MODULATING EMISSIONS FROM ELECTRIC GENERATING UNITS AS A FUNCTION OF METEOROLOGICAL VARIABLES

William Battye*
EC/R Incorporated, Chapel Hill, NC

William Warren-Hicks, Ph.D.
EcoStat, Inc., Mebane, NC

Steve Fudge
EC/R Incorporated, Chapel Hill, NC

Thomas Pierce+
Air Resources Laboratory, National Oceanic and Atmospheric Administration, Research Triangle Park, NC

1. INTRODUCTION

The National Air Quality Forecast System is being developed by the National Oceanic and Atmospheric Administration (NOAA) and the U.S. Environmental Protection Agency (EPA) to provide air quality forecasts for photochemical ozone (smog), fine particulate matter ($PM_{2.5}$) and other pollutants. NOAA is currently using the Community Multiscale Air Quality (CMAQ) model coupled with the Eta meteorological model to predict photochemical ozone concentrations, and plans to extend the model to include $PM_{2.5}$ in the near future. CMAQ requires hourly emissions estimates of pollutants from each stack (as well as from nonpoint and mobile sources). These hourly estimates are generally developed using temporal allocation factors, which reflect measured or inferred variations in emissions due to seasonal, weekly, and diurnal variations. However, these temporal patterns are subject to considerable uncertainty. In particular, emissions from electric generating units (EGUs) are believed to vary depending on short-term demands for electricity. For instance, increased use of air conditioning on hot summer days is expected to cause increases in electricity demand, and consequent increases in EGU emissions. EGUs are an important source of emissions of nitrogen oxides (NO_x), which react with volatile organic compounds (VOC) in the presence of sunlight to form ozone. NO_x and sulfur dioxide (SO_2) emissions from EGUs also react to produce $PM_{2.5}$.

*Corresponding Author: William Battye, EC/R Incorporated, 6330 Quadrangle Drive, Chapel Hill, NC 27517; e-mail: battye.bill@ecrweb.com
+In partnership with the U.S. Environmental Protection Agency's National Exposure Research Laboratory.

The U.S. EPA Clean Air Markets Division has assembled continuous emissions monitoring systems (CEMS) data from the U.S. Department of Energy for more than 5000 EGUs as part of EPA's Acid Rain and NO_x Budget Programs (EPA, 2005). These data are currently being used to provide hourly emissions data for CMAQ model evaluation studies, and to calculate retrospective seasonal average emission rates for other CMAQ modeling efforts. The purpose of the current effort is to draw on CEMS data to relate EGU NO_x emissions to temperature and other meteorological variables. These relationships will be used in the National Air Quality Forecast System to adjust for the influence of changing meteorological conditions on daily EGU emissions.

2. METHODS

Autoregressive time-series models were used to relate EGU NO_x emissions to meteorological parameters and to day type (weekday and weekend day or holiday). These models were developed at various levels of geographic resolution. A multi-step approach was used to build the models, relating regional-specific averaged daily NO_x emissions to regional-specific averaged meteorological variables.

2.1 Processing and Analysis of Emissions Data

Preprocessed CEMS data files were obtained from the EPA for calendar years 2002 and 2003. The CEMS data include the measured hourly NO_x emissions, SO_2 emissions, and boiler heat input for each boiler. The files are formatted for input to the Sparse Matrix Kernel Emissions (SMOKE) modeling system used with CMAQ; and EGU

