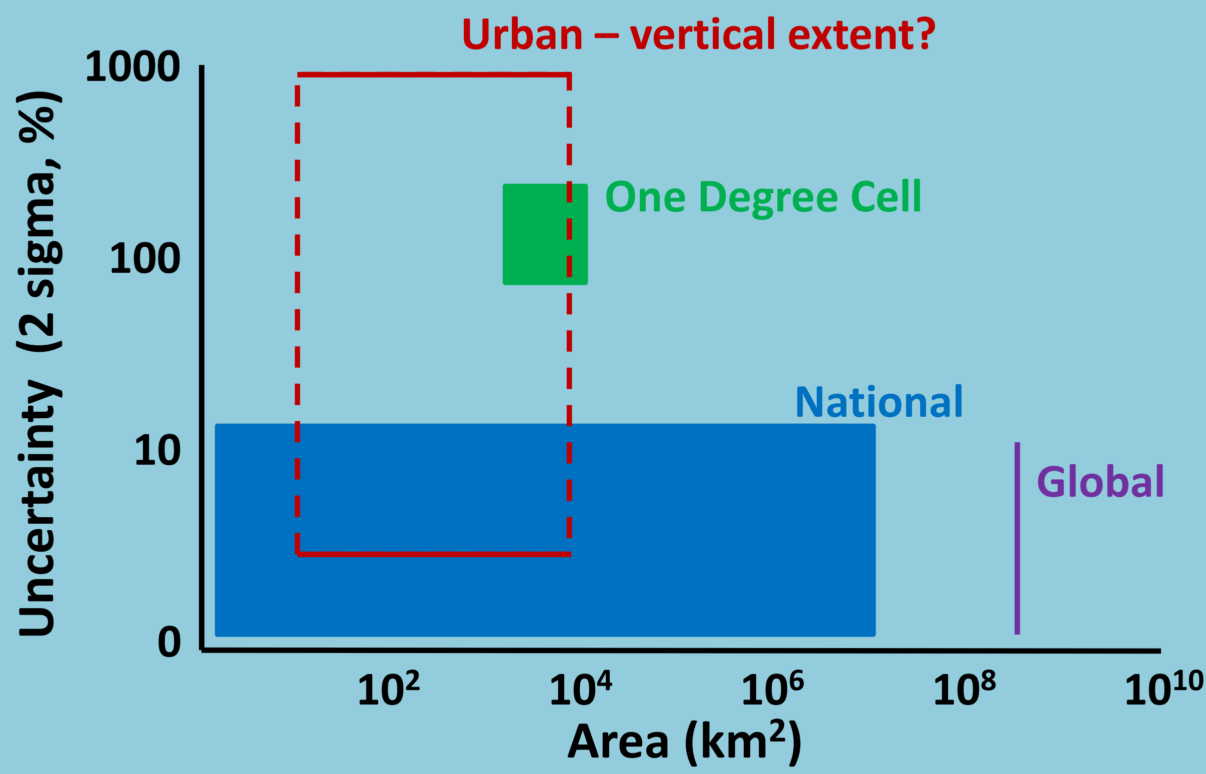




Basic Problem



Scales and Uncertainty



Conclusions/Implications

1. These assessments remind the community that FFCO2 emissions have a non-zero uncertainty associated with them.

2. That this uncertainty is significant, either in isolation or in relation to other components of the global carbon cycle.

FFDAS
Prior and posterior emissions uncertainties

S. Asefi-Najafabady, P. J. Rayner, K. R. Gurney, A. McRobert, Y. Song, K. Coltin, J.Huang, C. Elvidge, K. Baugh

Arizona State University

University of Melbourne

National Geophysical Data Center (NGDC)

National Oceanic and Atmospheric Administration (NOAA)

Prior uncertainties

IEA national emissions:

- IEA national emissions do not include an estimate of uncertainty.
- Using analysis on national/global energy consumption and CO2 emissions [Macknick, 2011] in which five data sets (British Petroleum, CDIAC, U.S. Energy Information Administration, and two variants of the IEA data) were harmonized to remove known differences such as categorical definitions and parameter assumptions.
- We defined the **span of the five estimates** for the 26 countries (top CO2-emitting nations) analyzed by Macknick [2011].
- For countries beyond the 26 analyzed by Macknick [2011], 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.
- On average, the mean percentage span value for the world is ~ 16%. The smallest 1997 to 2010 mean percentage span is 7.5% for Mexico, 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.

Prior uncertainties

Power plant point sources:

We have built upon and improved existing power plant emissions database (CARMA) [Ummel, 2012].

We have improved:

- locations & emissions via national datasets and GE search.
- Re-generated a multivariate regression model that generates improved emissions and **uncertainties** for each individual power plant.

Prior uncertainties

Nightlights:

- A total of six radiance calibrated products were produced by NGDC for 1997, 1999, 2000, 2003, 2004, 2006, and 2010.
- Using this product, we applied a linear interpolation to create a time series of nightlights for 1997 to 2010.
- Nightlight 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

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

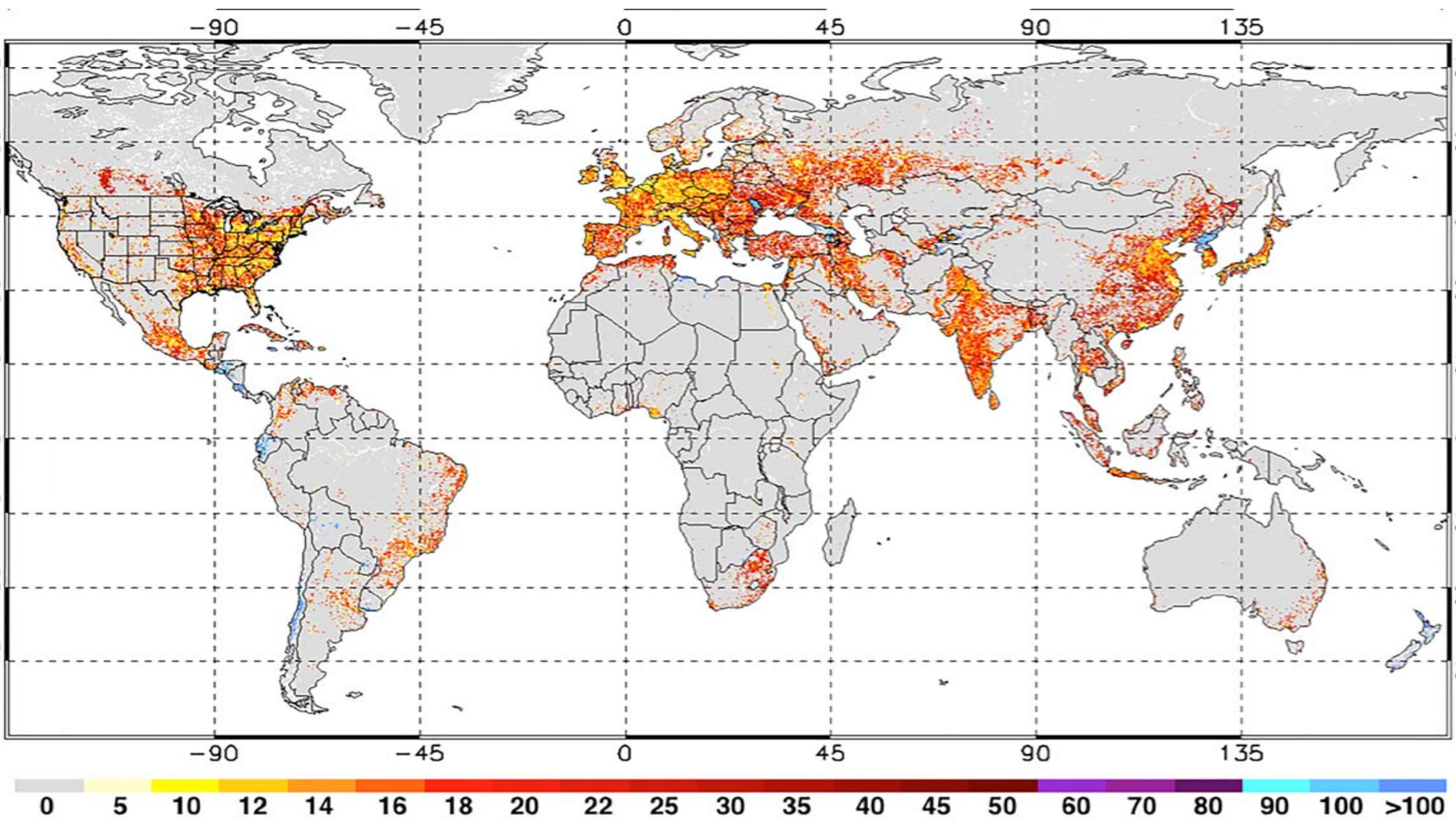
  where x_{NL} is the observed nightlight value.
- We inflate this uncertainty by 25% in years with interpolated nightlights to allow for interpolation error.

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.
- We calculate 10 Monte Carlo realizations of the posterior fluxes at the native 0.1° resolution for years with **observed** and **interpolated** nightlights.
- Users of FFDAS who need to aggregate uncertainties should aggregate realizations and not the uncertainties.

Posterior uncertainties

Posterior uncertainties associated with the total fossil fuel CO_2 emissions in 2003 (expressed as percentage standard deviation).



Evaluating model performance

- Uncertainties in the other sector combine those from nightlights and national emissions. The form for the nightlights uncertainty means that fractional uncertainties will be highest at low nightlight levels.

- 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.

