A Nonlinear Regression Model Estimating Single Source Concentrations of Primary and Secondarily Formed PM2.5

Kirk R. Baker

Kristen M. Foley

1 U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards
109 TW Alexander Dr., Research Triangle Park, NC 27711 (baker.kirk@epa.gov)

2 U.S. Environmental Protection Agency, Office of Research and Development
109 TW Alexander Dr., Research Triangle Park, NC 27711
Abstract

Various approaches and tools exist to estimate local and regional PM2.5 impacts from a single emissions source, ranging from simple screening techniques to Gaussian based dispersion models and complex grid-based Eulerian photochemical transport models. These approaches either lack a realistic chemical and physical representation of the atmosphere for secondary PM2.5 formation or, in the case of photochemical models, may be too resource intensive for single source assessments. A simple nonlinear regression model has been developed to estimate annual average downwind primary and secondary PM2.5 nitrate and sulfate from a single emissions source. The statistical model is based on single emissions sources tracked with particulate source apportionment technology in a photochemical transport model. This nonlinear regression model is advantageous in that the underlying data is based on single emissions sources modeled in a realistic chemical and physical environment of a photochemical model, and provides downwind PM2.5 impact information with minimal resource burden. Separate regression models are developed for total primary PM2.5, PM2.5 sulfate ion, and PM2.5 nitrate ion. Regression model inputs include facility emissions rates in tons per year and the distance between the source and receptor. An additional regression model input of receptor ammonia emissions is used to account for the variability in regional ammonia availability that is important for PM2.5 nitrate ion estimates.

Key words: air quality modeling, reduced form PM2.5 model, single source PM2.5 contribution
1. Introduction

Human exposure to particulate matter less than 2.5 microns in diameter (PM2.5) can result in a variety of adverse health impacts including reduced lung function and premature mortality (Pope, 2000). PM2.5 in the eastern United States is largely composed of organic carbon mass, ammonium sulfate, ammonium nitrate, and directly emitted species including oxidized metals. Under certain atmospheric conditions nitrogen oxides (NOx) emissions may be transformed to PM2.5 nitrate ion, and sulfur dioxide (SO2) emissions may be converted to PM2.5 sulfate ion. Emissions from stationary point sources account for most of the total SO2 emission inventory for the eastern United States and more than half of the NOx emission inventory (U.S. Environmental Protection Agency, 2010). Organic carbon emissions from large point sources are small compared to other sectors, and account for a small portion of the emission inventory. A variety of tools exist to assess the impacts of emissions from a single source on regional and local particulate matter. These types of tools range from screening approaches to simple Gaussian based dispersion models and complex 3-dimensional grid-based Eulerian photochemical transport models (Holmes and Morawska, 2006).

The downwind PM2.5 attributed to specific single sources are traditionally estimated with Gaussian/lagrangian models that do not have sophisticated secondary chemical formation or a realistic chemical environment (Holmes and Morawska, 2006). Some langrangian puff models include parameterized or explicit chemistry, but lack a well characterized spatial representation of ammonia and chemical oxidants that are important for secondarily formed particulate matter (Levy et al., 2002). Grid based
photochemical models are very useful for a realistic characterization of the chemical and
physical processes that transform precursor emissions of sulfur dioxide and nitrogen
oxides to secondary particulate matter (Baker and Scheff, 2007). However, the resource
intensive nature of photochemical modeling makes quick assessments of single source
impacts difficult. These types of evaluations are often useful to focus on a certain group
of sources or emissions control options before undertaking a more rigorous assessment of
downwind PM2.5 contribution estimates. Many State and Local agencies prioritize point
sources for permit review and emission inventory development based on facility
emissions totals or emissions divided by distance approaches. Regulatory agencies use
these and similar screening approaches that lack physically realistic PM2.5
concentrations at monitors of interest to determine which sources require more refined
permit modeling.

