Center for Global Development

Working Paper Number 145
May 2008

Calculating CARMA: Global Estimation of CO2 Emissions from the Power Sector

By David Wheeler and Kevin Ummel

Abstract

This paper provides a detailed description and assessment of CARMA (Carbon Monitoring for Action), a database that reports CO2 emissions from the power sector. We built CARMA to assist the millions of concerned global citizens who can act to reduce carbon emissions once they have timely, accurate information about emissions sources. CARMA also lays the groundwork for the global monitoring system that will be necessary to ensure the credibility of any post-Kyoto carbon emissions limitation agreement. CARMA focuses on the power sector because it is the largest carbon dioxide emitter (26% of the global total), and because power plants are much better-documented than many sources of carbon emissions. The CARMA database and website put anyone with web access a few keystrokes away from detailed knowledge about power plants and the companies that own and operate them. CARMA includes many aggregation tools, so it can be used for local, regional, national and international comparisons. The database also offers complete information about power plants and companies that do not emit carbon because they use non-fossil energy sources (nuclear, hydro, solar, wind, biofuels, geothermal, etc.). In this paper, we provide a description of CARMA's methodology, an assessment of its strengths and weaknesses, and some tests of its accuracy across countries and at different geographical scales. While CARMA performs well in these tests, we recognize that it is far from perfect. We therefore extend the following invitation to any power plant or company that disputes our estimates: Provide us with better data, verified by an appropriate third party, and we will incorporate them in CARMA.
.

The Center for Global Development is an independent think tank that works to reduce global poverty and inequality through rigorous research and active engagement with the policy community.

www.cgdev.org

Calculating CARMA: Global Estimation of CO2 Emissions From the Power Sector

David Wheeler
Kevin Ummel

Center for Global Development

May, 2008

For useful comments and suggestions, our thanks to Nancy Birdsall, Chris Bergeson, Ken Chomitz, Dennis de Tray, Ellen Giles, Robin Kraft, Lawrence MacDonald, Tom Tietenberg, Dave Witzel, Chris Woltz, the Forum One CARMA development team, and many commentators who have e-mailed their views since the CARMA website was launched. Our particular thanks to our colleague David Roodman for his extensive comments on a previous draft. All remaining errors are our own.

1. Introduction

The IPCC's Fourth Assessment Report warns that unabated carbon emissions will create a global environmental catastrophe, but political leaders remain deadlocked about which countries should limit emissions, and how much. At the same time, millions of concerned global citizens in developed and developing countries want to make their voices heard as consumers, investors, employees, voters and community activists. All they need is timely, accurate information about the sources of carbon emissions. We built CARMA (Carbon Monitoring for Action – www.carma.org) as a first step toward meeting that need.

CARMA focuses on the power sector because it is the largest carbon dioxide emitter (26% of the global total), and because power plants are much better-documented than many sources of carbon emissions. With resources provided by our independent think tank, the Center for Global Development, we have built a database and website that put anyone with web access a few keystrokes away from detailed knowledge about power plants and the companies that own and operate them. CARMA includes many aggregation tools, so it can be used for local, regional, national and international comparisons. The database also offers complete information about power plants and companies which do not emit carbon because they use non-fossil energy sources (nuclear, hydro, solar, wind, biofuels, geothermal, etc.).

While CARMA is primarily a tool for citizen action, it also begins laying the groundwork for the global regulation of greenhouse emissions that may be necessary for our common survival. Most proposals for regulation advocate cap-and-trade or emissions charge systems that share three features: They are global in scope, they depend on information about emissions from each polluter, and they are transparent to ensure credibility. CARMA offers all three features for the power sector, in order to provide a first model for the information systems that will accompany global regulation. CARMA is as comprehensive as we can make it, covering over 50,000 power plants, 20,000 companies and 200,000 locales.[1] The online database discloses publicly-reported or estimated current emissions, emissions in 2000, and future emissions from published capacity expansion plans. It provides quarterly updates, as well as tools for ranking and comparing power facilities, power companies, and geographic areas (countries, states/provinces, cities and, in the US, counties, congressional districts and zip codes). For the thousands of plants and companies that have yet to report their emissions publicly, CARMA estimates emissions using methods that we will discuss in the following section. It also invites non-reporting facilities and companies to submit audited emissions reports for publication.

CARMA has ample precedents: Plant-level carbon dioxide emissions for medium-size and large power facilities are disclosed online by the US, EU, Canada and India. Where such reports are available, we integrate them into CARMA. Two NGO initiatives -- The Carbon Disclosure Project and The Global Reporting Initiative – have established guidelines and facilities for voluntary emissions disclosure by major emitters, mainly at the firm level. The closest precedent is an international database maintained by the International Energy Agency (IEA), which provides estimated CO2 emissions from about

[1] CARMA's mapping system incorporates areas around 200,000 geographic centroids, so plants may appear on more than one map when locales are close to one another.

5,000 power facilities worldwide. However, this database is only sporadically updated and it is available only in spreadsheet format.[2]

If all the emissions numbers in CARMA were UN-certified and political leaders were in agreement, global regulation of the power sector could begin immediately. CARMA obviously meets neither condition, but we hope that it does convey a simple, powerful message: If a small team at the Center for Global Development can come this far in a year of development, surely the global community can muster the resources to develop a system that qualifies for UN certification.

How far have we come? This paper attempts to provide the answer, with a description of the methodology we have employed so far, an assessment of its weaknesses, and an evaluation of its predictive accuracy. The remainder of the paper is organized as follows. In Section 2, we describe CARMA's approach to estimating the emissions of power plants worldwide. Section 3 provides a critical analysis of the estimation methodology, while Section 4 tests its accuracy at various levels of geographic aggregation. Section 5 includes detailed treatments of CARMA's aggregation tools and power industry coverage. In Section 6, we provide a summary and conclusions.

2. Building the CARMA Database

CARMA draws on three primary data sources: Plant-level public emissions reports from the US, EU, Canada and India;[3] global plant- and company-level data from the

---

[2] The entries in the current (2006) version of the IEA database have the following distribution by date: 3,185 (1998 or earlier), 1,340 (1999-2000), 407 (2001-2004), 59 (2005-2007)).

[3] The US Environmental Protection Agency's eGRID database is available online at http://www.epa.gov/cleanenergy/energy-resources/egrid/index.html. The European Pollutant Emission Register is available online at http://eper.ec.europa.eu/eper/flashmap.asp.. The Register includes CO2 emissions reports for several hundred major emitters in the EU. Environment Canada provides greenhouse gas emissions reports at http://www.ec.gc.ca/pdb/ghg/onlinedata/DataAndReports_e.cfm. The Indian

World Electric Power Plants Database (WEPP), a subscription information service[4], and country-level power production data from the US Energy Information Agency.[5] We combine these databases to produce plant-level reports of CO_2 emitted (in short tons), energy produced (MWh), and emissions intensity (CO_2/MWh). Our approach merges public disclosure and WEPP data at the plant level, and this requires information-based matching because the two databases do not use the same identification codes for plants. We perform the match on plant names and locations, using a fuzzy-logic algorithm suggested by our CGD colleague, David Roodman.[6] We supplement the results with visual inspection, using information from the public and WEPP databases on plant names, locations and operational scale. Although this process has been quite time-consuming, we have undoubtedly missed some plants in the matching exercise.

The current version of CARMA includes information for 51,373 global power facilities in the WEPP database, of which 25,339 emit CO_2 and 26,034 do not because their power sources are nuclear, hydro, or other renewables. Among these power plants, we have matched 4,435 publicly-reported reported facilities – 2,922 CO_2 emitters and 1,513 non-emitters. The matched, publicly-reported emitters by geographic region are as follows: US 2,469, Canada 104, EU 260, India 89.

Our matched information for the US and India includes publicly-reported CO_2 emissions and power outputs, while the matched data for Canada and the EU include only

---

Power Ministry reports emissions at http://www.cea.nic.in/planning/c%20and%20e/Government%20of%20India%20website.htm.

[4] This database is available online at http://www.platts.com/Analytic%20Solutions/UDI%20Data%20&%20Directories/World%20Electric%20Power%20Plants%20Database/.

[5] Available online at http://www.eia.doe.gov/emeu/international/electricitygeneration.html.

[6] See Eugene Myers, "An O(ND) Difference Algorithm and its Variations," Algorithmica Vol. 1, No. 2, 1986. Available online at http://www.xmailserver.org/diff2.pdf.

CO_2 emissions. Where the matched information is available, we reproduce it in CARMA or modify it to account for plant-level changes in capacity, technology or fuel use since the public reporting date.

The public databases, plus our information on plant-level fossil and renewable energy sources, enables us to account for emissions from 28,956 of CARMA's 51,373 facilities without estimation. As previously noted, 26,034 of these plants emit no CO_2 because they use nuclear, hydro or other renewable energy, and 2,922 have publicly-reported CO_2 emissions. We estimate emissions for the remaining 22,417 facilities using a regression analysis of our matched sample of 2,469 CO_2-emitting facilities in the US. For plant-level power production, the public data sources enable us to account for 4,071 facilities (3,869 in the US, 202 in India). We estimate power production for the remaining 47,302 facilities, as described below.[7]

The need to estimate power production for so many plants is the weakest link in CARMA. From the WEPP database, we have accurate and up-to-date plant-level information on operational capacity, technology and energy sources (WEPP is updated quarterly). However, this database provides no information on the capacity factor – the annual percentage of potential operating hours during which the plant actually produces power. A plant's capacity factor is significantly affected by its scale, technology and energy sources. For any plant, however, the capacity factor can vary greatly from year to year as supply/demand conditions change and necessary maintenance is performed.

Our inability to observe locally-determined capacity factors for so many plants introduces significant random errors into our CARMA estimates, as we will show in the following section. However, we do have an important control for country-level elements

[7] For a detailed discussions of regression results and variables, see Appendix A.

that can also affect capacity factors: economic conditions that affect the general demand for power, and distinctive features of the public/private power mix. For example, several countries in Eastern Europe have lower-than-normal capacity factors because they are still adjusting from the era of central planning,

Combining plant- and country-level information sources, we use a five-step procedure to estimate power production: (1). Estimate plant-level capacity factors using a regression analysis of matched data for facilities in the US; (2). Multiply estimated capacity factors by operational plant capacities reported by the WEPP database; (3) Combine estimated and publicly-reported plant-level power to obtain total power production by energy source for each country in CARMA; (4) For each energy source (e.g. fossil, hydro, other renewables), divide this total by the corresponding total from the US Energy Information Agency to obtain an adjustment factor; (5) For each country and energy source, multiply each estimated plant-level power output by the relevant adjustment factor (publicly-reported power outputs are not adjusted).