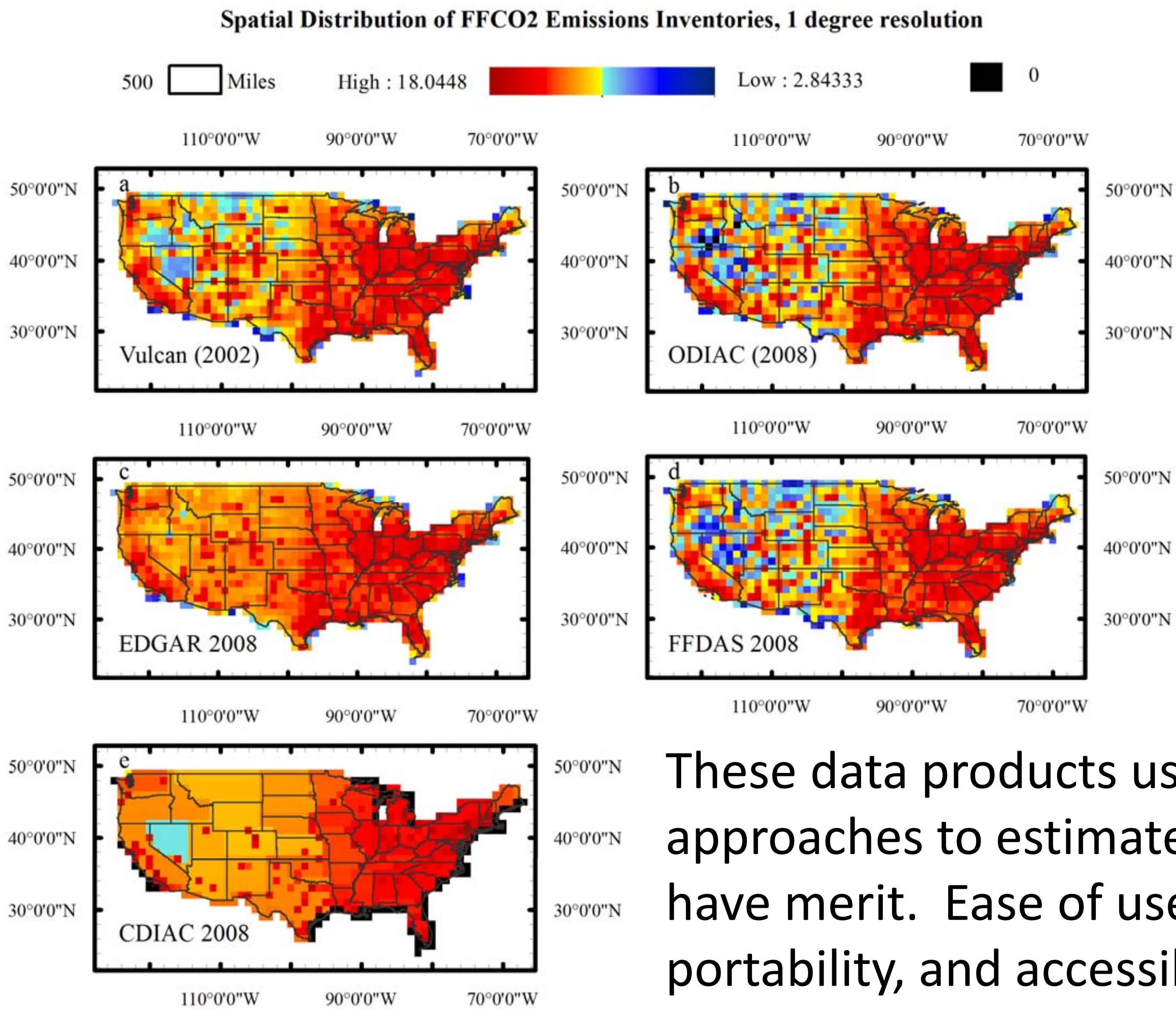
- In order to evaluate the model performance in optimizing posterior power plant emission estimates, we calculated the root-mean-square error (RMSE) of the difference between the prior and posterior values as

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

- 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.

Variance tuning

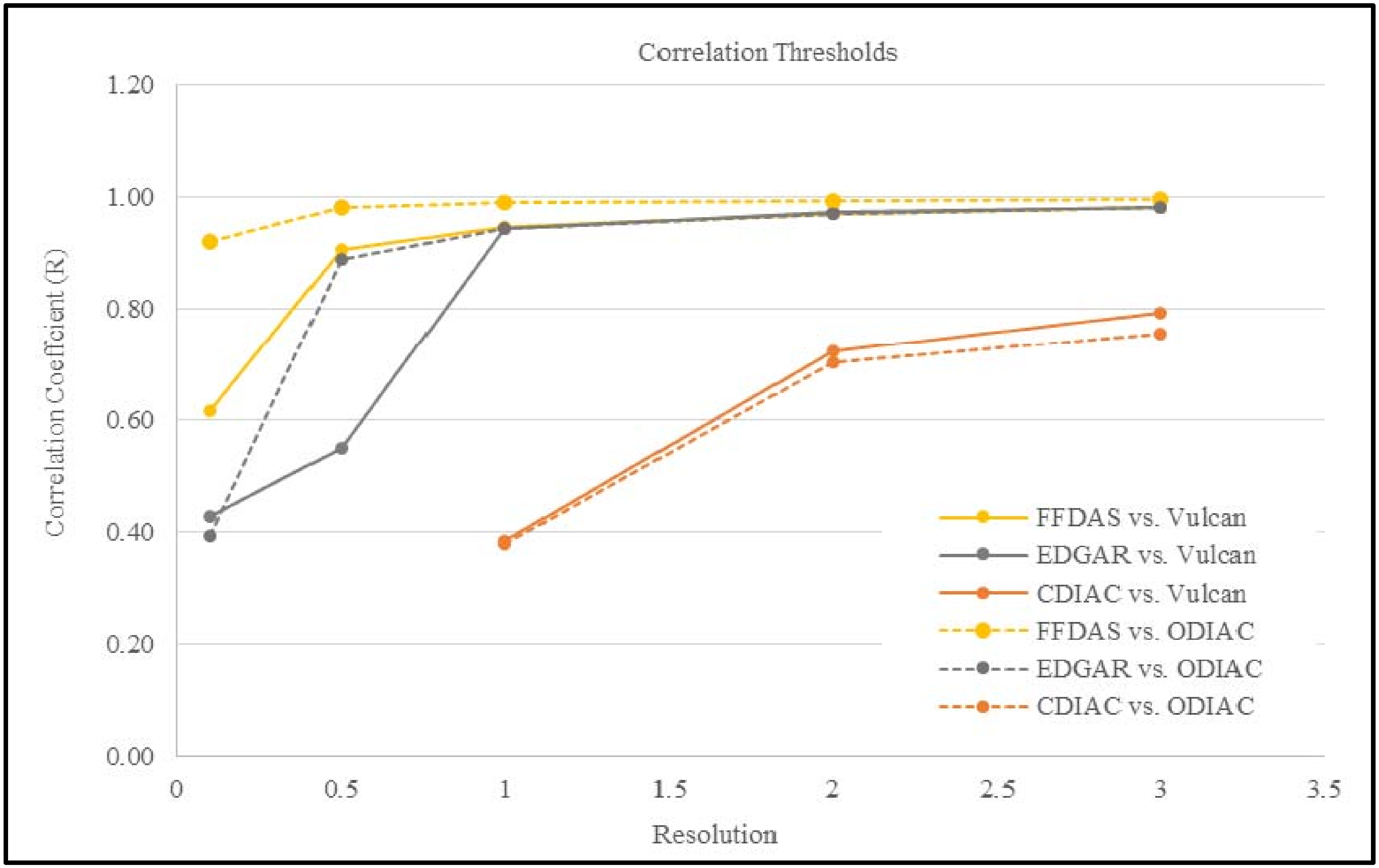
- The statistics of the differences between model outputs and observations must agree with the distributions assumed in the statistical formulation, e.g., normally distributed and independent. Similar conditions apply to the prior estimates and also to any subpopulation of the data.
- Here we limit ourselves 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 posterior uncertainties.



These data products use different approaches to estimate emissions. All have merit. Ease of use, purpose, format, portability, and accessibility all contribute to the use of the different data products. The differences however are apparent.