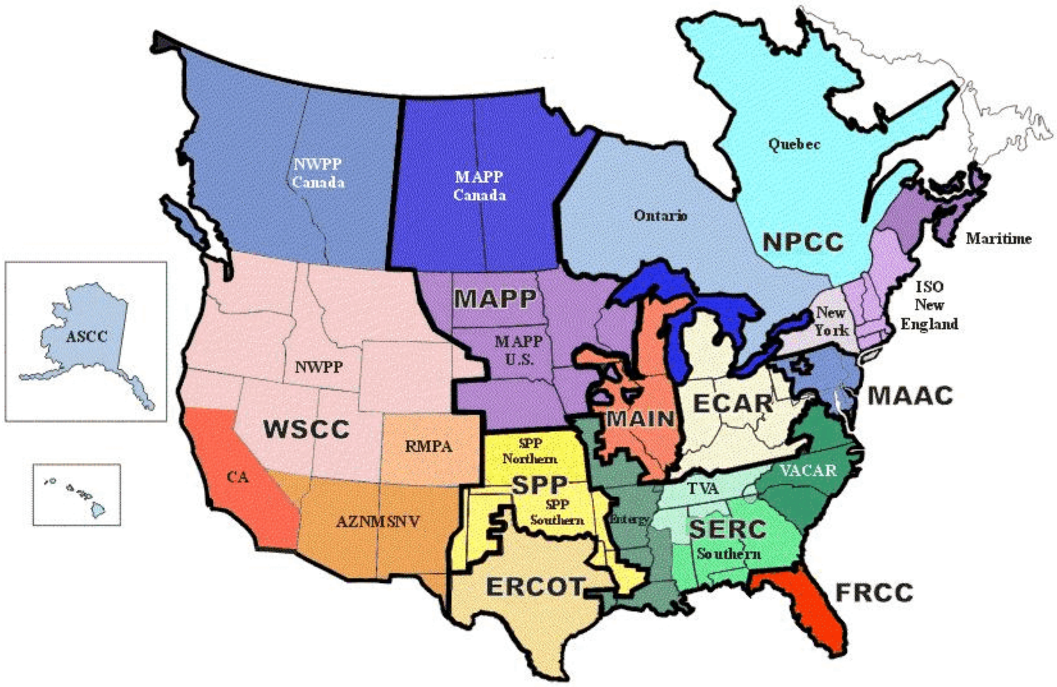

Figure 1. North American Electric Reliability Council Regions and Subregions (NERC).

identification codes have also been cross-referenced to the plant identification codes used in the EPA National Emissions Inventory (NEI).

The hourly CEMS data were aggregated to obtain daily emissions and daily average boiler heat inputs. We used daily instead of hourly values to reduce the variability in the CEMS data set that is unrelated to meteorology. In particular, power demands are subject to diurnal variations related to the work schedule that are independent of diurnal temperature patterns.

EGU emissions were aggregated to geographical regions of different sizes for regression analysis. First, the EGUs were grouped based on their nearest metropolitan statistical area. The EGUs were aggregated to the state level, and to larger geographical areas based generally on the North American Electric Reliability Council (NERC) regions. The NERC regions are shown in Figure 1. Total emissions and boiler heat inputs were calculated for each region, and emissions were also aggregated for the major fuel types (coal, oil, natural gas, and other) within each region.

Figure 2 shows sample NO_x emissions data for the Commonwealth of Virginia. The figure shows a considerable amount of variability in the daily emissions, but also some seasonal patterns that are typical of many other states. Emissions are generally higher in the winter than in the summer months, but tend to increase during the hotter months of summer. Emissions are also reduced in the summer of 2003 compared to previous years because of the installation of control devices to comply with the NO_x Budget Program. These controls are often used on a seasonal basis. The distribution of emissions by fuel type shown in Figure 2 is also typical. State level NO_x emissions are generally dominated by a single fuel, in this case coal.

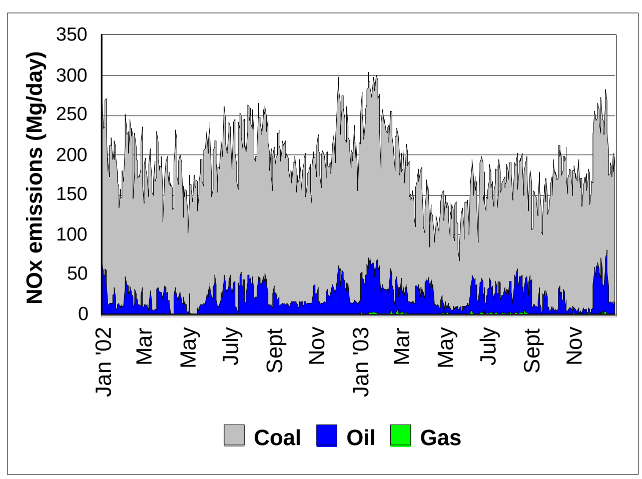

Figure 2. NO_x emissions data for Virginia.