For each plant in CARMA, we divide CO2 emissions by power production to obtain emissions intensity. From the 3,869 matched observations for the US, we develop sample statistics for emissions intensities for small, medium and large power plants. We use these statistics to set upper and lower limits for plausible emissions intensities, and trim outliers to the limits for each size class. For the trimmed cases, we have to adjust the original numbers for total emissions and total power produced. We re-estimate emissions from power using trimmed intensities when both emissions and power are estimated, and re-estimate power from emissions when only emissions are reported.[8]

---

[8] In our database there are no cases in which we observe power but not emissions.

3. Estimation Assessment

Our estimates are built on two core regressions, fitted to large datasets that match plant-level information from the US EPA's eGRID database and the WEPP database. The two dependent variables for our analysis – emissions and power – are drawn from eGRID, and the independent variables from WEPP. Our global estimation exercise requires this approach, since we only have WEPP data for estimation outside of the US, and many US power plants covered by the WEPP database are not covered by eGRID. The WEPP data are provided at the generator level, so we aggregate to the plant level for matching with eGRID. Once we have estimated the two core regressions using the matched data, we apply the estimated parameters to WEPP data for plants whose emissions and power production are not reported.

3.1 Emissions Estimation

For the emissions model, we regress CO2 emissions on power produced, and capacity shares for fuel sources and detailed combustion technologies (e.g., gas turbine in combined-cycle; internal combustion engine with heat recovery; steam turbine with steam sendout).[9] Appendix Table A3 reports the regression results. We obtain high significance levels for power produced, fuel sources, and a subset of combustion technologies. Figure 1 plots actual and predicted values for plant-level CO2 emissions. The overall regression fit for 2,166 observations is excellent, with an adjusted R^2 of .96 and a regression F-statistic that is significant at the .0001 level. Our results reflect normal differences in emissions intensity for coal, oil and gas, and the highly-significant emissions elasticity of power production suggests an important scale economy –

[9] After controlling for these variables, the average age of a plant's generators is not significant. We calculate plant-level capacity shares from fuel- and technology-specific information for individual generators. We aggregate to plant-level capacity shares using generator capacities as weights.

emissions increase .916% with each 1% increase in power production, controlling for other variables. Figure 1 indicates that the quality of the fit is equally good for small and large facilities. In the middle ranges there are some outlier predictions, but the great majority are near the regression line.[10]

In summary, our large-sample results indicate that knowledge of a US facility's power production, fuel source and combustion technology is sufficient to predict its CO2 emissions with about 96% accuracy.[11]

3.2 The Role of Variable Capacity Factors

Unfortunately, our emissions regression result is an upper bound on CARMA's accuracy because we do not have power production data for most plants outside the United States. A plant's power production in megawatt-hours (MWh) is the product of its capacity factor (% of potential capacity actually employed), its capacity (MW), and its potential operating hours per year (generally 8,760(=24*365)). We have the operational capacity of each plant in the CARMA database, but we only have capacity factors for a subset of plants in the matched public data for the US and India (overwhelmingly the former).

This gap is undoubtedly the weakest link in the CARMA estimation chain. To convey a sense of the variability in capacity factors, we match power plant data for 2000 and 2004 using common identification numbers in the US EPA's eGRID database.[12] We control for interim changes by restricting the matched set to plants whose capacities in

[10] Figure 1 is in log format, so the approximately-constant deviations from the regression line across size classes also imply roughly-constant percentage deviations.

[11] In large part, we attribute the high R^2 to the straightforward chemical relationship between fuel combustion and CO2 emissions. Once fuel sources and power outputs are specified, most of the remaining variation in CO2 emissions across plants comes from variations in energy efficiency. Our results suggest that our controls for combustion technologies are sufficient to capture most of this variation in efficiency. Our thanks to our colleague David Roodman for his comments on this issue.

[12] The 2004 capacity factor numbers are from eGRID 2007, which does not include more recent data.

2000 and 2004 are identical. These identical plants yield the scatter plot in Figure 2, which reveals a pattern of significant variation despite a mere 4-year interval between observations.[13] The implications for CO2 emissions are displayed in Figures 3 and 4. Figure 3 plots emissions in 2000 and 2004 for matched plants that have identical capacities; emissions are in log form to permit comparison for small and large plants simultaneously. Here the effect of variable capacity factors clearly translates to substantial variation in CO2 emissions for identical plants, particularly in the smaller size range. Figure 4 relaxes the equal-capacity condition and displays the emissions scatter for all matched plants. Here we see the effect of variable capacity factors increased by variation in capacity itself (from generator retirements, additions, and changes in operational status).

For CARMA, this variation in capacity factors identifies an important limit to estimation accuracy. At the same time, it highlights what CARMA is actually trying to measure. CARMA's plant-level emissions and power reports are best understood as long-run average estimates for these variables under standard operating conditions for plants with the same capacity, combustion technology, energy sources and country-specific circumstances. Even if power information for a particular year were available, actual emissions would be estimated with some error. This error is significantly compounded when the power estimate depends on a capacity-factor estimate that is itself subject to large error.

---

[13] The regression R^2 is .73, which indicates that about 25% of the interplant differences in 2004 capacity factors cannot be explained by their differences in 2000.

3.3 Capacity Factor Estimation

For the capacity-factor model, we regress capacity factors in our matched US sample of fossil-fuel-fired plants on plant capacity (multiplied by 8760 to obtain potential annual production), the average age of generators in the plant, the percent of generator capacity on standby (as opposed to operational) status, and capacity shares for combustion technologies and fuels. Appendix Table A3 reports the regression results. For 2,296 observations; the adjusted regression R^2 is .52. We find that the capacity factor increases significantly with plant scale and average generator age, decreases significantly as the share of capacity on standby increases, and varies significantly across technologies and fuel types. We have limited estimation to observed capacity factors between 0 and 1.

For renewable energy plants, we find that capacity factors are not sensitive to scale, age and standby status. We therefore use average capacity factors by energy source from the US renewable plant data, and adjust these for other countries using the EIA data.

3.4 Backcasting and Forecasting

Given the great year-to-year variability in capacity factors, we cannot realistically expect highly-accurate estimation of plant-level CO2 emissions without current information on power output. In any case, it is not clear to us that yearly point estimates actually convey the most useful information for CARMA users . To illustrate, consider the case of a large coal-fired facility that takes half its generators offline in a particular year, for reasons related to long-run maintenance and local economic fluctuations. During this year, the plant has emissions far lower than its capacity, technology and energy source would imply. But in subsequent years, the plant will bring its generators back online and greatly increase its emissions. CARMA is designed for a variety of users who are concerned about carbon emissions, including investors, journalists, local community leaders, educators, NGO's and researchers. In most cases, we believe that these users are interested in a plant's emissions under standard conditions, not yearly point estimates that may reflect transient factors.

We follow the same logic in our approach to forecasting and backcasting.[14] CARMA's regression-based model uses data for thousands of US power plants provided by the US EPA's eGRID system for the year 2004. CARMA combines the regression parameters with information from the WEPP database to estimate emissions for 2000, 2008 and 10 years into the future. The future estimates are based on plans for new construction that are tracked by the WEPP database.

For plants that require estimation of both emissions and power, we proceed in two steps. First, we use the WEPP data to construct independent regression variables (capacity, generator age, operational status, combustion technology, energy sources) for 2000, 2008 and 10 years into the future. Then we apply the regression model parameters to the independent-variable values to obtain estimated CO2 emissions and power output. The 2008 estimates in CARMA are based on current data. For the 2000 estimates, we adjust each plant's generator stock using WEPP data on generator retirements and additions between 2000 and 2008. For the future estimates, we adjust each plant's generator stock (and generate information on new plants) using WEPP data on planned generator installation. We do not adjust for future retirements because the WEPP future retirement database is very spotty.

Projecting time trends for publicly-disclosed facilities requires a different approach. The public data for Canada, the EU and India are quite recent, so we take their online

[14] We backcast to the year 2000 in order to provide consistent information on recent trends.

estimates as current. We cannot do this for the US, since the most recent estimates are for 2004. In all cases, we adopt the same generic approach to forecasting and backcasting. For the dates of the most recent entries (current for the EU, Canada and India; 2004 for the US), we use the capacity factor and emissions models to develop emissions and power estimates for those dates. Then we adjust as previously described, accounting for generator retirements and additions to develop estimates for 2000, 2008 (identical to the online observations for the EU, Canada and India) and 10 years hence. From the benchmark-year estimates, we calculate the ratio of estimated to reported power and emissions for each plant. Then we complete the estimates for 2000, 2008 and the future by multiplying the emissions and power estimates by the ratios. This ensures reproduction of the available online reports, while preserving the relationships with past and future emissions that are indicated by our modeling approach.

In effect, our approach addresses the following question: Adjusting for generator retirements and additions, what past and future power output and emissions would this plant have if its operating environment remained constant (at the benchmark year level – current for the EU, Canada and India; 2004 for the US)? Because of episodic changes in capacity factors (and emissions intensities, which are themselves generally estimated by plant engineers or government technicians), we do not expect our backcasts to accurately match any public reports available for the same plants in the year 2000. We could have adopted another approach for CARMA, transcribing past reports as well as current reports in our online database. We considered this option, but decided that the standardized approach would be more informative for CARMA's users. We continue to think about alternatives, however, and we may opt for transcription if it seems warranted by user requests and our own continued reflection on this problem.

4. Estimation Accuracy

CARMA uses regression models fitted to plant-level variables in a large US dataset. Using the regression parameters and WEPP data, we estimate emissions and power capacity for plants from very detailed information about their capacity, generator age, combustion technology and energy sources. The remaining random error may be large from plant to plant, as we have previously noted, but we would expect these errors to balance as we aggregate from plants to broader geographic units. We test these propositions in several ways. First, at the plant level, we compare CARMA-estimated emissions with observed emissions from matched plants in India, China, the EU and Canada. Then we perform aggregation tests using available benchmarks at three scales: across countries, US states and US counties.

4.1 Matched-Plant Comparisons

Figures 5-7 provide comparative information on observed emissions from power plants in India, China, the EU and Canada. In each case, the figure displays the values of CARMA estimates and observed emissions for matched plants. In the Indian and EU/Canada cases, the observed emissions are from public databases maintained by government agencies. For China, comparative emissions have been assessed by a collaborating Chinese team, using a confidential database maintained by the State

Environmental Protection Agency (SEPA). We have the matched data for India, the EU and Canada in our data files, but we only have the results report from the SEPA team.[15]

Matched sample sizes are 81 for India, 314 for China, and 364 for the EU and Canada. Figures 5-7 are displayed in log units for comparison of results for plants in all size classes. Correlations between CARMA estimates and reported emissions are quite high in all cases: .93 for India, .85 for China and .88 for the EU and Canada. For India and EU/Canada, there are large relative estimation errors for some small plants.[16] In the Chinese case, the large relative errors are more clustered near the center of the size distribution. In all three cases the fit is quite good for the largest plants, with the exception of a few large negative outliers whose capacity factors are far below normal levels for power plants with their characteristics.