Comparisons

2

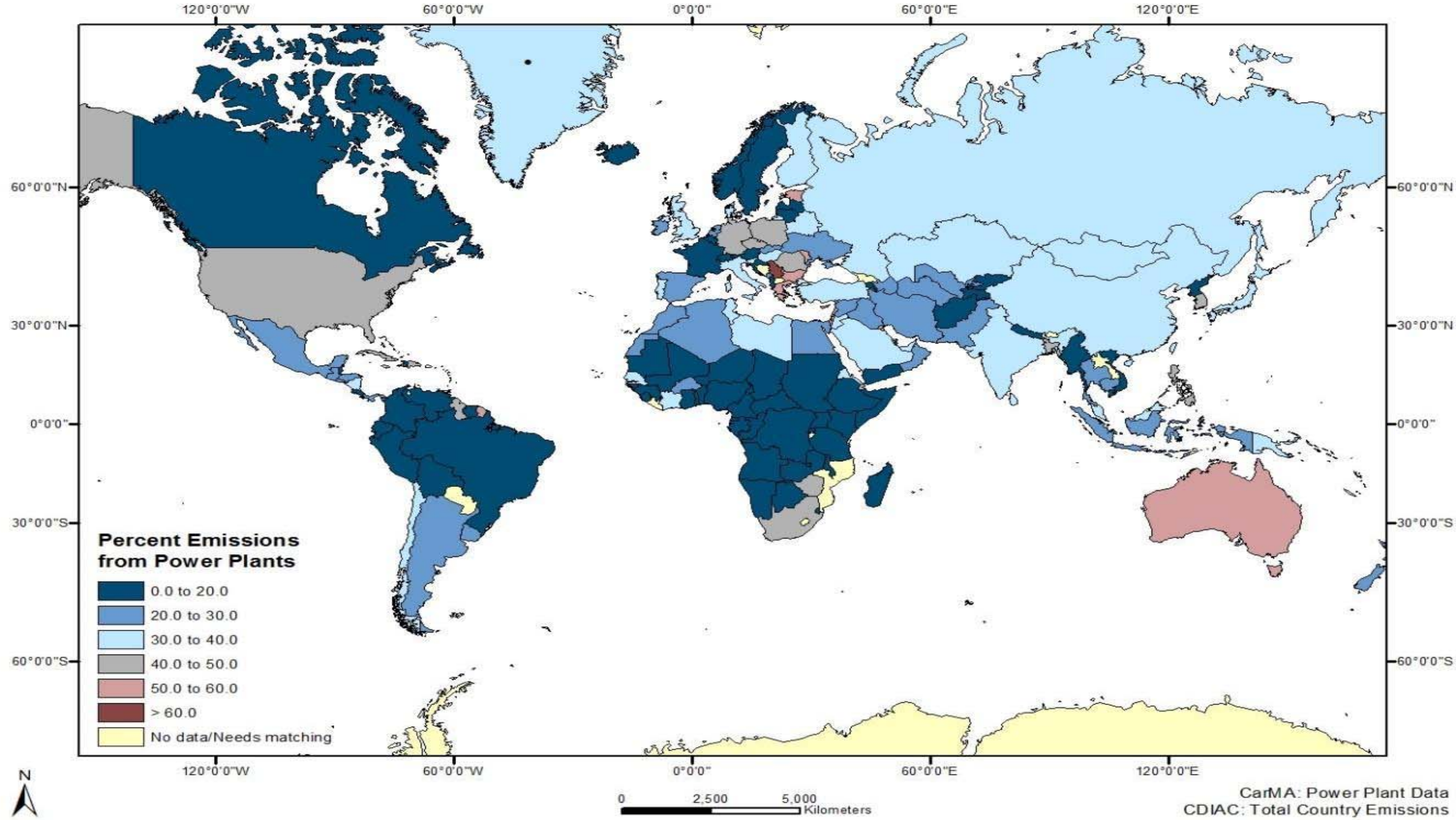


Components of Uncertainty

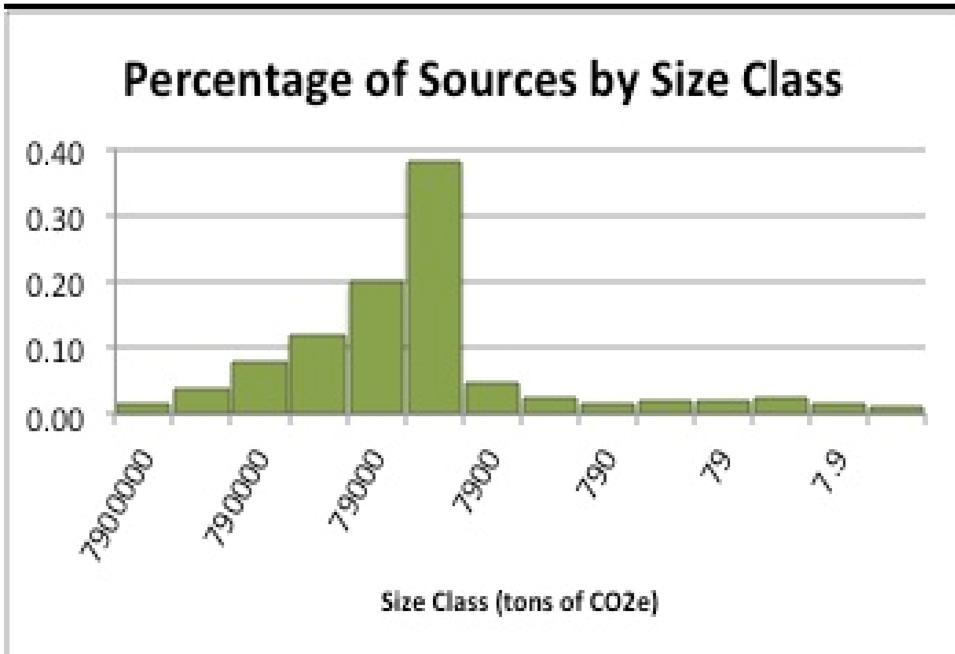
3

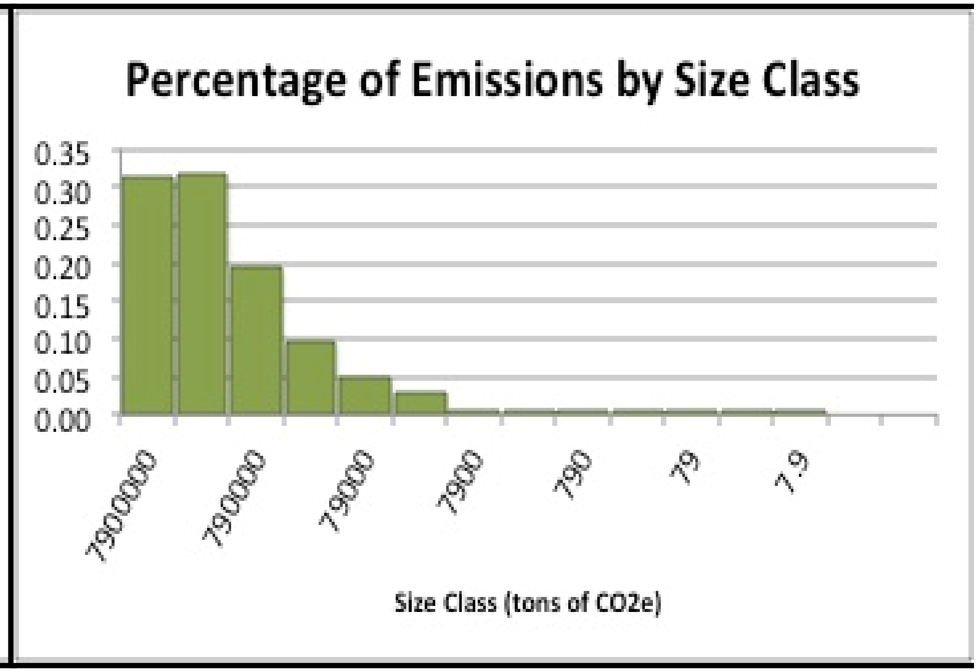
- magnitude uncertainty in country totals.
- spatial and magnitude uncertainty in directly measured emissions.
- spatial and magnitude uncertainty in estimated or modeled emissions.
- spatial and magnitude uncertainty in measured proxies.
- uncertainty in the proxy assumption.

