1. Multivariate model for estimating power plant CO\textsubscript{2} emissions:

In 2009, The Center for Global Development, a non-governmental organization, produced a global power plant emissions dataset. Under the name “CARMA”, the database represents a complete compilation of all power plants in the world with location and emissions, in addition to other energy metrics. Though aimed primarily at the policymaking community and the public, the CARMA database has been utilized by the carbon cycle science community as a component of the total fossil fuel CO\textsubscript{2} emissions estimation [Oda and Maksyutov, 2011, Rayner et al., 2010].

Though an important accomplishment for the purposes intended, there remain drawbacks to the CARMA database in its application to scientific problems. 1) CARMA does not report prediction intervals or standard errors for each plant. Hence, uncertainty at the facility level is unknown. 2) CARMA’s model is created by only using data from U.S. power plants. Whether a plant is in the U.S. or not is a significant predictor of the plant’s emissions. A more accurate model can be obtained by also including publicly disclosed CO\textsubscript{2} emissions data from Europe (http://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-4), India (http://www.cea.nic.in/reports/planning/cdm_co2/cdm_co2.htm), South Africa
and through independent location information (http://globalenergyobservatory.org/). 3) CARMA’s model is a regression model for predicting the capacity factor, heat rate, and CO₂ emission factor of each power plant, and then calculating CO₂ emissions based on these inputs. We believe that better estimates, and especially standard errors, can be obtained by creating a statistical model that predicts CO₂ emissions directly. 4) CARMA does not report an explicit final form of their models but only a general model form and a list of predictor variables that were considered. These two facts make their model difficult to interpret or test. We use data from the WEPP database for the predictor variables in our model. WEPP contains information on each power plant including fuel type, generating capacity, age, and emissions controls, among others. The CO₂ emissions data used to fit the model was retrieved from the power plant matched pairs created by CARMA using publicly disclosed national datasets covering the U.S. Canada, India, South Africa, and the EU. Data used was from 2009, the most recent common year for which data was available.

The following variables were used to predict CO₂ emissions for each power plant:

Capacity: The annual electricity generating capacity of a plant, in megawatts (MW).

Relative capacity: The percentile of a plant’s capacity relative to all plants on the same regional or national electricity grid. A power transform was used to force an approximately linear relationship between relative capacity and CO₂ emissions.

Capacity factor: Capacity factor of a plant’s fuel, averaged over all plants using that same fuel on the same grid. As with relative capacity, a power transformation was used.

Generator type %: This reflects the percent of the plant’s total capacity associated with
each fuel category and prime mover: coal, petroleum with steam turbine, other petroleum, gas with steam turbine, gas with combustion turbine, other gas, and other. The “other” category was less than 1% of the total. US: An indicator variable for whether the plant is in the U.S. or not. Steam type: For steam-based units only, indicates whether the steam generator is supercritical or subcritical.

To account for the size of the power plant, we added interaction terms between the relative size of the power plant and the other variables. Power plants were binned according to whether their capacity in megawatts was less than 200, between 200 and 1,000, between 1,000 and 2,000, or greater than 2,000.

The model used is:

\[
\text{carbon} = \beta_0 + \beta_1 \text{capacity} + \beta_2 \text{relative capacity}^{10} + \beta_3 \text{fuel category} \%
+ \beta_4 \text{capacity factor (fuel category)} + \beta_5 \text{capacity factor (fuel category)}^2 + \beta_6 \text{US}
+ \beta_7 \text{steam type} + b [\text{size : (capacity, relative capacity, capacity factor, fuel type)}]
+ \text{error}
\]

where, the colon represents a variable for size plus first-order interaction terms between size and the variables indicated, and \(b\) is a vector of coefficients on those terms.

Estimated “error” resulted from the regression model is used as prior uncertainty of each plant in FFDAS.

1.1 Estimating standard error of disclosed plants

To estimate the standard error of the CO\(_2\) emissions of plants for which disclosed emissions data is available, we use the IEA and EPA/CAMD datasets (http://www.eia.gov/electricity/data/eia923/, http://ampd.epa.gov/ampd/). Because these two sets contain estimates of emissions from the same plants for the same months, the
variability between them provides evidence of the accuracy of these disclosed measurements of emissions.

Let $X_{it}$ and $Y_{it}$ be the reported emissions for plant $i$ in month $t$ from CAMD and EIA, respectively. We then run the simple linear regression:

$$Y_{it} = b_0 + b_1 X_{it},$$

and compute $s = \text{average}(s_{it})$, where $s_{it}$ is the standard error of the prediction on plant $i$ at time $t$. The value $s$ thus gives us an estimate of the true standard errors that would result from taking many estimates of the emissions at a given plant and time. We use $s$ as our estimate of the standard error for all plants with disclosed emissions, including those from non-U.S. plants for which only one data set of disclosed emissions exists. This estimate gives an adequate approximation of the standard error given the limited data available.

2. Short-term variations of nightlights

The following analysis and figure demonstrates that observed patterns of FFDAS CO$_2$ emissions anomalies (discussed in Section 3.4 of the manuscript) closely follow patterns of nighttime lights anomalies.

Short-term variations of nighttime lights about the local trend is calculated as follows:

- Remove the long-term trend;
- Calculate the standard deviation of the resulting time series;
- Normalize the time series values by the standard deviation.
We examined the years 2006 and 2010; years for which non-interpolated nightlights observations are available.

**Figure 1**: Annual nighttime lights anomalies for two years on either side of the Global Financial Crisis. (a) 2006; (b) 2010. Units: weighted standard deviation (see main text for details on units).