4.2 Cross-Country Accuracy

The World Resources Institute (WRI) provides online country-level estimates of CO2 emissions from the power sector.[17] WRI's estimates are principally based on national fuel-use data. For comparison, we aggregate CARMA's plant-level information for the 2004 benchmark year and compare country totals with WRI data for 2003 (the latest available). A regression of the WRI emissions on CARMA emissions for 121 countries yields an extremely good fit, with a regression coefficient of 1.00, a t-statistic of 64.8, and an adjusted R^2 of .97. Figure 8 displays the same information in a scatter diagram.

---

[15] In all cases, it is important to note that reported plant-level emissions are generally calculated by multiplying fuel consumption by standard engineering parameters.

[16] The median percent difference between actual and predicted values for all sample plants is 8.2% for India and 31.3% for EU/Canada.

[17] Both datasets include information for both electricity and heat – WRI explicitly, and CARMA because it includes emissions from combined heat and power plants.

Per our expectation, the random plant-level errors balance almost exactly at the country aggregation level.[18]

4.3 Cross-State Accuracy in the US

For this test, we use data from matched plants for CARMA and the US EPA's eGRID system for 2004. We estimate CO2 emissions for those plants using the full CARMA methodology (capacity utilization and emissions both estimated from the regression models). In the matched dataset, we compute state totals from the CARMA estimates and the eGRID emissions reports. Then we regress the eGRID total on the CARMA total for all 51 states. We obtain a highly-significant regression coefficient of .995 (t-statistic 22.6) and an adjusted R^2 of .91. Figure 9 displays the same information in a scatter diagram, which indicates that random plant-level errors balance very well at the state level.

4.4 Cross-County Accuracy in the US

For this test, we again use data from matched plants for CARMA and the US EPA's eGRID system for 2004. We compute county totals from the CARMA estimates and the eGRID emissions reports. Then we regress the eGRID total on the CARMA total for counties with more than 5 power plants. For those 27 counties, we obtain a highly-significant regression coefficient of 1.06 (t-statistic 6.87) and an adjusted R^2 of .64. Figure 10 displays the same information in a scatter diagram, which indicates that random errors balance well for smaller geographic aggregates that have more than a few power plants.

---

[18] Our calculation of country energy adjustment factors (Section 2) is important in this context. These adjustment factors bring CARMA's country-level power totals into line with US EIA totals. The power adjustment also affects country-level emissions totals, bringing them into closer alignment with the WRI totals.

5. Constructing CARMA's Geographic and Corporate Components

Following the development of plant-level emissions estimates, we added and refined the available data to construct the geographic and corporate totals found in CARMA. On the geographic side, in particular, this exercise resulted in a considerable extension of existing public information on the location of power plants worldwide. In the following sections, we describe that process and add some caveats about the data.

5.1 Assigning Geographic Information to Individual Plants

Almost all of the power plants listed in the U.S. EPA's eGRID database are assigned geographic coordinates (i.e., latitude and longitude), as are many plants listed in similar European public disclosure databases. We have used such data whenever possible to geo-locate the plants in CARMA. These sources provide initial coverage for about 36% of power plants worldwide and 91% of plants in the U.S.

More general geographic data are provided by the WEPP database, including the country where a plant is located (for 100% of plants), state/province (~75%), and city (~70%). More refined, plant-specific spatial data can be purchased from private sources, but because of their proprietary restrictions and CARMA's open-source nature, we chose to extend the geographic coverage ourselves using only publicly-available data.

Our primary goal in extending plant-level geographic data was not to provide precise geographic coordinates for local-scale mapping, but, instead, to allow for regional- and global-scale mapping to reveal broad patterns in the spatial distribution of plants. For this purpose, the data provided by the WEPP database were often sufficient to identify the geographic coordinates of the city where a plant is located.

This was done by employing the "fuzzy match" algorithm suggested by our colleague David Roodman[19] to match the city name listed in the WEPP database to a global database of cities and their geographic coordinates. We used MaxMind's World Cities Database (http://www.maxmind.com/app/worldcities) – containing about 2.7 million records – and merged it with state/province names according to the specified ISO-3166-2 (in the U.S.) or FIPS 10-4 code. Considerable effort was invested in standardizing the spellings of regions in the WEPP data so as to maximize concordance with the FIPS 10-4 region names.

Once these data were sufficiently cleaned, the algorithm was used to identify plants with a potential match on the city, region, *and* country fields of both datasets. Out of the candidate matches, the best-matched as determined by the fuzzy match algorithm were selected and, from that group, only unique matches were retained. The latitude and longitude of the matched city were then assigned to the associated plant(s). In cases where the WEPP data contained city but not state/province information (about 15% of all plants), a match process was carried out on only the city and country fields. Again, only unique matches were retained for reasons of quality assurance. In these cases, the matching process allowed us to identify previously-missing state/province data, thereby extending not just the coordinate coverage but also the more general geographic information.

The overall quality of the matching process depends, in part, on how tightly or loosely the algorithm is set.[20] We experimented with different thresholds, finally settling

[19] See the reference in footnote 6.

[20] Relative tightness in this context is defined by the distance limit specified for the algorithm. The distance between two character strings is the number of character additions or deletions needed to change one into the other. For example, the distance between "hat" and "had" is 2.

on a tiered approach that requires a perfect match on the country name, looser for the state/province field, and looser still for city name. In practice, however, the primary quality-assurance mechanism is the unique match requirement, which effectively eliminates any cases of obvious ambiguity.

In the initial public release of CARMA, the fuzzy match process resulted in the inclusion of city-center geographic coordinates for more than 12,000 plants and added additional state/province information for about 4,000 of those. All told, CARMA's first incarnation contained geographic coordinates for about 60% of all power plants, representing about 80% of all power production and 90% of all power sector CO2 emissions.

Note that because of the common assignment of city-center geographic coordinates rather than plant-specific locations, some plants in CARMA share the same latitude and longitude. This also means that plants formally located in sprawling cities may actually be quite distant from the assigned coordinates. This is not a problem when plotting plants at the regional or global level, but can lead to misplaced accuracy at higher resolution. That said, for larger plants the approximate coordinates can often lead to identification of the actual plant location via CARMA's Google Maps capabilities, as we have discussed online (http://carma.org/blog/the-eye-in-the-sky-carma-identifies-plants-from-above/).

CARMA's quarterly updates provide an opportunity to extend the reach and quality of the database's geographic information over time. As new plants are brought into operation and additional plants are proposed for future construction, we will update the locational information. It should be noted that CARMA's global coordinate mapping of proposed plants is, to our knowledge, unavailable anywhere else.

5.2 Calculating Total Emissions for Geographic Entities

The term "geographic entity" refers to continents, countries, states/provinces, cities, metro areas, counties, and congressional districts (the latter three are U.S. only). For each of these entities, we calculate and publish online the total emissions, power production, and carbon intensity of all plants physically located within its borders. For continents, countries, and states/provinces, this process needs no explanation. For metro areas, counties, and congressional districts within the U.S., we use plant-level zip code data from the WEPP provider to assign plants to the appropriate entity. From there we simply aggregate.

Cities are treated differently. Our primary objective in including city-specific totals was to facilitate the use of CARMA in the identification of emission sources in one's local area. In this spirit, we calculate city totals as the sum of all plants within a 15-mile (~24-km) radius of the city center. This aggregation relies upon the geographic coordinate assignment process detailed in the preceding section. The initial public release of CARMA included entries for more than 26,000 cities worldwide. Subsequent revisions of the data have extended this number to nearly 190,000.

It is obvious – but perhaps important to emphasize – that the geographic totals in CARMA do *not* measure emissions related to the consumption of people within a region. Instead, they measure emissions produced by power plants physically located within the relevant boundary. The total emissions reported for the city of San Francisco, for example, do not reflect emissions produced as a result of electricity use by San Francisco residents but, instead, measure total annual emissions generated by power plants physically located within 15 miles of San Francisco. The goal of CARMA's city-level data is not to provide a carbon "footprint" accounting. The data are designed to help users identify plants or companies of interest in their immediate vicinity.

Despite the considerable extension of plant-specific geographic information as a result of the CARMA project, coverage is still not complete. The published totals may, in some cases, underestimate actual totals for geographic entities smaller than continents or countries, where geographic information for the relevant plants may be unavailable. In total, however, emissions assigned to states/provinces constitute more than 95% of total emissions from the global power sector.

5.3 Country-Specific Geographic Information Coverage

Table 1 shows CARMA's geographic data coverage by country for operating plants in 2007. We report both the proportion of plants and corresponding proportion of total current emissions that have been assigned to a state/province, city, or coordinate location. These figures are calculated from a revision of the CARMA data made subsequent to the initial public launch, but it does not differ much from the original version.

5.4 Treatment of Company Totals and Corporate Hierarchies

For the purpose of compiling aggregate data on companies represented in CARMA, it is helpful to distinguish internally between "operators" – companies that control the day-to-day operations of plants – and "parent" companies that ultimately own operators. While every plant has an operating company, not all have an associated parent company. These data, including information on the headquarters countries of companies, are part of the information provided by the WEPP database. Its supplier monitors the global power sector to keep track of constantly-changing and increasingly-complex ownership hierarchies.

We simply aggregate emissions, power, and intensity data for all companies. In most cases, this is a straightforward process and one easily interpreted on the CARMA site. The only area of potential confusion is the inclusion of companies that are both operators and parents. For example, AES Corp. both owns plants through subsidiaries and operates plants itself. In the original version of CARMA, we effectively treated AES Corp. (operator) as distinct from AES Corp. (parent). In subsequent versions, we have collapsed the data for simplicity. In this approach, the totals presented for a given company name will always represent the highest aggregated totals (i.e. the "parent" totals in the original approach).

6. Summary and Conclusions

In this paper, we have provided a description of CARMA's methodology, an assessment of its strengths and weaknesses, and some tests of its accuracy across countries and at different geographical scales. Although it is clear that CARMA performs well, we recognize that neither CARMA nor any other carbon emissions database will ever be perfect. Our operating philosophy is maximum possible transparency, with a standing offer to all power plants and power companies in our database: We have done our best to represent their emissions fairly and accurately, but they can immediately improve our estimates if they think revisions are warranted. All they have to do is provide us with third-party-verified data on their CO2 emissions and power generation, or the types and quantities of fuels consumed if they do not have emissions data. We can do the rest with standard engineering parameters.