2.2 Meteorological Data

Meteorological data for 2003 for this analysis were obtained from the National Climatic Data Center (NCDC) compilation of Integrated Surface Hourly Observations (NCDC). From the hourly observations, the following daily values were calculated: maximum, minimum and average temperature; maximum, minimum and average dew point; maximum, minimum and average wind speed; hours of clear sky; hours of overcast; hours of precipitation; and cooling degree days. Days having fewer than 20 hours of observations were dropped. The average temperature, dew point, and wind speed were calculated by averaging all observations in the given day. The cooling degree day number was calculated in accordance with NCDC standard calculation methods, as follows:

$$C = \left[\left(T_{max} + T_{min}\right) / 2 \right] - 18.3 \qquad \textbf{(1)}$$

where C is the cooling degree number (°C), T_{max} is the maximum temperature for the day (°C) and T_{min} is the minimum temperature (°C). The cooling degree day number is zero if the daily average temperature is below 18.3°C (65°F).

2.3 Regression Analysis

We defined a modulation factor, F_d, to express day to day variations in NO_X emissions in relation to seasonal average emissions:

$$F_d = \left(E_d - E_{avg}\right) / E_{avg} \qquad \textbf{(2)}$$

where E_d is the day-specific NO_X emission rate for EGUs in a given region (g/sec), and E_{avg} is the average NO_X emission rate in the region (g/sec) over the ozone season (from May through September).

Linear regression models were used to calculate F_d as a function of meteorological parameters and day type for the 2003 area-specific data. A multi-step approach was used to build the final models, and to select the final model structure. The analysis focused on the eastern U.S., which is currently included in the operational version of the Eta-CMAQ forecasting system; however, regression analyses were also carried out for the contiguous western states.

2.3.1 Preliminary analysis

The initial regression analysis included all of the meteorological parameters listed in the *Meteorological Data* section, as well as a binary variable differentiating weekend days and holidays from weekdays:

$$F_d = \alpha + \sum \beta_i P_i + \varepsilon_d \qquad \textbf{(3)}$$

where α, β_i and β_n are parameters calculated in the regression model; P_i are the meteorological parameters and the binary day-type variable (1 for a weekend day or holiday and 0 for a weekday); and ε is the model error. Natural logarithms were taken for each of the meteorological parameters. (Cooling degree days less than 1 were also rounded to 1.0 prior to taking the logarithm.)

Separate regression analyses were carried out for each fuel type within each of the metropolitan areas. Regression analyses were also carried out for the state-level data sets and the larger NERC-based regions.

The results of the initial regression analysis were used to narrow the list of meteorological variables. Sets of interrelated variables, such as maximum temperature, minimum temperature, and average temperature were evaluated to identify the most important member of the set. The final set of parameters was selected based on a conceptual model of the expected relationships of the model parameters to NO_X emissions, the observed correlations among the candidate model parameters, and preliminary information on initial model fits. The final list of regression variables was narrowed to the following five: maximum temperature, average dew point, average wind speed, cooling degree day number, and day type. It is recognized that this list still contains three temperature parameters which are interdependent (maximum temperature, dew point, and cooling degree days). However, these parameters provide different approaches for quantifying the potential demand for air conditioning of residential and office space. Further, the three temperature-related parameters have different influences in different geographical regions.

The initial regression analysis indicated that the correlation was not improved by separating the fuel types used in a given metropolitan area or region. One reason for this result may be the emissions in a given region are generally dominated by one fuel (such as coal in the Southeast and oil in the Northeast)

Initial model fits indicated that the conceptual model structure worked well for some areas, but not others. The correlation was improved by expanding the geographic region size from the metropolitan area level to the state level. In general though, the larger NERC-based regions did not produce significant improvements. An exception was the New England states, which were combined in the final analysis. The initial state-by-state model runs for New England gave explained variances (r^2) ranging from 0.39 (for Maine) to 0.69 (for Massachusetts). Combining these states the overall r^2 to 0.79. The impacts on r^2 were generally smaller for other power sharing regions, and any improvements in r^2 were generally outweighed by a dilution of the impact of meteorological parameters as a result of increased region size.