A variety of approaches have been used to quickly translate PM2.5 precursor
emissions to estimated downwind PM2.5 concentration or assess strength of source. The
Q/D (Q=emissions rate and D=distance from emissions release point) method of dividing
emissions by distance has been used as a screening method, but it only provides useful
results when comparing the relative influence of a variety of source (U.S. Department of
the Interior, 2008; U.S. Environmental Protection Agency, 2005; U.S. Forest Service,
2010). This approach does not directly relate emissions to ambient PM2.5 concentration,
since non-linear chemical and physical processes important for secondary PM2.5
formation are not considered. An additional limitation of this approach is that all PM2.5
precursor emissions have an equivalent strength of influence estimate.
Another common approach is the use of transfer coefficients which are multipliers used to convert an emissions rate to a concentration. Transfer coefficients have been used for a variety of applications including plume trajectory models and to assess inhalation exposure to PM2.5 (Greco et al., 2007; Tsai and Chen, 2006; Tsuang et al., 2003).

Another typical use for transfer coefficients is to efficiently assess the air quality impacts from a variety of pollution control options for numerous sources. These terms are estimated for specific source-receptor combinations and lack generalization to sources and receptors not included in the original study.

A simple statistical model has been developed to provide estimates of primary and secondarily formed sulfate and nitrate PM2.5 downwind from stationary point sources based on single sources modeled with a 3-dimensional photochemical transport model. Single source PM2.5 contributions have been tracked with a photochemical transport model enhanced with source apportionment. This type of model extension allows for tracking of precursor emissions from specific sources to PM2.5 transported downwind and formed through secondary chemical reactions (Wagstrom et al., 2008).

A nonlinear regression model relating single source precursor emissions, distance between source and receptor, and primary and secondary PM2.5 sulfate and nitrate ion concentration at the receptor location is applied for 99 facilities. This nonlinear regression model has the advantage of being based on stationary point sources modeled in the realistic chemical and physical environment of a photochemical model and provides information with minimal resource burden. Regression models are developed separately for sulfate, nitrate, and primary PM2.5 at each plant to account for differences in chemical formation, deposition, and transport between these species and locations. A
hierarchical clustering analysis is used to group similar facilities to provide a form of the nonlinear regression model that would be generalizable to sources not included in the original modeling analysis.

Single source assessments often examine contribution from an incremental change in emissions or compare downwind contribution from an existing emissions rate with modified emissions rate (Levy et al., 2002; U.S. Environmental Protection Agency, 2005; U.S. Forest Service, 2010). The change in downwind concentration due to change in precursor emission, or relative response, of the nonlinear regression model is compared to changes in photochemical model contribution due to emissions adjustments. This comparison provides additional information about how well the nonlinear regression model captures the relative changes in sulfate, nitrate, and primary PM2.5 and what level of emissions changes are best represented.

2. Methods

2.1 Photochemical Model Application

The Comprehensive Air Quality Model with Extensions (CAMx) version 4.5 is a publically available three-dimensional Eularian photochemical transport model that treats the physical and chemical processes that form ozone and PM2.5 (Baker and Scheff, 2007; Nobel et al., 2001; Russell, 2008). CAMx was applied with ISORROPIA inorganic chemistry (Nenes et al., 1999), a semi-volatile equilibrium scheme to partition condensable organic gases between gas and particle phase (Strader et al., 1999), Regional Acid Deposition Model (RADM) aqueous phase chemistry (Chang et al., 1987), and
Carbon Bond 05 (CB05) gas-phase chemistry module (ENVIRON, 2008; Gery et al., 1989).

Particulate matter source apportionment technology (PSAT) implemented in CAMx estimates the contribution from specific emissions source groups to PM2.5 using reactive tracers (ENVIRON, 2008; Wagstrom et al., 2008). The tracer species are estimated with source apportionment algorithms rather than by the host model routines. PSAT tracks contribution to PM2.5 sulfate, nitrate, ammonium, secondary organic aerosol, and inert primarily emitted species. Non-linear processes like gas and aqueous phase chemistry are solved for bulk species and then apportioned to the tagged species. Emissions of nitrogen oxides are tracked through all intermediate nitrogen species to particulate nitrate ion. Ammonia emissions are tracked to particulate ammonium ion. Even though ammonium nitrate is chemically coupled, the apportionment scheme does not attempt to determine which species is limiting the formation, but directly attributes precursor gases to specific particulate ions. This method only attempts to apportion the emissions and does not provide emissions sensitivity information that would require estimating whether ammonia or nitric acid is limiting PM2.5 ammonium nitrate formation.