Although we make no pretense of perfection, we want to assure the concerned global citizens who visit CARMA that they have the best information we can provide.

We will constantly update our database, improve our methodology, and rapidly post revised information. We encourage any company listed in CARMA to provide us with better data if they dispute our estimates. We will be happy to incorporate corrections right away, under two conditions: (1) The company provides the information for all of its plants; (2) All the data are vetted by an independent auditor with recognized competence.

We hope that many companies will come forward, particularly those whose plant-level emissions are not disclosed by government-operated public disclosure sites (we use that information already). Since global regulation will ultimately happen anyway, why not get ahead of the process? If such audited disclosure becomes universal, we will have the database we need for global regulation of carbon emissions from the power sector. At that point (or much sooner, if asked), we will be happy to turn CARMA over to the international agency that has been delegated the task of regulating carbon emissions.

Figure 1: Actual vs. Predicted CO2 Emissions

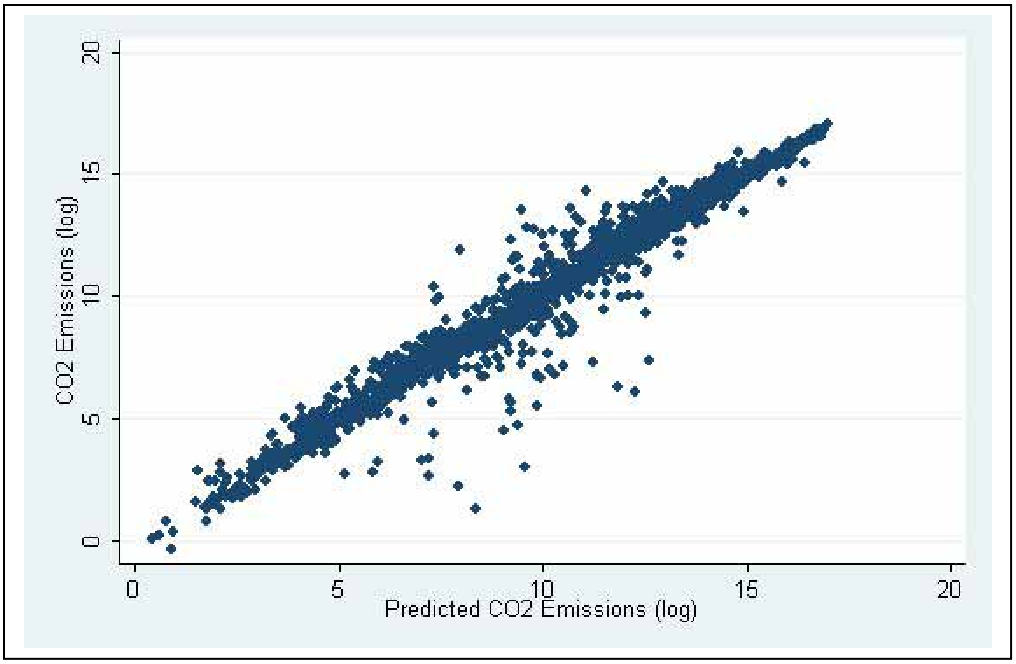
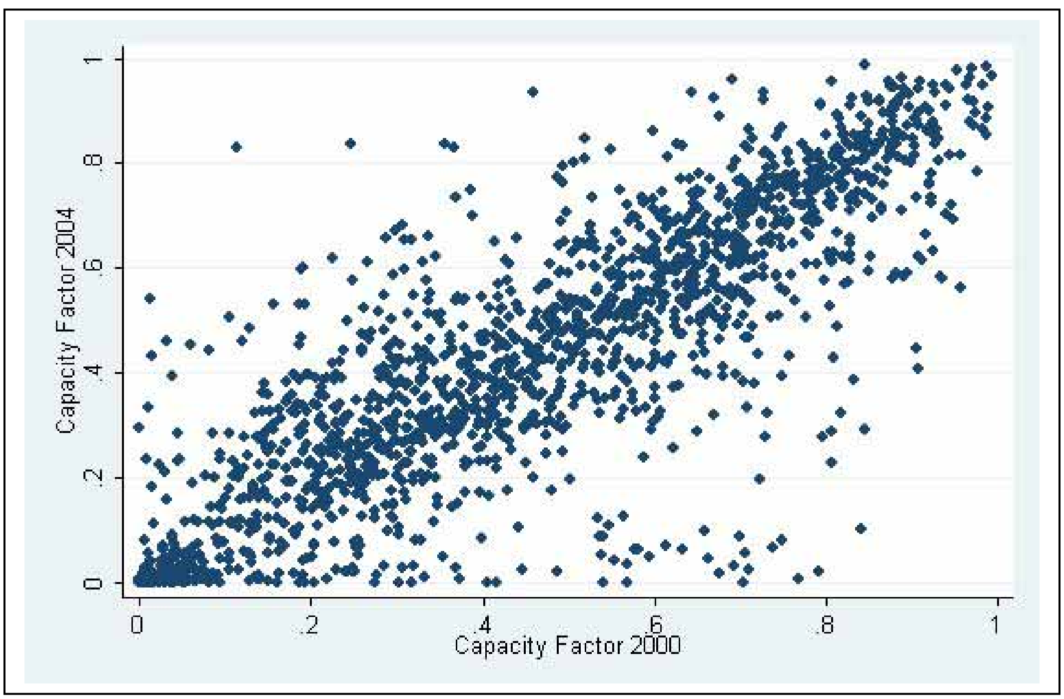
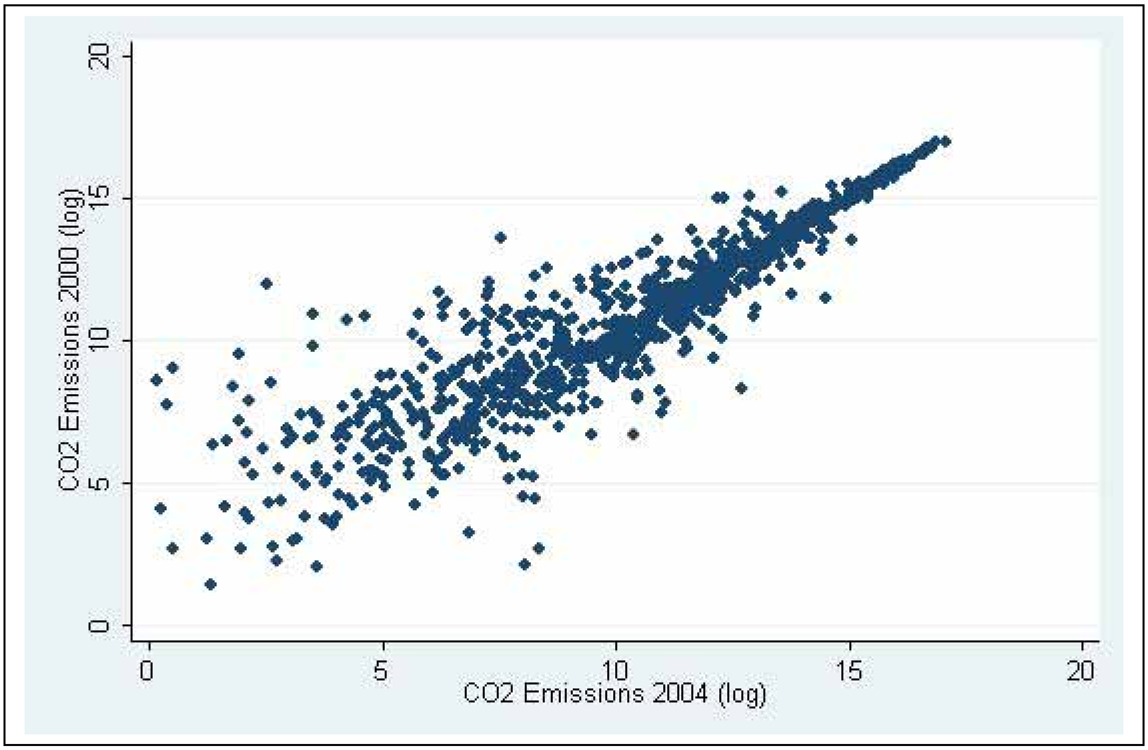


Figure 2: Capacity Factors for Identical US Power Plants, 2000 and 2004



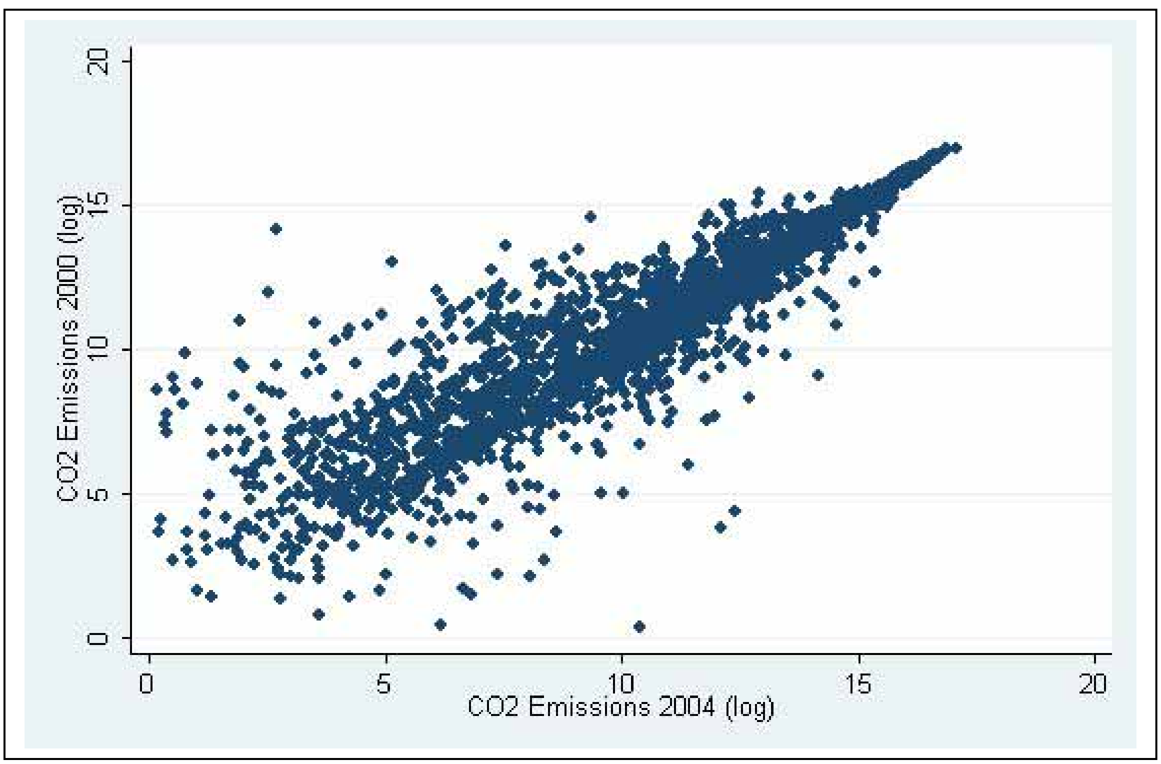
Source: US EPA, eGRID Database (2000,2007)