Issues in the initial model fits included incorrect signs on the model parameters and autocorrelated model errors. Incorrect signs indicate the model is not consistent with scientific understanding. For the purpose of forecasting future emissions, a scientifically valid model was considered an important model selection criterion. The most severe issue was the violation of the normality and independence assumptions of the model errors.

2.3.2 Final analysis

As a result of the findings of the preliminary analysis, fuel types were combined in the final regression analysis. In addition, the final analysis focused on state level NO_X emissions, with some exceptions. The six New England states were grouped together into two small regions – northern and southern New England. In addition, the District of Columbia was lumped together with Maryland, and the Upper Peninsula of Michigan was included with Wisconsin.

To correct for serial correlation of the model errors found in the initial analysis, an autoregressive error model was used. The model goodness-of-fit was evaluated, and the model residuals were tested for normality and independence (i.e., the standard assumptions for the distribution of the model errors).

The conceptual model for the final regression analysis took the following form:

$$\begin{aligned} F_d &= \alpha + \beta_1 \ln\left(T_{max}\right) + \beta_2 \ln\left(Dp_{avg}\right) \\ &\quad + \beta_3 \ln\left(W_{avg}\right) + \beta_4 \ln(C) + \beta_5 Dt + \varepsilon_d \end{aligned} \quad \textbf{(4)}$$

where α and β_0 through β_5 are parameters calculated in the regression model; Dp_{avg} is the average dew point (K); W_{avg} is the average wind speed (km/hour); Dt is the binary day-type variable (defined above), and ε_d is an error term for the day (the difference between the model prediction and the input data).

The autoregressive error model to correct for serial correlation was as follows:

$$\varepsilon_d = -\varphi_1 \varepsilon_{d-1} + \nu_d \quad \textbf{(5)}$$

where φ_1 is the lag 1 autoregressive error model parameter estimated from the data, and ν_d is the nonserial error component.

Testing of the resulting model errors indicated that the assumptions of normality and independence of the model errors were consistently met. In the eastern U.S., regression results were reviewed in detail. In those cases where the model parameter signs were not consistent with the conceptual model, the parameter was set to zero (i.e., the predictor variable was dropped from the model). Examination of the model fit statistic in these cases (r^2) showed no decline in the goodness of model fit, and many times resulted in a small increase in the value of r^2. Model forecasts quickly converge to the mean prediction (after a few time steps).

3. RESULTS AND DISCUSSION

Since future values of the response variable are unknown, the model kernel function is used for to predict the modulation factor:

$$\begin{aligned} F_d &= \alpha + \beta_1 \ln\left(T_{max}\right) + \beta_2 \ln\left(Dp_{avg}\right) \\ &\quad + \beta_3 \ln\left(W_{avg}\right) + \beta_4 \ln(C) + \beta_5 Dt \end{aligned} \quad \textbf{(6)}$$

It must be noted that this analysis has focused on the eastern U.S. (east of the Mississippi River). Results for the western U.S. have not been reviewed as thoroughly, and are therefore subject to greater uncertainty. Because of the complexity of the relationship between NO_X emissions and meteorological parameters, the model is not