All photochemical modeling is done with a lambert conformal projection centered at -97 degrees longitude and 40 degrees latitude using square 12 km sized grid cells. The vertical atmosphere up to approximately 15 km above ground level is resolved with 14 layers. The layers are smaller inside the planetary boundary layer (mixing layer) to capture the important diurnal variations in the mixing height. Point source emissions of SO₂, NOₓ, and primary PM2.5 species from 99 plants in the eastern United States are
tracked with PSAT for the entire year of 2005. Specific facilities were selected for source
contribution assessment (shown in Figure 1) if total facility emissions of NO\(_x\), SO\(_x\), or
primary PM2.5 were among the top 5% of all stationary point sources located east of -97
longitude and were located in one of the 3 existing photochemical modeling domains:
midwest, mid-Atlantic, and southeast (Figure 1). This criteria resulted in selecting point
sources that had emissions of NO\(_x\) > 7,000 tpy, SO\(_x\) > 20,000 tpy, or PM2.5 > 1,100 tpy
based on the 2005 emissions inventory. Each point source has a symbol corresponding to
the modeling domain used for tracking that particular source (Figure 1). All point sources
were also tracked with emissions reductions of 50% and 90% to assess the relationship
between a percentage emissions reduction and the resulting change in source contribution
estimates. This sensitivity was performed for the entire 2005 simulation over the
southeast and midwest model domains.

The emissions used for the photochemical modeling were based on the 2005
National Emission Inventory for stationary point, onroad and nonroad sources. Hour
specific continuous emissions monitoring (CEM) data from 2005 are used for emissions
of point sources that have CEM data available. Day specific biogenic emissions were
estimated using hourly gridded day-specific meteorology for 2005. Other area sources
were projected from the 2002 National Emission Inventory to approximate the level of
emissions expected in 2005 (Strum, 2008). All emissions were processed using the
Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System version 2.5
(Houyoux et al., 2000). Speciation profiles allocate selected facility specific emissions
based on state and county location and plant identification codes to unique emissions
species that are tracked in the source apportionment photochemical model.
2.2 Nonlinear Regression Model

Annual average concentrations of PM2.5 sulfate, nitrate, and primary PM2.5 attributable to specific facilities output from CAMx/PSAT are used to estimate separate nonlinear regression models (Equations 1 and 2) that relate downwind PM2.5 concentration to source emissions and distance between source and receptor. All single source contribution estimates greater than 1,000 km from the source or located in the same grid cell as the source are not included in the data used for estimating the regression model. Equation 1 is used to estimate downwind concentrations of PM2.5 sulfate ion and primary PM2.5 species. Downwind concentrations of PM2.5 nitrate ion are influenced by the presence of ammonia emissions at the receptor location of interest, which is included as an additional term in Equation 2.

\[
Concentration = \frac{\beta_1 \cdot \text{emissions}}{1 + (\beta_2 \cdot \text{dist})^{\beta_3}} \quad (1)
\]

\[
Concentration = \frac{\beta_1 \cdot \text{emissions} + \beta_4 \cdot \text{ammonia}}{1 + (\beta_2 \cdot \text{dist})^{\beta_3}} \quad (2)
\]

Source (facility) and ammonia emissions are expressed in tons per year and distance is expressed in km. Concentration is estimated in units of \(\mu g/m^3\) and the coefficients \(\beta_1, \beta_2, \beta_3,\) and \(\beta_4\) are constants estimated using nonlinear least squares. Annual sulfur dioxide emissions are input to Equation 1 to estimate downwind concentrations of PM2.5 sulfate ion. Annual primary PM2.5 emissions are used in Equation 1 to estimate concentrations.
of downwind primary PM2.5 species. Source specific annual nitrogen oxides emissions 
are used in Equation 2 with annual emissions of ammonia at the receptor location to 
estimate downwind concentrations of PM2.5 nitrate ion. Annual average county total 
ammonia emissions where the receptor location resides are an example of an appropriate 
input to Equation 2. The coefficients are estimated for each facility for PM2.5 sulfate, 
nitrate, and primary PM2.5 species using the nls function in the R statistical software 
package (R, 2011).