Figure 3: CO_2 Emissions, 2000 and 2004 (Matched Plants with Identical Capacities)



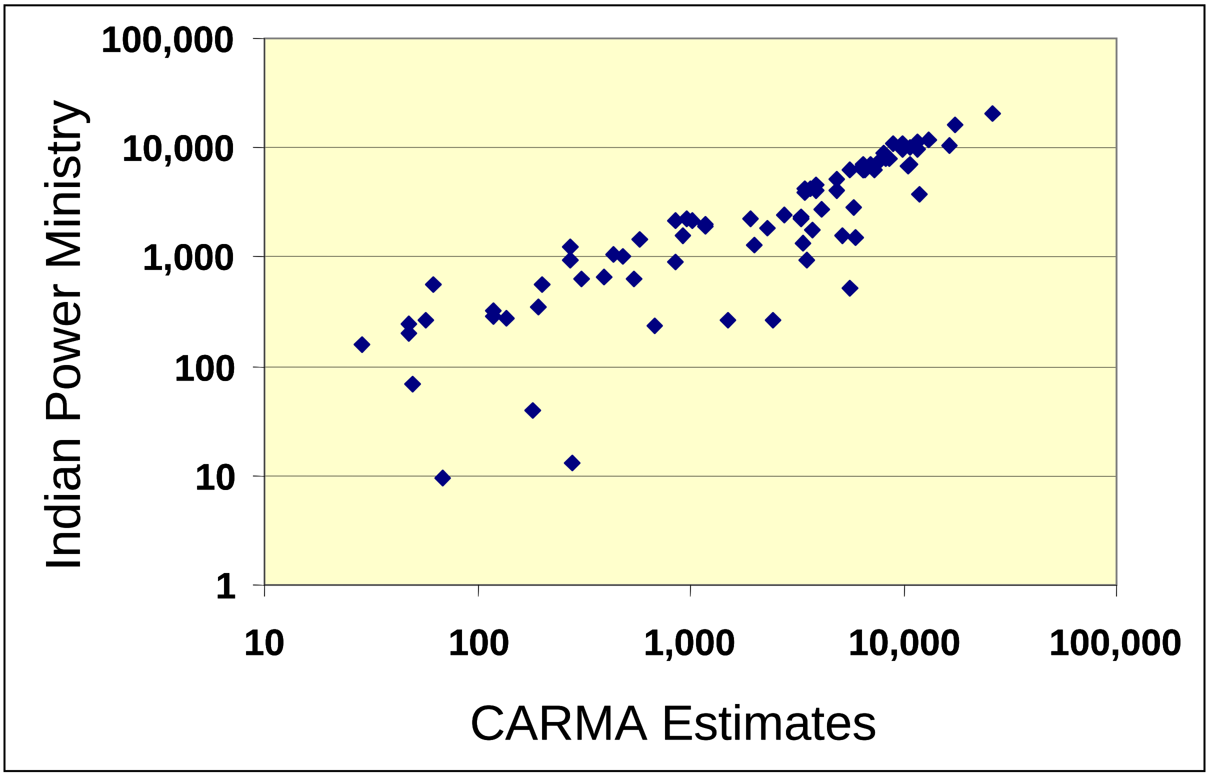
Source: US EPA, eGRID Database (2000,2007)

Figure 4: CO_2 Emissions, 2000 and 2004 (Matched Plants)



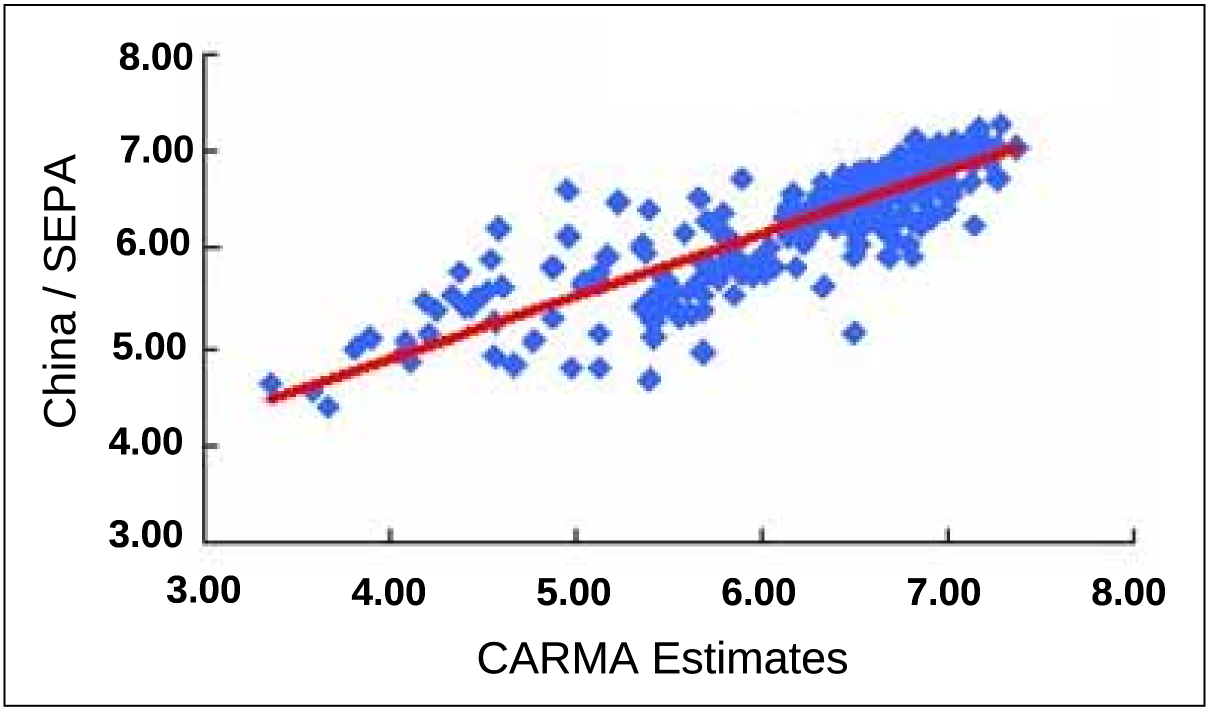
Source: US EPA, eGRID Database (2000,2007)

Figure 5: Reported vs. Estimated CO2 Emissions ('000 tons), India (ρ = .93)



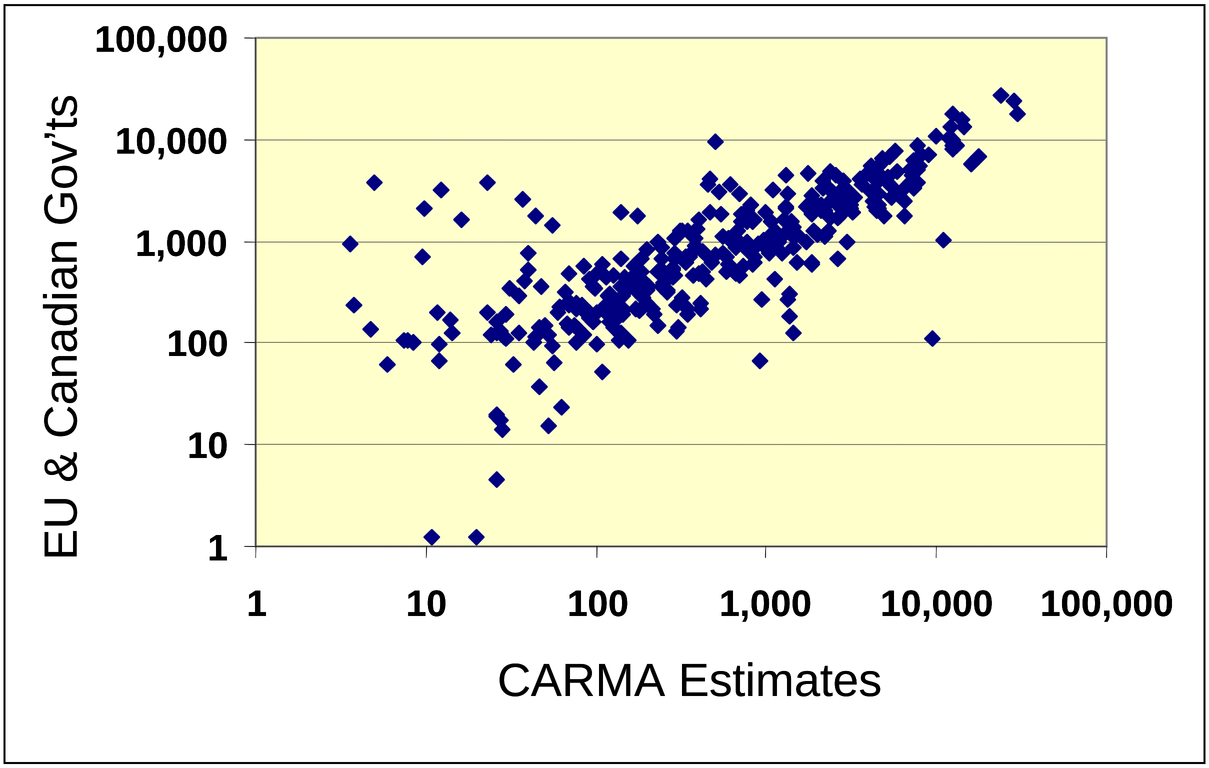
Sources: Indian Power Ministry

Figure 6: Reported vs. Estimated CO2 Emissions (logs), China (ρ = .85)



Source: China State Environmental Protection Administration

Figure 7: Reported vs. Estimated CO2 Emissions ('000 tons), EU/Canada (ρ = .88)



Sources: European Environment Agency and Environment Canada

Figure 8: Country-Level CO2 Emissions From the Power Sector (Million Tons)