Table 1. Regression Analysis Results

Area	Explained error (r^2)	F-factor predictions 10th percentile	F-factor predictions 90th percentile	F-factor limits Minimum	F-factor limits Maximum
Alabama	0.72	-0.08	0.07	-0.33	0.20
Arizona	0.83	-0.12	0.09	-0.27	0.18
Arkansas	0.72	-0.24	0.16	-0.71	0.43
California	0.94	-0.72	0.65	-0.90	1.34
Colorado	0.80	-0.11	0.11	-0.25	0.19
Delaware	0.77	-0.50	0.46	-0.84	1.24
Florida	0.74	-0.09	0.08	-0.29	0.23
Georgia	0.83	-0.19	0.16	-0.54	0.61
Idaho	0.67	-0.80	0.94	-1.00	3.68
Illinois	0.93	-0.32	0.28	-0.70	0.67
Indiana	0.83	-0.14	0.13	-0.42	0.29
Iowa	0.86	-0.18	0.14	-0.37	0.34
Kansas	0.74	-0.20	0.21	-0.56	0.44
Kentucky	0.75	-0.17	0.17	-0.49	0.43
Louisiana	0.73	-0.14	0.13	-0.44	0.39
Maryland and DC	0.87	-0.44	0.39	-0.80	0.88
Michigan, lower	0.89	-0.27	0.54	-0.54	0.88
Minnesota	0.67	-0.10	0.08	-0.51	0.26
Mississippi	0.68	-0.15	0.16	-0.35	0.47
Missouri	0.84	-0.21	0.16	-0.46	0.40
Montana	0.54	-0.17	0.12	-0.68	0.35
Nebraska	0.84	-0.20	0.17	-0.52	0.29
Nevada	0.82	-0.29	0.15	-0.55	0.27
New England	0.79	-0.22	0.25	-0.46	1.13
New Jersey	0.85	-0.55	0.46	-0.86	1.40
New Mexico	0.78	-0.13	0.11	-0.51	0.18
New York	0.82	-0.30	0.36	-0.54	0.75
North Carolina	0.86	-0.27	0.19	-0.66	0.45
North Dakota	0.84	-0.17	0.16	-0.39	0.28
Ohio	0.75	-0.14	0.10	-0.34	0.27
Oklahoma	0.79	-0.16	0.16	-0.36	0.37
Oregon	0.92	-0.84	0.34	-1.00	0.73
Pennsylvania	0.86	-0.21	0.18	-0.55	0.67
South Carolina	0.70	-0.11	0.09	-0.33	0.32
South Dakota	0.81	-0.08	0.12	-1.00	0.27
Tennessee	0.84	-0.23	0.19	-0.45	0.32
Texas	0.89	-0.13	0.12	-0.28	0.36
Utah	0.61	-0.08	0.06	-0.35	0.18
Virginia	0.83	-0.23	0.20	-0.57	0.48
Washington	0.80	-0.31	0.60	-1.00	1.28
West Virginia	0.73	-0.15	0.14	-0.57	0.56
Wisconsin and upper MI	0.90	-0.20	0.20	-0.50	0.49
Wyoming	0.70	-0.13	0.12	-0.37	0.22

valid for extrapolation beyond the range of conditions analyzed in the 2003 data set. Table 1 shows the explained variance, r^2, for the regression models. The r^2 statistic ranges from 0.54 in Montana to above 0.9. The r^2 statistic is above 0.75 for most of the regions. The table also shows the 10th and 90th percentile predictions of the modulation factor, F, which give an indication of the variations in emission predictions for various regions. Finally, the table shows the maximum and minimum modulation factors which have been established for each region based on the range of conditions in 2003.

Figure 3 shows the estimated impact of a 10°C increase in temperature on emissions in different regions of the continental U.S. Figure 4 compares predicted emissions with observed emissions for the summer of 2003 in North Carolina. The horizontal line at the center of the figure reflects the average daily emission rate for the summer season. As the figure shows, the regression model gives a significant improvement in predicted emissions over seasonal average emission rates. On average, the mean error was reduced from 21% to 9%. Figure 4 also illustrates the impact of meteorological parameters, with emissions varying by up to 20% from the seasonal average. However, the figure also shows that the daily emission rate is subject to considerable variability that is not explained by the meteorological regression model. Other sources of variability include day-to-day variations in the mixture of fuels burned and the average efficiency of control devices. In addition, the regression model makes only a simple distinction between weekdays and weekends and holidays. Industrial and residential electricity demands will vary between Saturday and Sunday, and during the course of the work week. In addition, power demands will vary during the course of the summer season as a result of vacation schedules and product demands.

4. SUMMARY

This effort has demonstrated that NO_X emissions from EGUs can be correlated to meteorological parameters. In addition, regression results indicate that meteorological parameters can have a significant impact on NO_X emissions (see Figures 3 and 4).