Percentage of Total National CO_2 Emissions from Power Plants



4

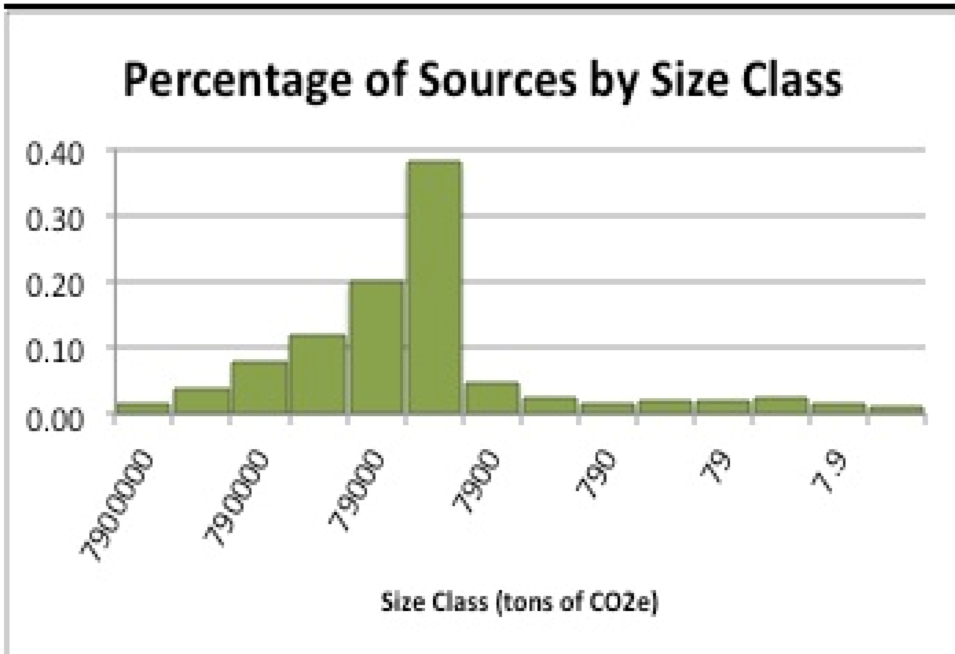


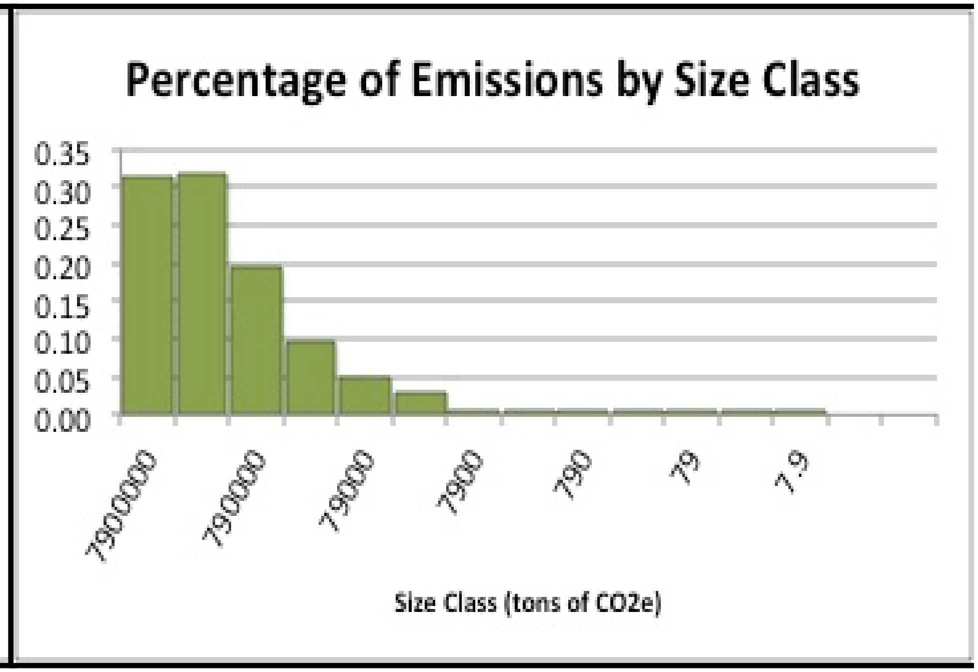


Not so many big power plants

Produce lots of emissions

5

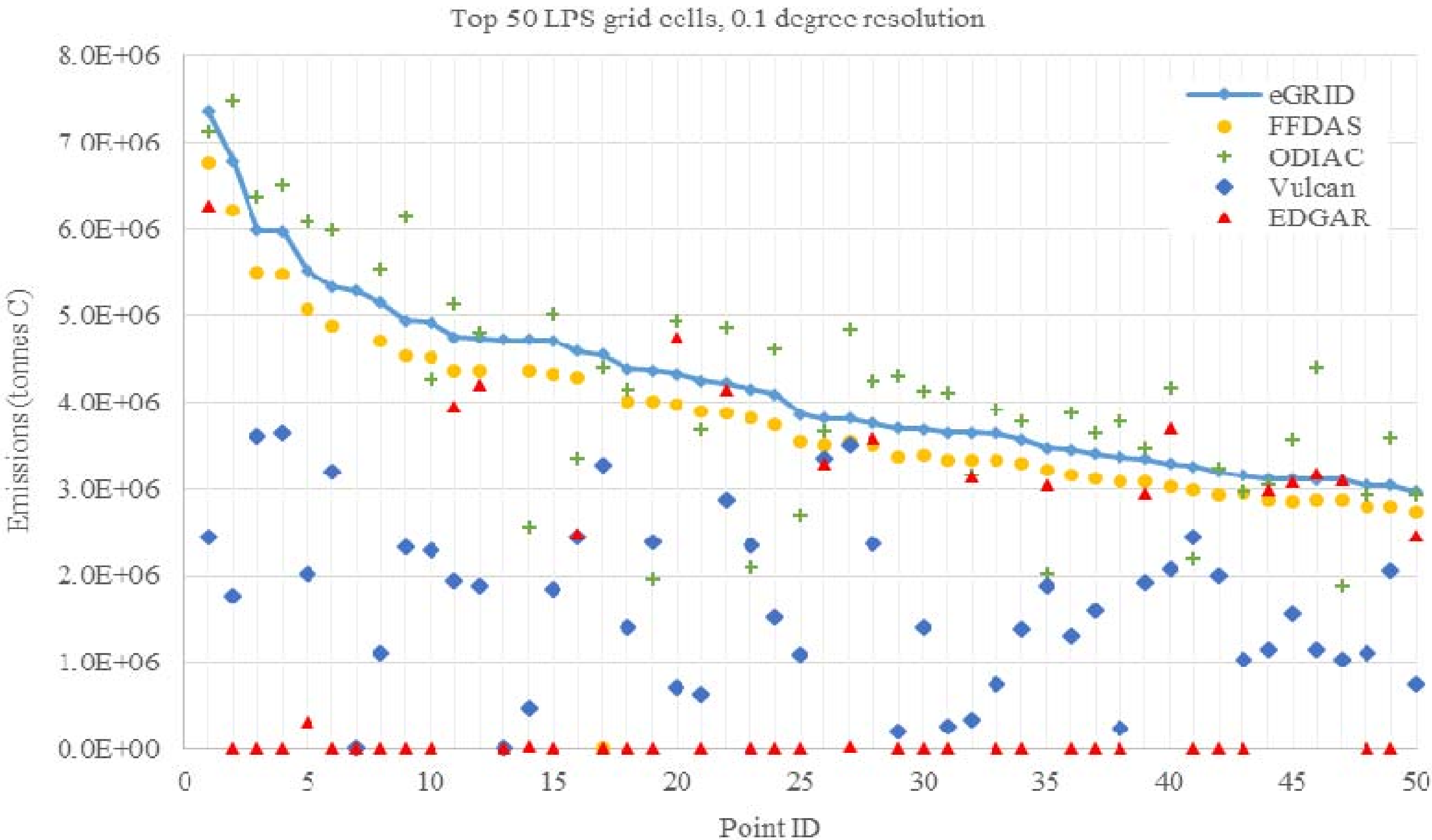




In the U.S. one third of all emissions come from only 311 sites.

6

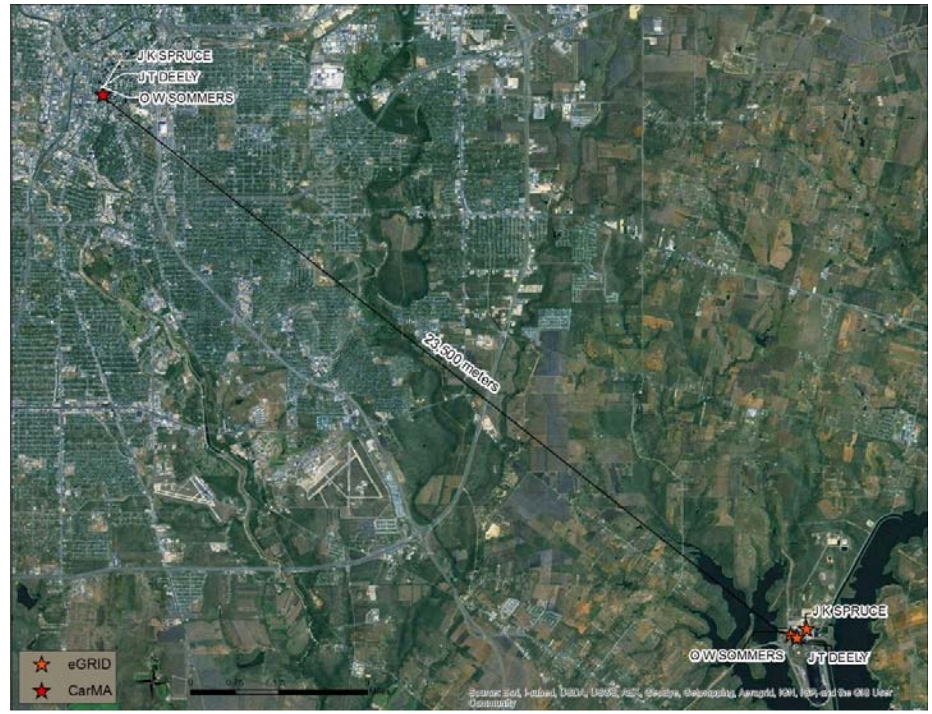
7



Top 50 LPS grid cells, 0.1 degree resolution
8.0E+06
7.0E+06
6.0E+06
5.0E+06
4.0E+06
3.0E+06
2.0E+06
1.0E+06
0.0E+00
Emissions (tonnes C)
0
5
10
15
20
25
30
35
40
45
50
Point ID
eGRID
FFDAS
ODIAC
Vulcan
EDGAR

Uncertainty in Locations

8



Default allocations, naming conventions, typos, company offices, multiple stacks, large sites, unexplained (?)

Analysis of 500 random points:

- 19% within 10km of state border
  (compared to 11% of state area)
- 68% within 2km of water
  (compared to 40.67% of state area)
- Average difference in location: 0.84km
- Average difference in location excluding zero: 1.97km
- Maximum spatial difference: 105.85km

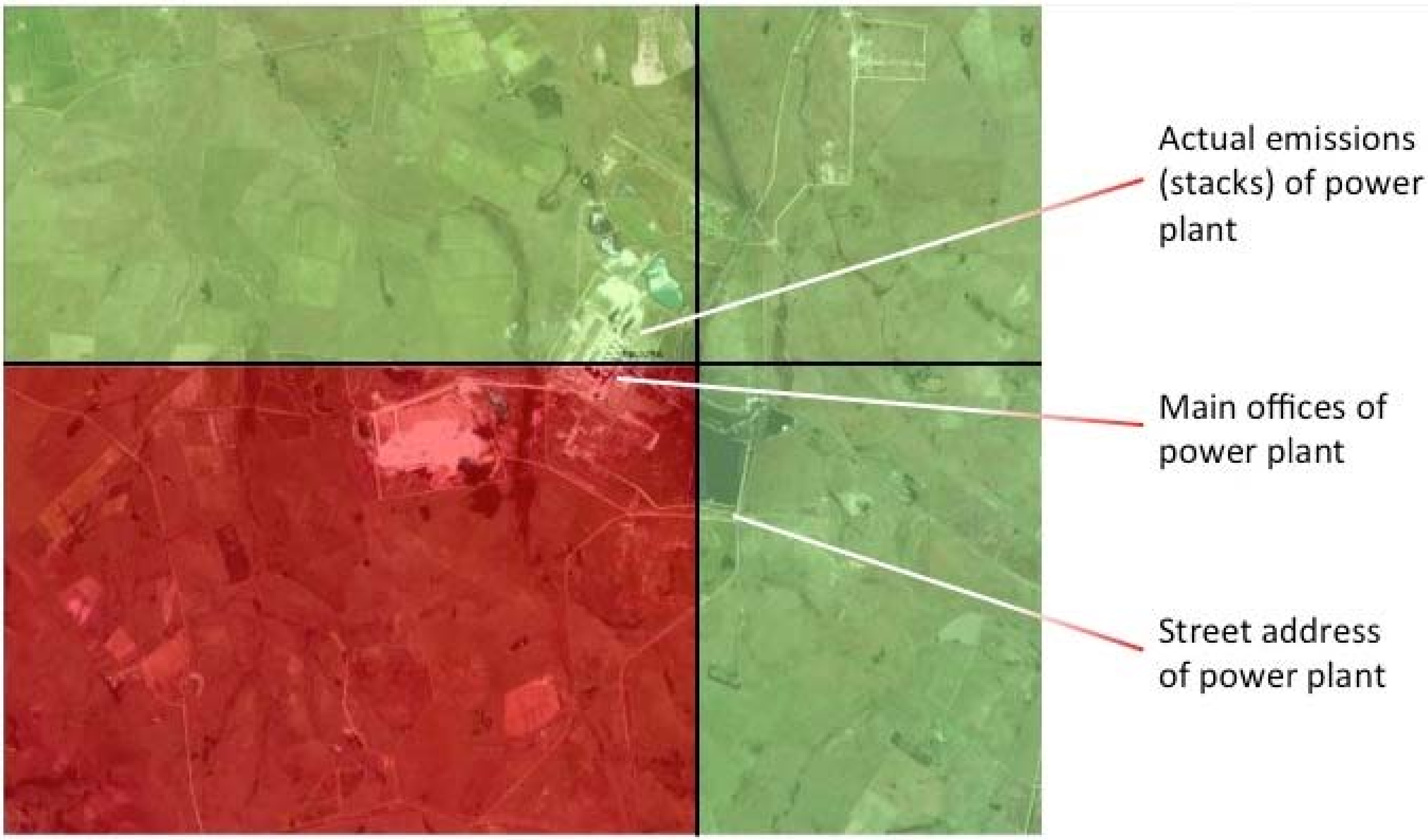
Analysis of 500 random points:

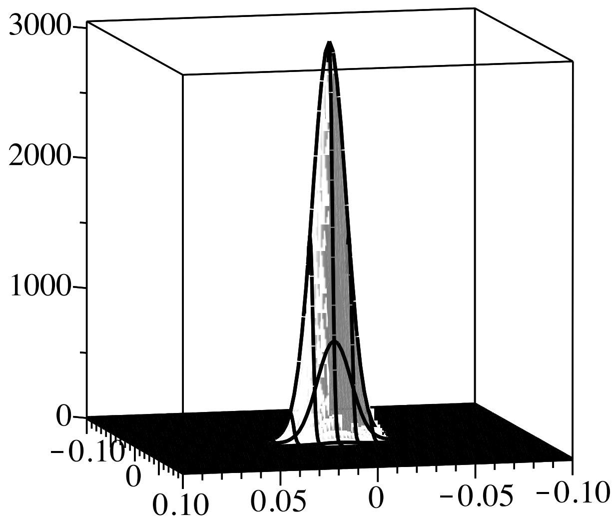
- 19% within 10...
  (co...
- ...
- ...xcluding zero: 1.97km
- ...erence: 105.85km

Of the top 81 emitters in eGRID:

- 70% are incorrect
- 60% are farther than 1km from actual
- Mean difference in location: 7.94km
- Maximum spatial difference: 121.83km

One thing is certain ...

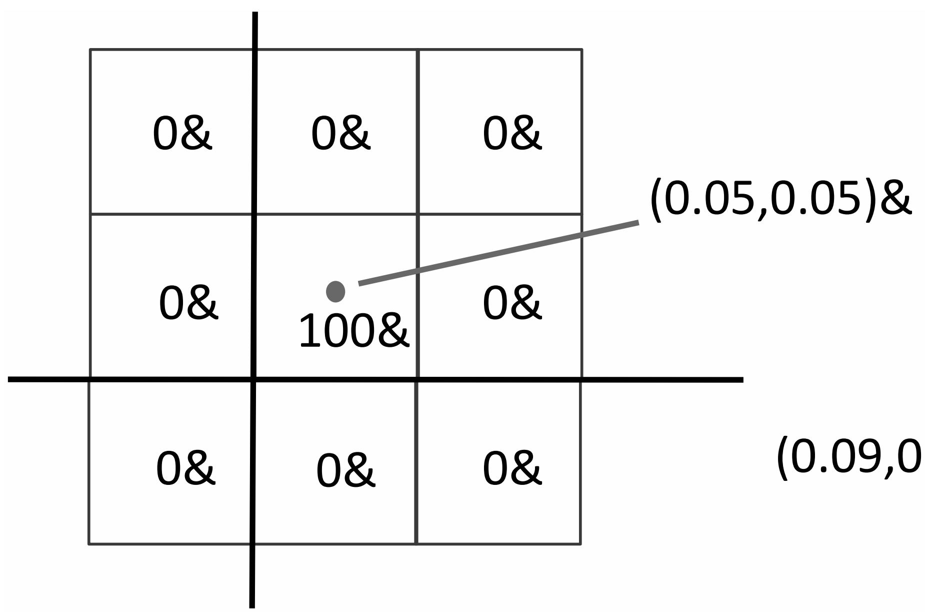


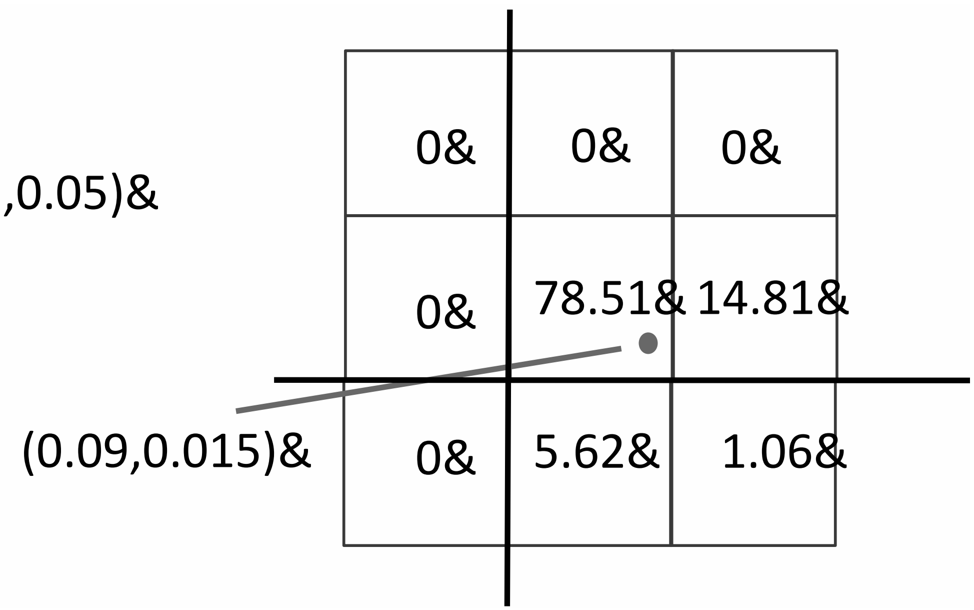


Simulated means are calculated by Monte Carlo simulation based on the distance between the reported and actual locations of large point sources. Uncertainty levels can be classified based on characteristics of the data such as fuel type and proximity of water to the reported location.

1

2

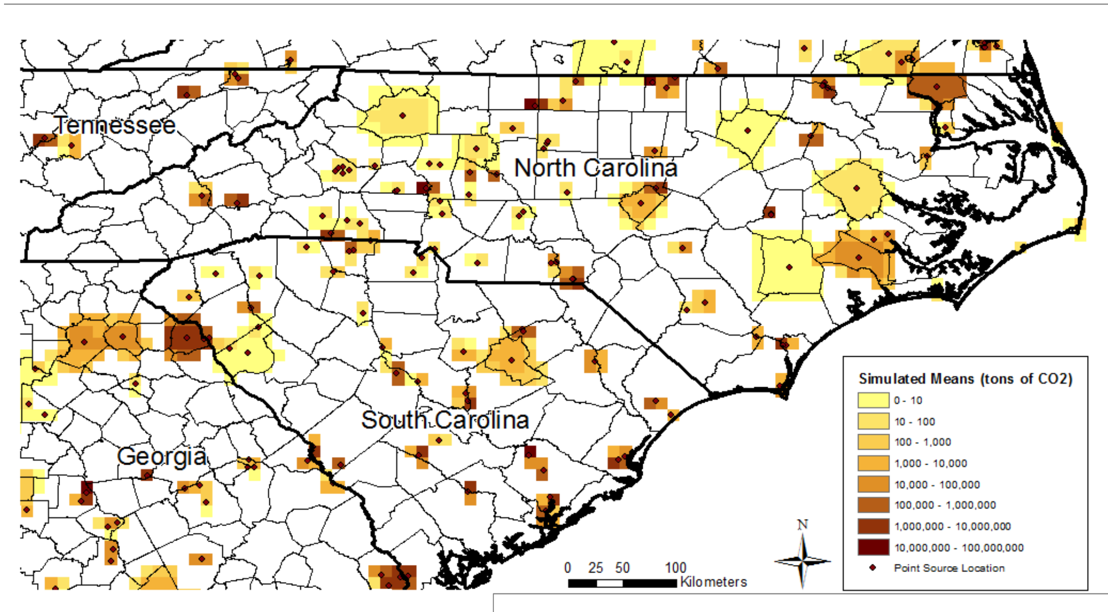




power plants reported to the centroid of a county are assumed to be arbitrarily placed.