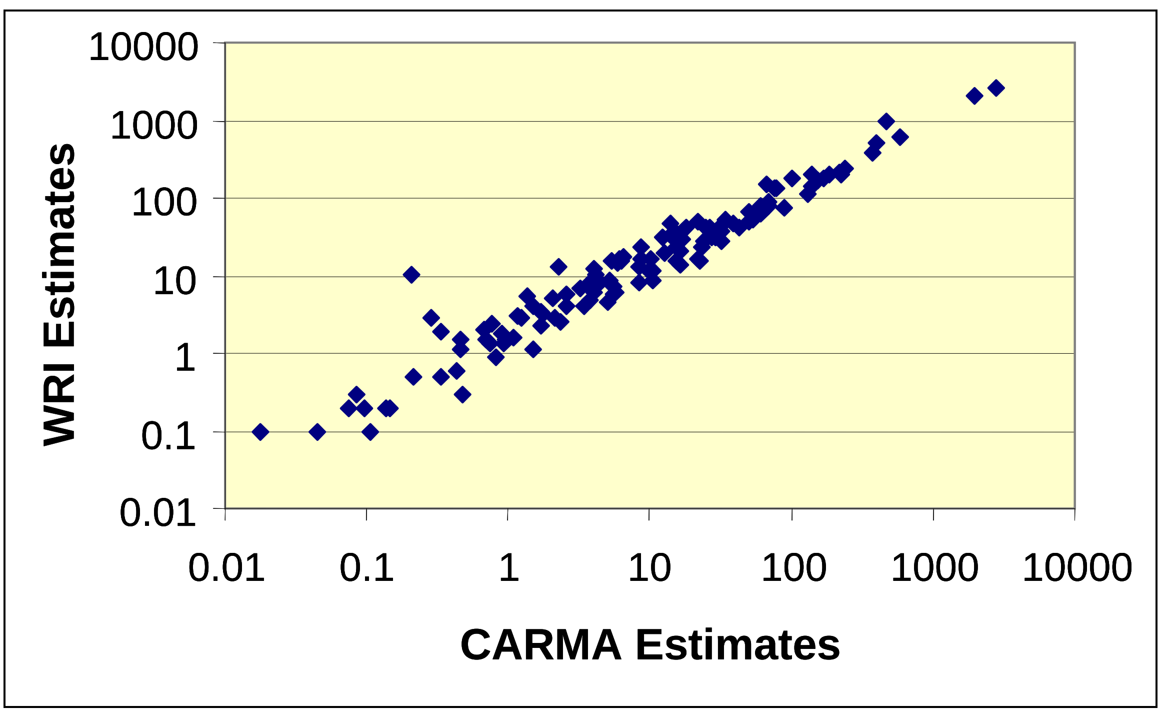


Figure 9: US State-Level CO2 Emissions ('000 tons), 2004 (log scale)

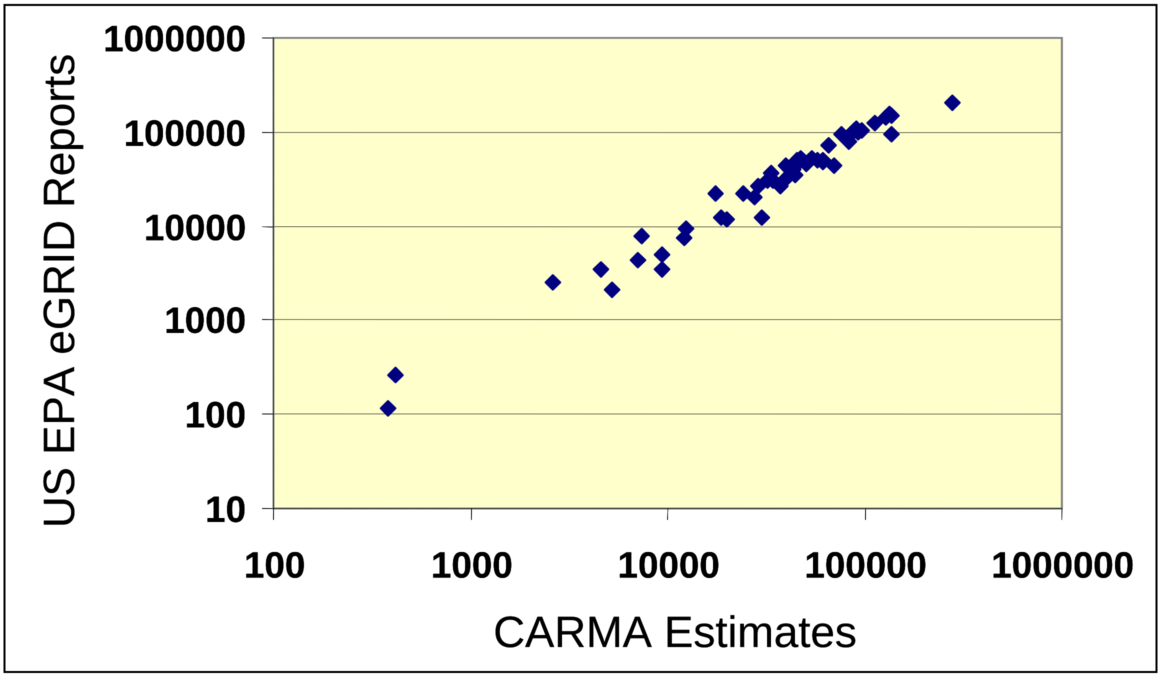


Figure 10: US County-Level CO2 Emissions ('000 tons), 2004 (log scale)

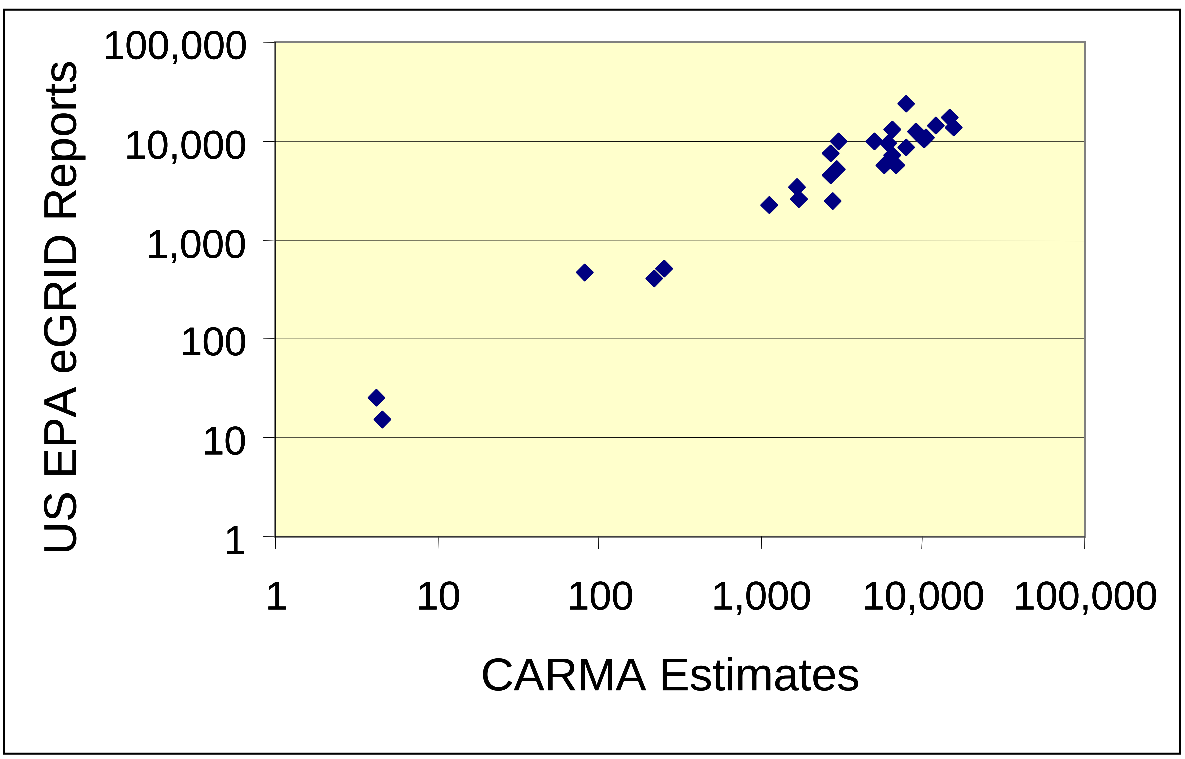


Table 1: Geographic Coverage at the Plant Level in CARMA

	PROPORTION OF PLANTS			PROPORTION OF TOTAL EMISSIONS		
Country	**State / Province**	**City**	**Lat/Long**	**State / Province**	**City**	**Lat/Long**
Afghanistan	0.82	0.82	0.53	0.85	1.00	0.85
Albania	0.64	0.46	0.36	0.61	0.79	0.79
Algeria	0.52	0.91	1.00	0.70	1.00	1.00
Angola	0.89	0.89	0.98	0.99	0.97	1.00
Antigua & Barbuda	0.17	0.50	0.17	0.30	0.65	0.30
Argentina	0.95	0.61	0.56	1.00	0.95	0.88
Armenia	1.00	0.84	0.78	1.00	1.00	1.00
Australia	1.00	0.86	0.64	1.00	1.00	0.83
Austria	0.79	0.55	0.44	0.99	1.00	0.98
Azerbaijan	0.64	0.64	0.46	0.87	1.00	0.86
Bahamas	0.89	0.21	0.14	1.00	0.71	0.92
Bahrain	0.81	0.75	0.44	0.88	0.98	0.50
Bangladesh	0.88	0.90	0.43	1.00	1.00	0.51
Barbados	0.00	0.83	0.00	0.00	1.00	0.00
Belarus	0.91	0.94	0.63	1.00	1.00	0.86
Belgium	0.83	0.92	0.77	0.96	0.96	0.94
Belize	1.00	0.94	0.35	1.00	0.85	0.65
Benin	0.17	1.00	0.83	0.06	1.00	0.94
Bermuda	0.75	0.75	0.75	0.99	0.99	0.99
Bhutan	0.98	0.29	0.11	0.94	0.20	0.14
Bolivia	0.69	0.62	0.42	0.99	0.88	0.87
Bosnia-Herzegovina	0.96	0.76	0.54	1.00	1.00	0.96
Botswana	0.60	1.00	1.00	1.00	1.00	1.00
Brazil	0.99	0.89	0.74	1.00	1.00	0.64
Brunei	0.88	0.88	0.00	0.99	0.99	0.00
Bulgaria	0.97	0.71	0.60	1.00	1.00	1.00
Burkina Faso	0.13	0.30	1.00	0.70	0.71	1.00
Burundi	0.48	0.84	1.00	0.92	1.00	1.00
Cambodia	0.97	0.94	0.38	0.98	1.00	0.55
Cameroon	0.87	0.87	0.93	1.00	1.00	1.00
Canada	1.00	0.80	0.64	1.00	1.00	0.94
Cape Verde	1.00	0.69	1.00	1.00	0.76	1.00
Cayman Islands	1.00	0.50	0.00	1.00	0.98	0.00
Central African Republic	0.56	1.00	1.00	0.19	1.00	1.00
Chad	0.80	1.00	1.00	1.00	1.00	1.00
Chile	0.92	0.70	0.51	1.00	0.99	0.69
China	0.98	0.78	0.66	1.00	0.96	0.88
Colombia	0.96	0.79	0.66	0.99	0.99	0.94