This nonlinear function is selected to match the shape of the underlying source 
contribution data, which tends to be highest nearest the source at 12 km grid resolution 
and then steadily decreases with increasing distance. The rate of decrease in source 
contribution by distance from the source is different for each species shown in Figure 2. 
In Equation (1), when distance approaches zero the concentration is equal to $\beta_1 *$ 
emissions. Similar to the use of transfer coefficients, the estimated beta term is used to 
scale the emissions rate at the source to a concentration value. The additional term in the 
numerator of Equation (2), $\beta_4 * ammonia$, is used to account for the spatial variability in 
regional ammonia that is important for PM2.5 nitrate ion formation. The $\beta_2$ and $\beta_3$ 
parameters control how quickly concentrations decrease with increasing distance.

An important limitation of these empirical equations is that the resolution of the 
underlying data is 12 km. Application of these equations to distances less than 12 km is 
not recommended. The source apportionment data used to estimate the beta coefficients 
extends out to a maximum of 1000 km, which is the furthest distance from the source this 
equation would potentially be applicable. Beta coefficients are estimated for all facilities 
in all model domains. Several different forms of these equations were considered before
220 choosing Equations (1) and (2) as the best fit of the underlying data. The $\beta_3$ parameter
221 was removed from Equations (1) and (2) creating a model with 2 beta coefficients in
222 Equation (1) and a model with 3 beta coefficients in Equation (2). In addition, Equation
223 (1) was used to estimate PM2.5 nitrate ion concentrations and these predictions were
224 compared to the results from Equation (2) that incorporate regional ammonia emissions.
225
226 2.3 Hierarchical Clustering Analysis
227 A cluster analysis is used to group similar facilities to provide a general form of the
228 nonlinear regression model that may be applied to sources not included in the original
229 modeling analysis. To calculate the similarity/dissimilarity of the estimated regression
230 models at different facilities, the estimated beta coefficients at each site are used to
231 predict pollutant concentrations at all of the sites in the domain based on source specific
232 emissions. A Euclidean distance metric is then computed between the predicted
233 concentration values from each pair of regression model results (i.e. for N total facilities,
234 a total of $\binom{N}{2}$ distance metrics are computed) (Johnson and Wichern, 2002). A
235 hierarchical cluster analysis is performed based on these distance metrics using the hclust
236 function in the R statistical software package (R, 2011). A separate cluster analysis is
237 done for the sulfate, nitrate and primary PM2.5 regression models. The purpose of this
238 analysis is to identify facilities or groups of facilities that may have unique micro-scale
239 meteorology, orography, or other physical facility characteristics that would result in
240 regional spatial patterns of PM2.5 contribution not characteristic of most plants in the
241 eastern United States. Primary PM2.5 and PM2.5 sulfate ion spatial patterns are expected
242 to be similar at different plants in the eastern United States since directly emitted PM2.5
has no chemical transformations in the atmosphere and sulfur dioxide to PM2.5 sulfate ion transformations are simple and have minimal regional dependence. In contrast, PM2.5 nitrate ion concentrations at individual facilities may have very different regional patterns since NOx transformation to PM2.5 nitrate is dependent on the availability of ammonia emissions and temperature. Once the clustering analysis has identified groups of similarly behaving facilities, new beta estimates are calculated using all of the available data from these groups.

3. Results & Discussion

3.1 Evaluation of Photochemical Model

Photochemical model estimates are matched to observations of chemically speciated PM2.5 in time and space. Speciated PM2.5 data from the EPA Chemical Speciation Network (CSN) and Interagency Monitoring of Protected Visual Environments Network (IMPROVE) are used to estimate operational model performance. Metrics used to describe model performance include mean bias, gross error, fractional bias, and fractional error (Boylan and Russell, 2006). The bias and error metrics describe performance in terms of measured concentration units and the fractional metrics are expressed as a percentage bounded by 200%. The best possible performance is when the metrics approach zero. Model performance metrics are the average of all daily metrics over all three model domains by calendar quarter (Table 1).