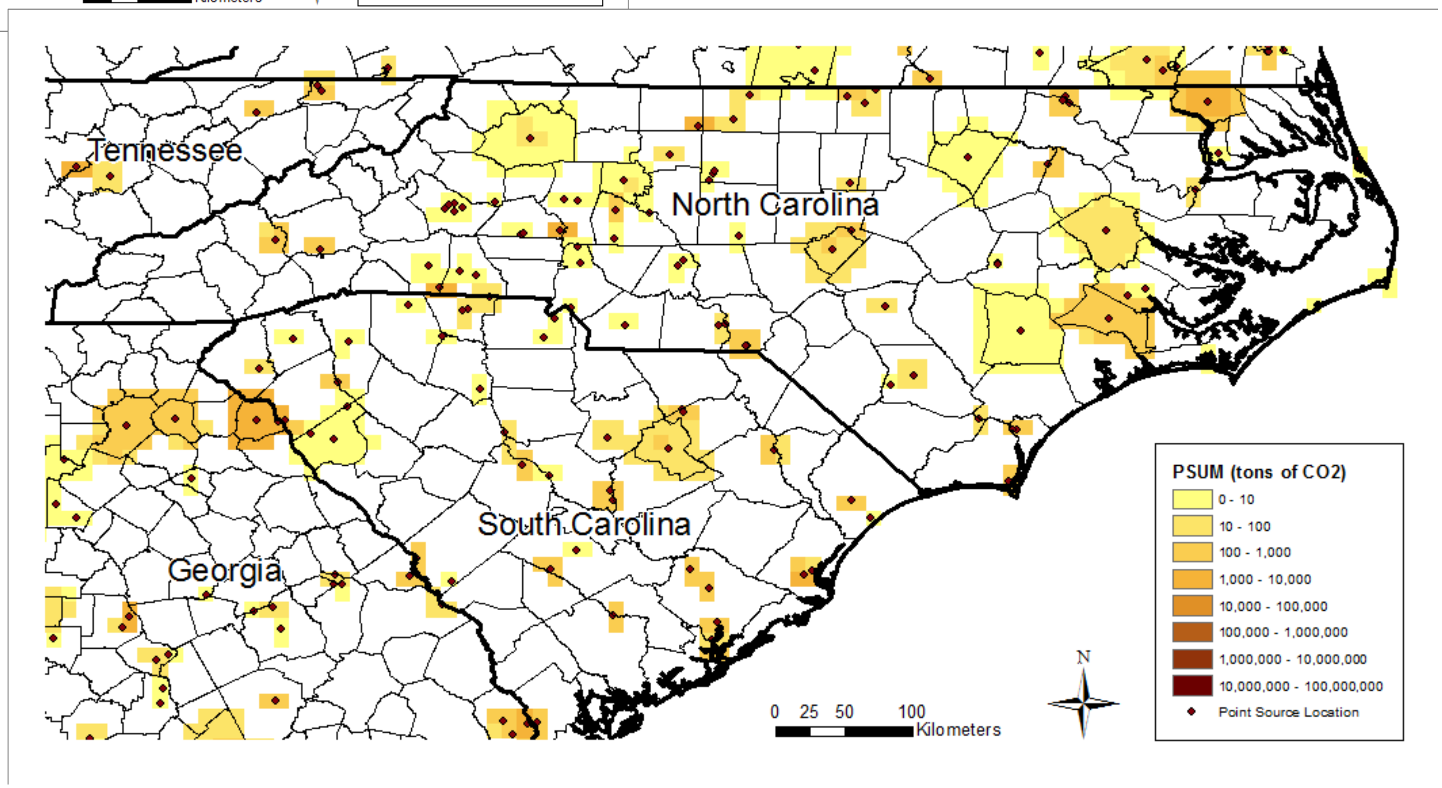
1

3

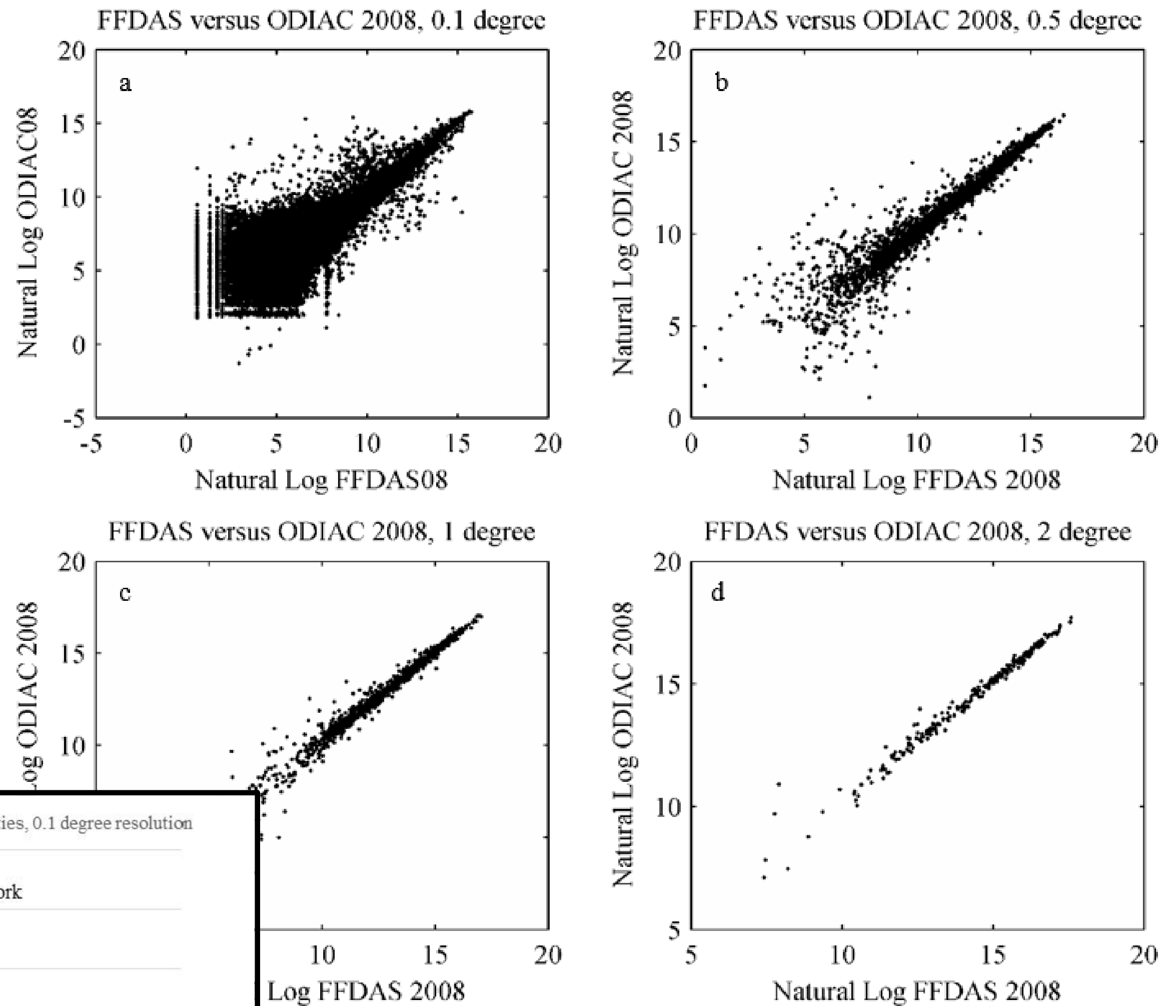


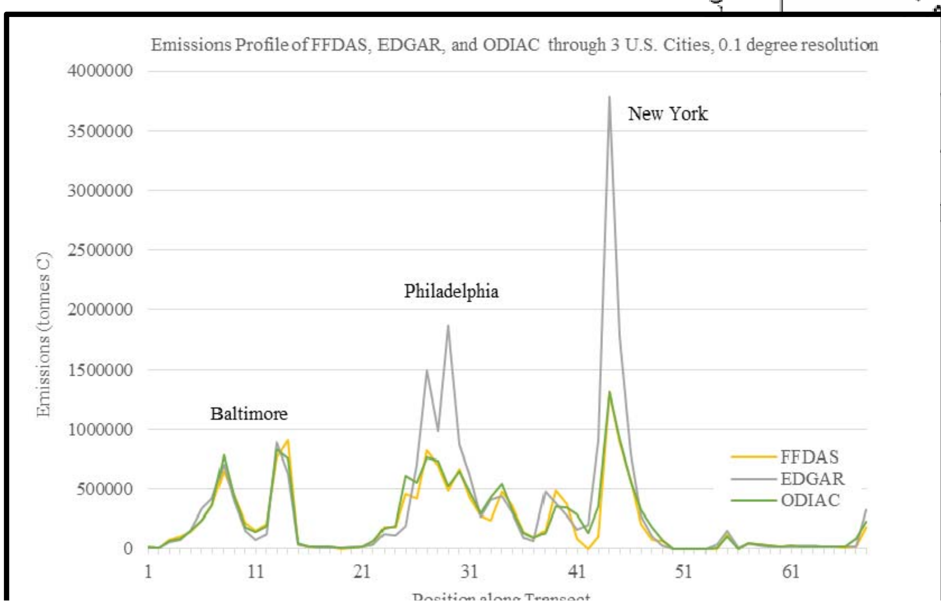
simulated means

PSUM uncertainty measure

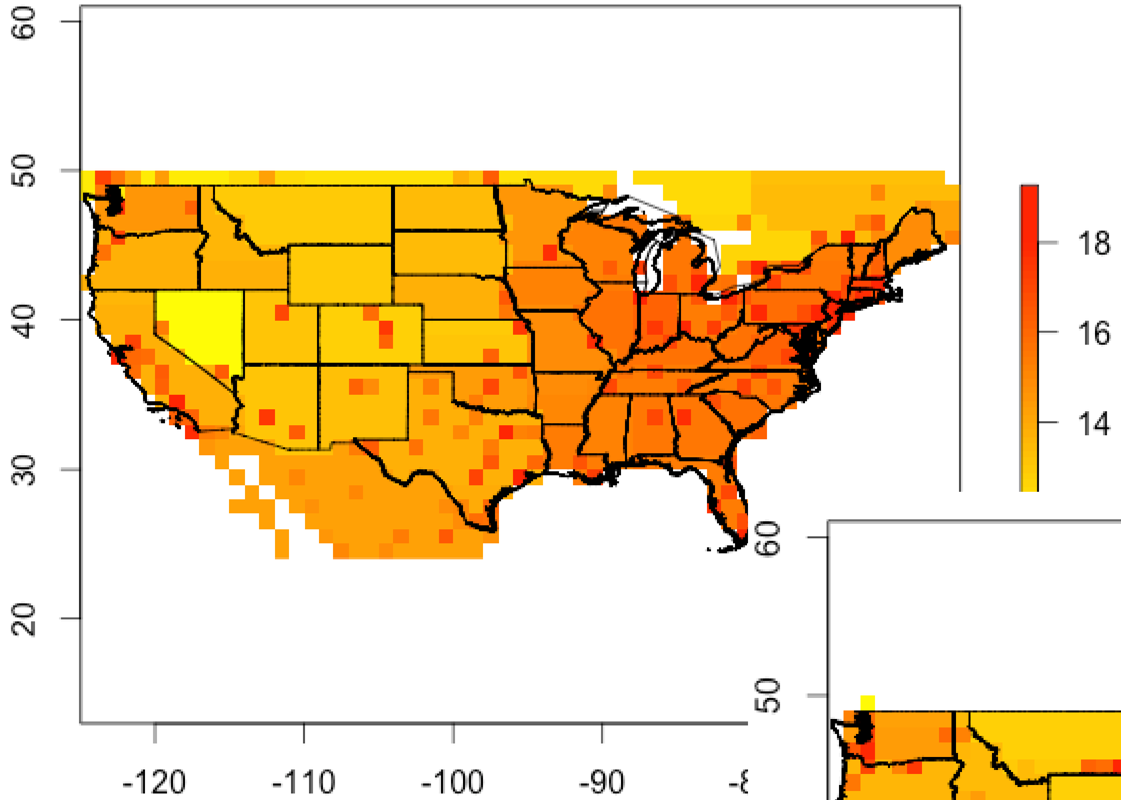


Lack of agreement at small spatial scales, potential saturation at high population or night lights levels for some methodologies, and banding in comparisons at low light levels suggest issues with the proxy for emissions.