	PROPORTION OF PLANTS			PROPORTION OF TOTAL EMISSIONS		
Country	**State / Province**	**City**	**Lat/Long**	**State / Province**	**City**	**Lat/Long**
Comoros	1.00	0.50	1.00	1.00	0.72	1.00
Congo	0.69	0.69	1.00	0.59	0.69	1.00
Congo Republic	0.89	0.89	1.00	0.52	0.52	1.00
Costa Rica	0.98	0.98	0.60	1.00	1.00	1.00
Cote D'Ivoire	0.42	0.75	1.00	0.53	1.00	1.00
Croatia	0.95	0.89	0.68	0.94	1.00	0.94
Cuba	0.53	0.67	0.40	0.84	0.85	0.75
Cyprus	0.71	1.00	0.71	0.98	1.00	0.98
Czech Republic	0.57	0.84	0.52	0.95	1.00	0.89
Denmark	0.61	0.86	0.68	0.68	0.99	0.69
Djibouti	1.00	1.00	1.00	1.00	1.00	1.00
Dominica	0.33	0.33	0.33	1.00	1.00	1.00
Dominican Republic	0.79	0.89	0.68	0.53	0.90	0.67
East Timor	0.00	1.00	0.00	0.00	1.00	0.00
Ecuador	0.69	0.54	0.38	0.78	0.91	0.72
Egypt	0.93	0.98	1.00	1.00	1.00	1.00
El Salvador	0.69	0.73	0.50	0.99	0.99	0.85
Equatorial Guinea	1.00	0.80	1.00	1.00	1.00	1.00
Eritrea	0.17	1.00	1.00	0.89	1.00	1.00
Estonia	1.00	0.86	0.49	1.00	1.00	0.99
Ethiopia	0.84	0.91	1.00	0.79	1.00	1.00
Faroe Islands	0.09	0.18	0.00	0.00	0.84	0.00
Fiji	0.86	0.78	0.51	0.95	1.00	0.97
Finland	0.59	0.64	0.50	0.94	1.00	0.96
France	0.63	0.69	0.45	0.64	0.97	0.51
French Polynesia	1.00	0.41	0.07	1.00	0.94	0.29
Gabon	0.63	0.97	0.97	0.45	0.89	1.00
Gambia	1.00	1.00	1.00	1.00	1.00	1.00
Georgia	0.62	0.93	0.75	0.93	1.00	0.93
Germany	0.76	0.74	0.59	0.94	0.99	0.90
Ghana	0.71	0.86	0.93	1.00	1.00	1.00
Greece	0.72	0.64	0.45	0.78	1.00	0.98
Grenada	0.00	1.00	0.00	0.00	1.00	0.00
Guatemala	0.93	0.91	0.73	1.00	1.00	0.90
Guinea	0.50	0.57	0.86	0.82	0.90	0.87
Guinea-Bissau	0.50	1.00	1.00	0.92	1.00	1.00
Guyana	0.25	0.75	0.30	0.11	0.48	0.11
Haiti	0.88	0.88	0.81	1.00	1.00	0.99
Honduras	1.00	0.88	0.83	1.00	1.00	1.00
Hungary	0.86	0.96	0.88	0.96	1.00	1.00

	PROPORTION OF PLANTS			PROPORTION OF TOTAL EMISSIONS		
Country	**State / Province**	**City**	**Lat/Long**	**State / Province**	**City**	**Lat/Long**
Iceland	0.03	0.28	0.05	0.00	0.00	0.00
India	0.99	0.79	0.60	1.00	0.98	0.83
Indonesia	0.58	0.55	0.35	0.99	0.99	0.81
Iran	0.91	0.83	0.64	0.99	0.99	0.76
Iraq	0.86	0.90	0.66	0.99	1.00	0.74
Ireland	1.00	0.94	0.62	1.00	1.00	0.94
Isle Of Man	0.00	1.00	0.40	0.00	1.00	0.08
Israel	0.62	0.79	0.54	0.95	1.00	0.99
Italy	0.77	0.75	0.49	0.96	0.99	0.81
Jamaica	0.97	0.72	0.59	1.00	1.00	0.98
Japan	0.81	0.38	0.29	0.99	0.92	0.83
Jordan	0.25	0.86	0.61	0.45	1.00	0.91
Kazakhstan	0.89	0.80	0.62	1.00	0.97	0.84
Kenya	0.39	0.65	1.00	0.83	0.98	1.00
Kiribati	0.25	0.75	0.25	0.02	0.98	0.73
Kuwait	0.47	0.76	0.47	0.49	1.00	0.49
Kyrgyzstan	0.53	0.27	0.27	1.00	1.00	1.00
Laos	0.89	0.80	0.31	1.00	0.95	0.12
Latvia	0.68	0.82	0.77	0.35	1.00	1.00
Lebanon	0.44	0.94	0.44	0.42	1.00	0.18
Lesotho	0.50	0.00	1.00	0.29	0.00	1.00
Liberia	0.00	1.00	1.00	0.00	1.00	1.00
Libya	0.27	0.63	1.00	0.43	0.96	1.00
Liechtenstein	0.75	0.75	0.75	0.56	0.56	0.56
Lithuania	0.56	0.78	0.56	0.75	1.00	0.75
Luxembourg	0.77	0.90	0.77	0.19	0.97	0.93
Macedonia	0.30	0.35	0.30	0.98	1.00	0.98
Madagascar	0.48	0.90	0.80	0.06	0.56	0.99
Malawi	0.67	0.93	1.00	0.78	1.00	1.00
Malaysia	0.97	0.79	0.59	1.00	0.99	0.82
Maldives	0.03	0.63	0.03	0.00	0.96	0.00
Mali	0.68	1.00	1.00	0.46	1.00	1.00
Malta	0.00	0.80	0.40	0.00	0.99	0.42
Marshall Islands	0.00	1.00	0.00	0.00	1.00	0.00
Mauritania	0.71	1.00	1.00	0.64	1.00	1.00
Mauritius	0.77	0.68	0.97	1.00	1.00	1.00
Mayotte	1.00	1.00	1.00	1.00	1.00	1.00
Mexico	0.99	0.98	0.84	1.00	1.00	0.96
Micronesia	0.00	0.90	0.20	0.00	1.00	0.65
Moldova	0.87	0.93	0.67	0.98	0.99	0.97

	PROPORTION OF PLANTS			PROPORTION OF TOTAL EMISSIONS		
Country	**State / Province**	**City**	**Lat/Long**	**State / Province**	**City**	**Lat/Long**
Monaco	0.50	1.00	0.00	0.00	1.00	0.00
Mongolia	0.78	0.89	0.48	1.00	1.00	0.90
Montenegro	0.00	0.30	0.10	0.00	1.00	1.00
Morocco	0.98	0.92	1.00	1.00	1.00	1.00
Mozambique	0.32	1.00	1.00	0.05	1.00	1.00
Myanmar	0.95	0.53	0.28	1.00	0.95	0.23
Namibia	0.67	1.00	1.00	1.00	1.00	1.00
Nauru	0.00	1.00	0.00	0.00	1.00	0.00
Nepal	0.94	0.95	0.02	1.00	1.00	0.00
Netherlands	0.64	0.73	0.62	0.82	1.00	0.95
New Caledonia	0.33	0.47	0.13	0.01	0.99	0.22
New Zealand	0.30	0.63	0.34	0.03	1.00	0.88
Nicaragua	0.48	0.69	0.34	0.52	1.00	0.74
Niger	0.50	1.00	1.00	0.04	1.00	1.00
Nigeria	0.92	0.81	0.94	1.00	0.99	0.99
North Korea	0.78	0.70	0.34	0.78	0.78	0.19
Norway	0.76	0.46	0.29	0.99	0.91	0.70
Oman	0.91	0.91	0.52	0.97	1.00	0.35
Pakistan	0.98	0.92	0.75	1.00	1.00	0.82
Palestine	0.00	1.00	0.00	0.00	1.00	0.00
Panama	0.92	0.80	0.61	0.87	0.99	0.85
Papua New Guinea	0.31	0.53	0.23	0.40	0.42	0.18
Paraguay	0.30	0.70	0.20	0.05	0.94	0.02
Peru	0.99	0.99	0.83	1.00	1.00	0.94
Philippines	0.85	0.70	0.46	0.99	0.98	0.90
Poland	0.59	0.87	0.62	0.83	1.00	0.91
Portugal	0.84	0.76	0.41	0.94	0.99	0.94
Qatar	0.28	0.94	0.33	0.58	1.00	0.63
Romania	0.61	0.16	0.13	0.98	0.98	0.88
Russia	0.99	0.88	0.76	1.00	0.99	0.90
Rwanda	0.50	0.75	0.92	1.00	0.00	1.00
Samoa	0.80	0.30	0.00	1.00	0.74	0.00
Sao Tome & Principe	1.00	1.00	1.00	1.00	1.00	1.00
Saudi Arabia	0.57	0.92	0.54	0.87	0.98	0.74
Senegal	0.83	0.96	0.65	0.96	1.00	0.95
Serbia	0.56	0.61	0.11	0.99	1.00	0.75
Seychelles	1.00	1.00	1.00	1.00	1.00	1.00
Sierra Leone	0.25	0.63	1.00	0.82	1.00	1.00
Singapore	0.02	0.33	0.04	0.00	0.35	0.00
Slovakia	0.62	0.77	0.53	0.93	1.00	0.93