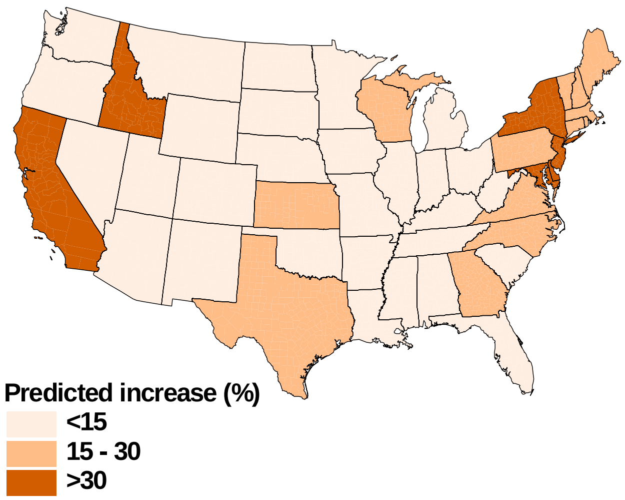

Figure 3. Estimated impact of a 10°C change in ambient temperature on daily NO_x emissions (percent increase above summer season average).

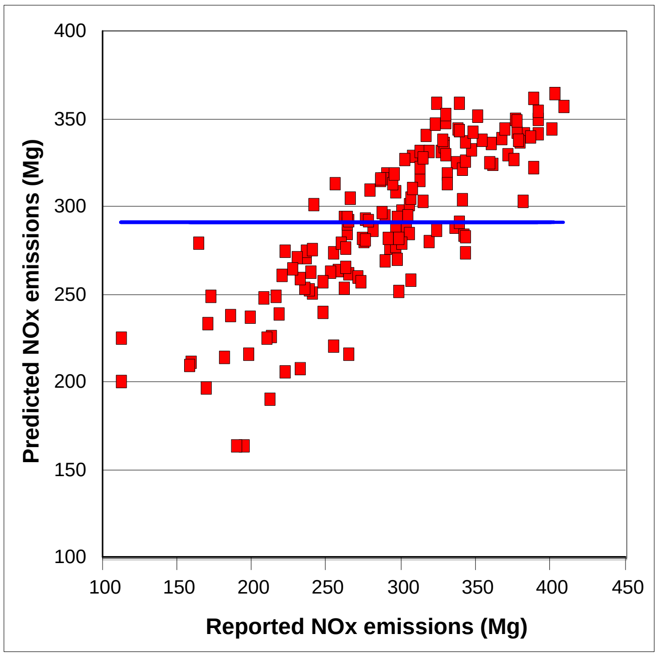

Figure 4. Comparison of predicted and observed daily NO_x emissions for North Carolina in the summer of 2003, where predictions are based on meteorology and day type.

NOAA is currently testing the regression models for NO_x emissions in its ETA-CMAQ ozone forecast model system. The regression models can also be used for summer NO_x emissions in $PM_{2.5}$ modeling. In addition, similar methodologies can be used to develop regression models for winter NO_x emissions, and for SO_2 emissions.

5. REFERENCES

EPA. 2005. Acid Rain/OTC Program Hourly Emissions Data. U.S. Environmental Protection Agency, Clean Air Markets Division, Washington, DC. www.epa.gov/airmarkets/emissions/raw/index.html

NCDC. Integrated Surface Hourly Observations, Volume 23: United States – 2002 and 2003. National Climatic Data Center, National Oceanic and Atmospheric Administration, Asheville, NC.

NERC. North American Electric Reliability Council Regions and Subregions. www.eia.doe.gov/cneaf/electricity/chg_str_fuel/html/fig02.html

6. ACKNOWLEDGEMENTS

We gratefully acknowledge the assistance of William Benjey in obtaining the CEM data, the programmatic support of Paula Davidson and Steve Fine of NOAA, and the counsel of Tanya Otte, Rohit Mathur, and George Pouliot.

7. DISCLAIMER

The research presented here was performed under the Memorandum of Understanding between the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to NOAA's Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.