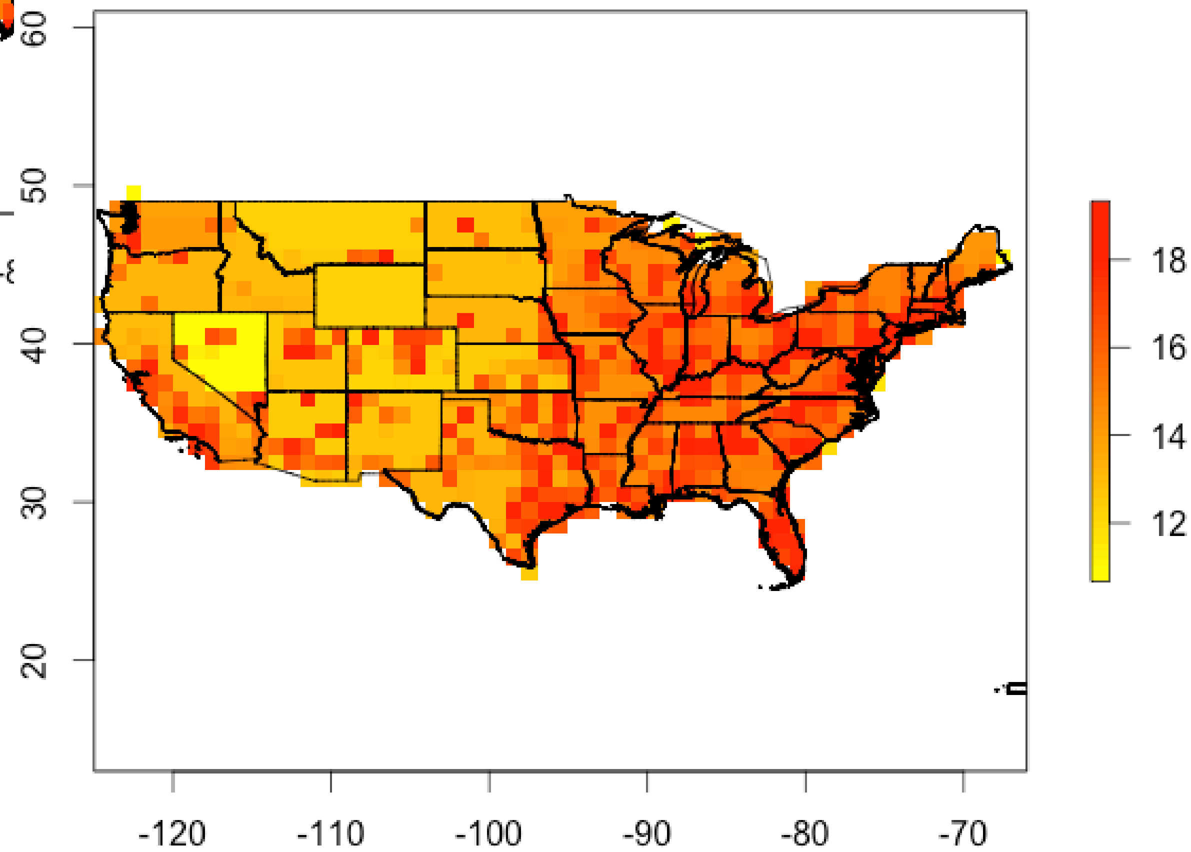


14

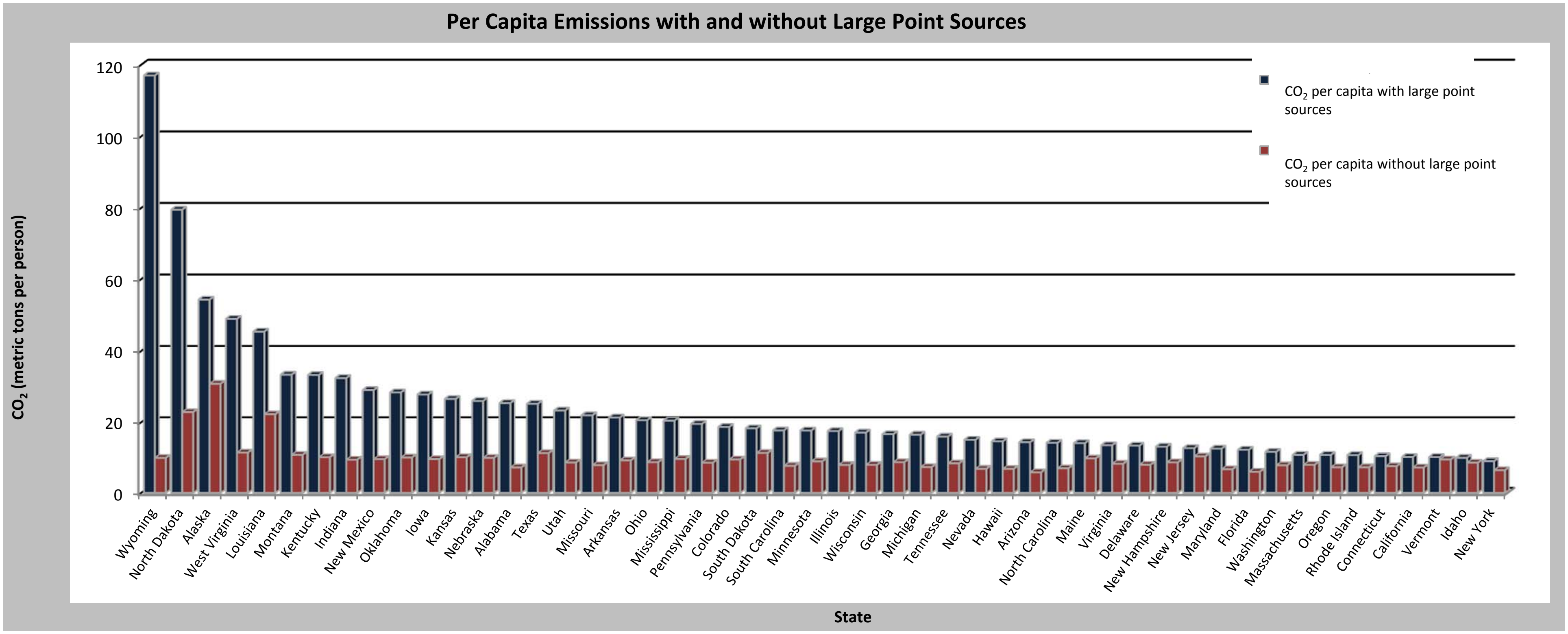


15

CDIAC Test Case

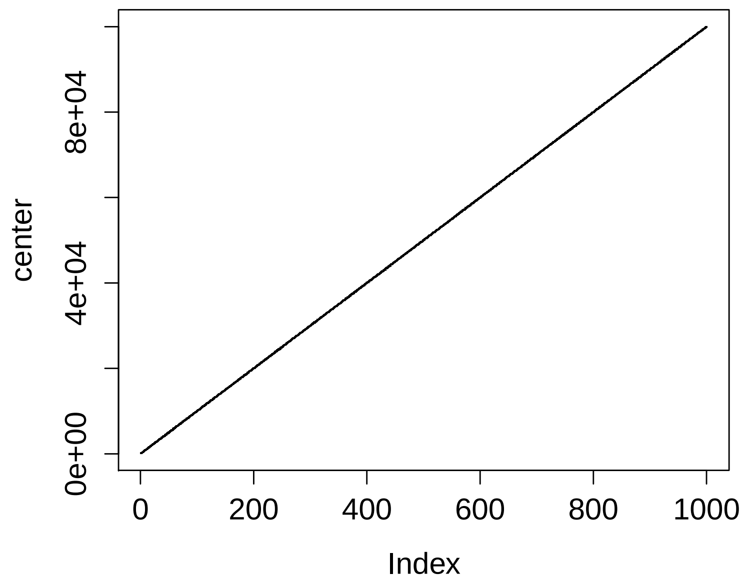


Oak Ridge National Laboratory

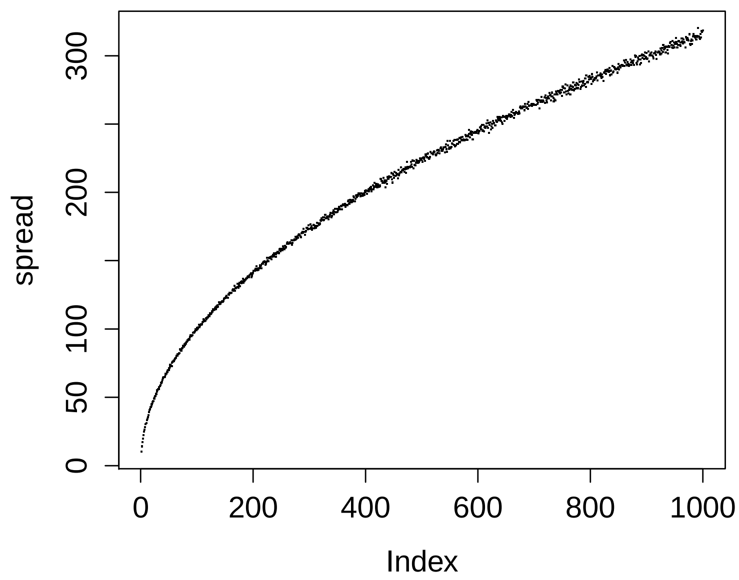


Separate treatment of large point sources drastically reduces the variability in per capita emissions values for the remaining emissions, particularly where resource availability (i.e. coal, wind, hydro) varies greatly.

16



Relationship between number of grid cells and the mean – a simple sum of the means



Relationship between number of grid cells and the SD – the square root of the sum of the SDs

The proxy error is computed by relating the variation in per capita emissions at the state level to the expected variation at the grid cell level.

$$Var_{state} \bullet N = Var_{cell}$$

where N is the average number of cells per state.