	PROPORTION OF PLANTS			PROPORTION OF TOTAL EMISSIONS		
Country	**State / Province**	**City**	**Lat/Long**	**State / Province**	**City**	**Lat/Long**
Slovenia	0.35	0.38	0.34	0.99	1.00	1.00
Solomon Islands	0.89	0.83	0.44	1.00	0.98	0.81
Somalia	0.71	0.86	1.00	1.00	1.00	1.00
South Africa	0.82	0.83	0.99	1.00	1.00	1.00
South Korea	0.77	0.76	0.29	1.00	1.00	0.81
Spain	0.86	0.79	0.59	0.99	0.99	0.75
Sri Lanka	0.98	0.83	0.12	1.00	0.99	0.58
St Kitts & Nevis	1.00	0.50	0.50	1.00	0.14	0.14
St Lucia	1.00	1.00	1.00	1.00	1.00	1.00
St Vincent & Grenadines	0.75	0.50	0.00	1.00	0.64	0.00
Sudan	0.76	0.97	0.94	0.49	1.00	0.99
Suriname	0.47	0.79	0.47	0.75	1.00	0.75
Swaziland	0.50	0.70	1.00	0.04	0.18	1.00
Sweden	0.62	0.74	0.46	0.67	0.85	0.92
Switzerland	0.83	0.57	0.45	0.78	0.82	0.73
Syria	0.57	0.69	0.57	0.73	0.97	0.82
Taiwan (China)	0.00	0.64	0.37	0.00	0.94	0.78
Tajikistan	0.85	0.62	0.54	1.00	1.00	1.00
Tanzania	0.97	0.99	0.99	0.99	1.00	0.99
Thailand	0.78	0.77	0.41	0.99	0.99	0.66
Togo	0.00	0.80	1.00	0.00	1.00	1.00
Tonga	1.00	0.25	0.00	1.00	0.63	0.00
Trinidad & Tobago	0.50	0.88	0.50	0.38	1.00	0.76
Tunisia	0.58	0.77	1.00	0.84	0.94	1.00
Turkey	0.99	0.61	0.50	1.00	0.99	0.76
Turkmenistan	1.00	0.92	0.77	1.00	1.00	0.90
Tuvalu	0.33	0.67	0.00	0.85	0.15	0.00
Uganda	0.32	1.00	0.95	0.52	1.00	0.56
Ukraine	0.91	0.91	0.54	0.96	1.00	0.82
United Arab Emirates	1.00	0.80	0.49	1.00	0.94	0.55
United Kingdom	0.96	0.84	0.60	1.00	1.00	0.77
United States	1.00	0.96	0.89	1.00	1.00	1.00
Uruguay	1.00	0.75	0.63	1.00	1.00	1.00
Uzbekistan	0.81	0.52	0.45	0.96	1.00	1.00
Vanuatu	0.80	0.60	0.40	0.63	0.94	0.57
Venezuela	0.97	0.93	0.82	1.00	1.00	0.94
Vietnam	0.97	0.86	0.36	1.00	0.98	0.54
Yemen	0.69	0.72	0.13	0.86	0.87	0.27
Zambia	1.00	0.83	1.00	1.00	1.00	1.00
Zimbabwe	0.65	0.88	1.00	0.28	1.00	1.00

	PROPORTION OF PLANTS			PROPORTION OF TOTAL EMISSIONS		
Country	**State / Province**	**City**	**Lat/Long**	**State / Province**	**City**	**Lat/Long**
World	**0.83**	**0.76**	**0.61**	**0.96**	**0.98**	**0.89**

Appendix A

The core regression models for CARMA estimation are based on observations of capacity factors and CO_2 emissions for over 2,000 small, medium and large power plants in the US. As we note in the paper, the regression models are fitted to information from the WEPP database on fuels, capacity, equipment age, and detailed specifications of combustion technologies. Table A3 reports regression results for capacity factors and CO_2 emissions. The overall fit for the CO_2 emissions regression with 2,166 observations is very good, with an R^2 of .96. Random variation in capacity factors, which we discuss in the paper, produces a significantly lower R^2 (.53) for 2,296 observations.

Fuels

We aggregate the fuels detailed in Table A1 into six categories: coal, gas, oil, other gases, other liquids and other solids. WEPP data are provided at the generator level; we use capacity-weighted shares for the six fuel categories, dividing coal into supercritical and subcritical (we exclude other solids to prevent perfect collinearity; the result for this variable is absorbed in the regression constant). In both regressions, fuel shares have generally high levels of significance. The results are in line with our prior expectations: Capacity factors are largest for supercritical coal, ceteris paribus, followed by subcritical coal, other gases, other liquids, gas and oil. CO_2 emissions intensities are highest for subcritical coal, followed by supercritical coal, other liquids, oil, gas, and other gases.

Scale

Both scale-related variables are highly significant. Capacity factors increase with potential power (based on plant capacity). The CO_2 emissions regression reveals significant scale economies (CO_2 emissions and power produced are in log form): Emissions increase .917% for each 1% increase in power produced, ceteris paribus. We also interact fuels with log power produced for all fuel groups, and find significance only for interactions with gas and other gases.

Vintage

We calculate weighted average generator age using dates of installation and generator capacities in each plant. Capacity factors are positively and significantly related to average generator age. This variable is not significant in the emissions equation

Combustion Technologies

Table A2 displays the combustion technology information provided by the WEPP database. With so many degrees of freedom available, we have been able to fit regressions to the full set of technologies, fuel/technology interactions and scale/technology interactions. Most of these variables are not significant. We include the significant results for technologies, technology/fuel interactions, and technology/scale interactions in Table A3. The core regressions for CARMA include only these significant variables.

Table A1: Regression Fuels and Fuel Groups

Group	Code	Name
COAL	COAL	Coal
GAS	GAS	Natural gas
OIL	GASOIL	Gasoil (intermediate refining product also called No 2 fuel oil)
OIL	OIL	Fuel oil
OIL	SHALE	Oil shale
OIL	WSTOIL	Waste oil
OTHGAS	BFG	Blast-furnace gas (approximately 10% of the heat content of pipeline natural gas)
OTHGAS	BGAS	Biogas from digestion of sewage sludge or agricultural waste or food waste or other organic material
OTHGAS	CGAS	Coal syngas (from coal gasification)
OTHGAS	COG	Coke oven gas (approximately 50% of the heat content of pipeline natural gas)
OTHGAS	FGAS	Flare gas or wellhead gas
OTHGAS	LGAS	Landfill gas
OTHGAS	LNG	Liquified natural gas
OTHGAS	METH	Coal-bed methane aka CBM
OTHGAS	MGAS	Mine gas (low-BTU waste gas or methane from coal mines)
OTHGAS	OGAS	Gasified crude oil or refinery bottoms or bitumen
OTHGAS	PETGAS	Petroleum coke synthetic gas
OTHGAS	REFGAS	Syngas from gasified refuse
OTHGAS	RGAS	Refinery off-gas
OTHGAS	TGAS	Top gas
OTHGAS	WOODGAS	Syngas from gasified wood
OTHGAS	WSTGAS	Waste gas from refinery or other industrial processes
OTHLIQ	BL	Bio-derived liquid fuel or biodiesel or bio-oil
OTHLIQ	CWM	Coal-water mixture
OTHLIQ	ETHANOL	Ethanol
OTHLIQ	JET	Jet fuel
OTHLIQ	KERO	Kerosene
OTHLIQ	LIQ	Pulping liquor (black liquor)
OTHLIQ	LPG	Liquified petroleum gas (usually butane or propane)
OTHLIQ	NAP	Naphtha
OTHLIQ	ORI	Orimulsion (emulsified bitumen)
OTHLIQ	RWST	Refinery wastes
OTHLIQ	WSTWSL	Wastewater sludge
OTHSOL	BITUMEN	Bitumen or asphalt
OTHSOL	COKE	Petroleum coke
OTHSOL	HZDWST	Hazardous Waste
OTHSOL	INDWST	Industrial Waste
OTHSOL	MANURE	Manure fuel
OTHSOL	MBM	Meat and bonemeal
OTHSOL	PWST	Paper mill waste or sludges
OTHSOL	REF	Refuse (unprocessed municipal solid waste)
OTHSOL	TIRES	Scrap tires
OTHSOL	TSAND	Tar sands
OTHSOL	WOOD	Wood or wood-waste fuel

Appendix Table A2: Fuel Combustion Technologies

Code	Name
CC	Combined-cycle
CCSS	Combined-cycle single shaft configuration
GT	Gas/combustion turbine
GT/C	Gas turbine in combined-cycle
GT/H	Gas turbine with heat recovery
GT/R	Gas turbine used for partial or complete steam-turbine repowering
GT/S	Gas turbine with steam sendout
GT/T	Gas turbine in topping configuration with existing conventional boiler and T/G set
IC	Internal combustion (reciprocating engine or diesel engine)
IC/C	Internal combustion engine in combined-cycle
IC/H	Internal combustion engine with heat recovery
RSE	Reciprocating steam engine
ST	Steam turbine
ST/C	Steam turbine in combined-cycle
ST/S	Steam turbine with steam sendout
SUBCR	Subcritical
SUPERC	Supercritical
TEX	Turbo expander/gas expander

Table A3: Regression Results

	Capacity Factor	Log CO2 Emissions
Gas	-0.296	1.773
	(9.80)**	(12.20)**
Oil	-0.339	1.823
	(10.41)**	(16.25)**
Supercritical Coal	0.146	2.842
	(3.15)**	(19.13)**
Subcritical Coal	0.057	2.959
	(1.98)*	(29.52)**
Other Gases	-0.173	0.557
	(3.70)**	(1.33)
Other Liquids	-0.230	2.157
	(5.33)**	(14.31)**
Potential Power	0.007	
[Capacity*24*365]	(5.58)**	
Log Age-Weighted Capacity	0.021	
	(3.31)**	
Standby Share of Capacity	-0.113	
	(6.36)**	
Combined-Cycle Share	0.238	
	(14.91)**	
Steam Turbine Share	0.020	
	(1.14)	
Steam Sendout Share	0.298	
	(20.77)**	
Heat Recovery Share	0.236	
	(7.58)**	
Log Produced Power		0.917
		(113.01)**
Gas Turbine Share		0.204
		(2.63)**
Gas Turbine Combined Cycle Share		0.292
		(4.22)**
Gas Turbine Steam Sendout Share		0.154
		(1.85)
Gas With Steam Turbine Share		0.552
		(4.31)**
Oil With Gas Turbine		0.405
		(5.24)**
Oil With Steam Turbine		0.964
		(7.40)**
Gas x Log Power Produced		0.013
		(2.50)*
Gas With Steam Turbine x Log Power Produced		-0.010
		(2.01)*
Other Gases x Log Power Produced		0.046
		(2.25)*
Constant	0.337	-1.166
	(10.02)**	(5.61)**
Observations	2296	2166
R-squared	0.53	0.96

Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%