17

Land Scan Spatial Uncertainty

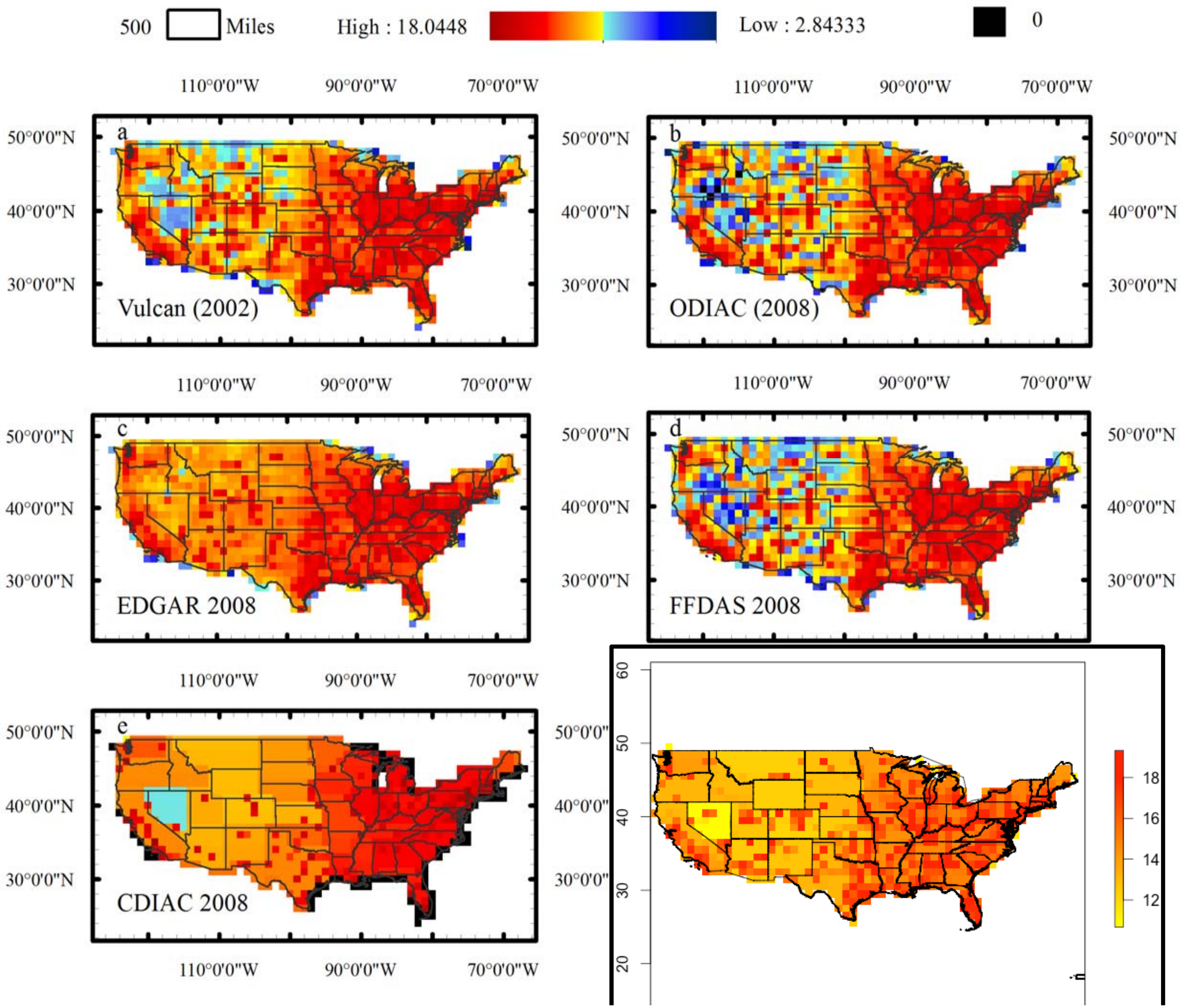
- Overestimating uncertainty based on variation of surrounding grid cells.
- Assume half a grid cell potential error (~0.5 km).

Land Scan Magnitude Uncertainty

- Overestimates based on percent error in state approximation by the US Census Bureau estimates.

LandScan has begun working on more detailed estimates of uncertainty.

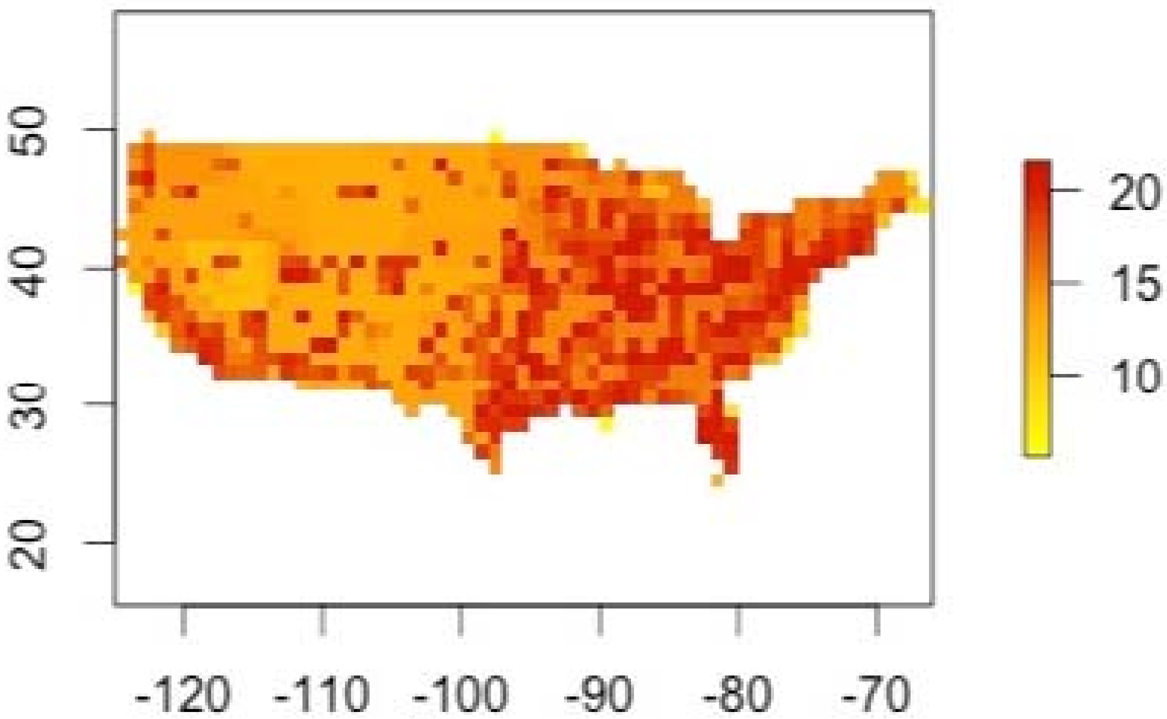
Spatial Distribution of FFCO2 Emissions Inventories, 1 degree resolution



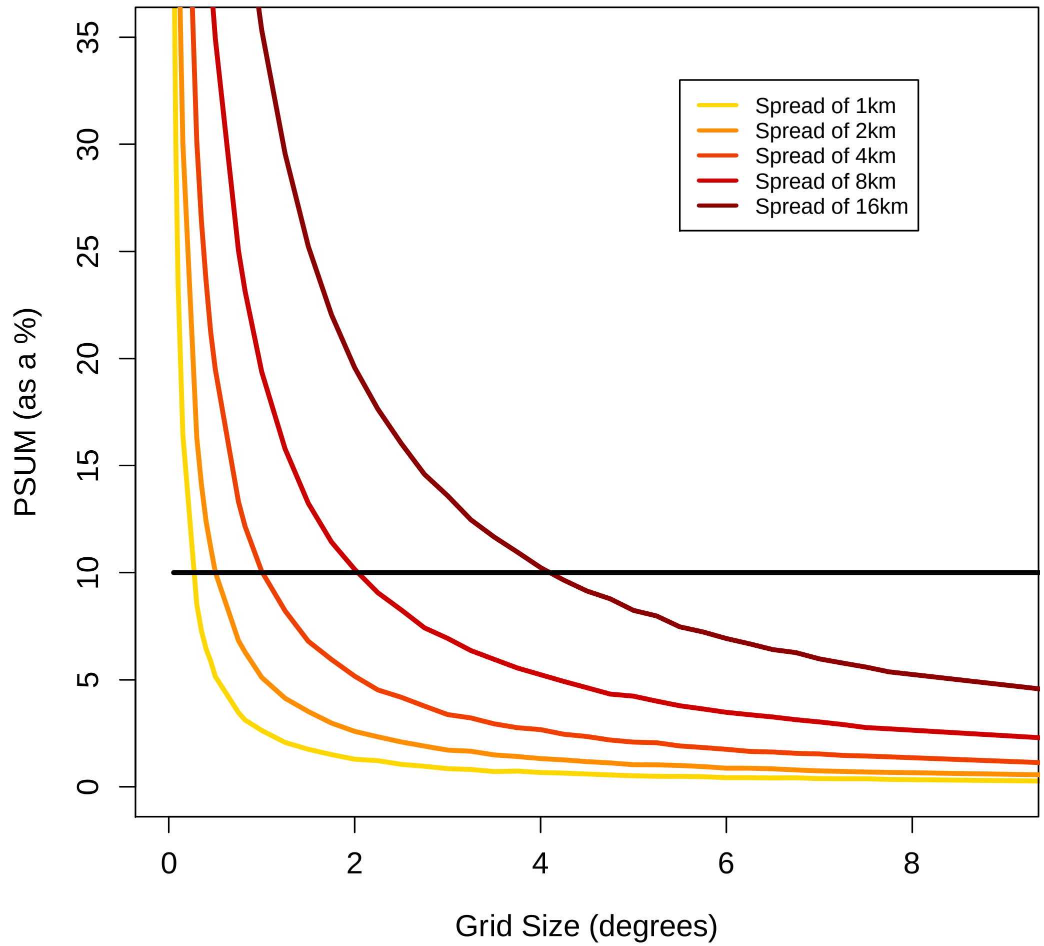
1
9

Overall Uncertainty

- Assuming each component is independent ...



These are large uncertainties, but it gives us a starting point.



Reducing Uncertainty

- Which component contributes the most?
- Which is easiest to reduce?
- What is the working spatial resolution?
- At what time resolution can we say something useful?
- With such large uncertainty, what really we really say?
- What questions can we answer?

L

Eric Marland, Gregg Marland, Susannah Hogue, Maya Hutchins, Dawn Woodard, Bob Andres

Research Institute for Environment, Energy, and Economics

NASA

Appalachian State University

Oak Ridge National Laboratory