Average model-observation bias is less than 0.5 μg/m³ for PM2.5 sulfate and nitrate ion. Under-estimation of organic carbon is related to under-estimated secondary
organic aerosol formation which is typically not directly related to large stationary point
source emissions. Better model representation of organic carbon formation could provide
some indirect impact on modeled sulfate and nitrate but would unlikely make a large
change to sulfate and nitrate contribution estimates. Model performance for speciated
PM2.5 is similar to other published photochemical modeling applications (Doraiswamy,
2010; Tesche et al., 2006). With the exception of summer organic carbon and summer
nitrate ion, fractional bias and fractional error metrics for other species aggregated by
quarter meet recommended model performance goals (Boylan and Russell, 2006). PM2.5
nitrate ion ambient concentrations and model predictions are very small in the summer
and the model appropriately predicts the elevated ambient PM2.5 nitrate ion
concentrations typical in the cooler months making model performance acceptable for
this application.

3.2 Nonlinear Regression Model

Nonlinear least squares regression is used to estimate the beta coefficients relating
to emissions ($\beta_1$ and $\beta_4$) and dispersion over distance ($\beta_2$ and $\beta_3$) (Neter, 1996). The $\beta_1$, $\beta_2$,
and $\beta_3$ coefficients are estimated for PM2.5 sulfate, nitrate, and primary PM2.5 while $\beta_4$
is only estimated for nitrate. Data for all sources (N=99) in all model domains are used to
estimate beta coefficients for each facility, which represent different dominant
meteorological regimes, chemical environments, and topography. The coefficients
estimated using annual average source contribution information are shown in Figure 3.
The sulfate and primary PM2.5 coefficients are estimated using Equation (1) and the
nitrate coefficients are estimated using Equation (2).
The estimated $\beta_1$ coefficients relating to emissions are very similar across all facilities for the sulfate and primary PM2.5 regression models (see far left panel of Figure 3). There is greater variability in the estimated emissions related beta coefficients ($\beta_1$ and $\beta_4$) for the nitrate regression model due to differences in downwind ammonia availability and regional climate having more of an impact on nitrate. A notable geographic pattern is seen in the coefficient term relating to NO$_x$ emissions ($\beta_1$). The nitrate $\beta_1$ coefficient for sources in the midwest region and eastern Pennsylvania are generally higher than the southeast and Appalachian mountain region. The midwest and mid-Atlantic regions are more conducive for PM2.5 nitrate formation due to ambient ammonia availability and the higher frequency of cold temperatures than the southeast (Baker and Scheff, 2007). The estimated dispersion related coefficient, $\beta_2$, tends to be smaller in the sulfate and nitrate models compared to the estimates for the primary PM2.5 model (second panel of Figure 3). A smaller $\beta_2$ value in Equation (1) will result in higher predicted concentrations at greater distances. This is consistent with Figure 2 which shows that the primary PM2.5 concentration values decreasing sharply within 200 km of the source and sulfate concentrations having a larger regional impact. PM2.5 nitrate dispersal is difficult to compare with the other species due to the influence of receptor ammonia on concentration estimates.

The coefficient of determination ($R^2$) and root mean square error (RMSE) are estimated using the source contribution data compared to the nonlinear regression model fitted concentrations. Metrics are based on contribution data for distances less than 500 km to focus the evaluation on the region nearer the source with the largest pollutant concentrations. Figure 4 shows both metrics for sulfate and primary PM2.5 using
Equation (1), nitrate using Equation (2), and all species using a simpler form of the nonlinear regression model where the denominator exponent coefficient is removed. In addition, a model using nitrate and Equation (1) is presented to compare the impact of including receptor ammonia emissions in the nonlinear regression model. Performance metrics improve when using Equations (1) and (2) compared to the simpler forms. Nonlinear regression model estimates of PM2.5 nitrate ion using Equation (2) have less error and higher correlation with the source contribution data compared to using Equation (1). This supports the inclusion of receptor ammonia emissions in the nonlinear regression model to appropriately estimate regional PM2.5 nitrate contributions from large sources of nitrogen oxides.

The spatial extent and gradient estimated by the regression model for a single source is illustrated for sulfate and primary PM2.5 in Figure 5 and nitrate in Figure 6. The magnitude and spatial pattern estimated by the nonlinear regression model compare well with the PSAT annual average contribution estimates. The spatial pattern of downwind PM2.5 nitrate ion concentrations estimated with Equation (2) show more directional variability due to the inclusion of receptor ammonia emissions in the regression model. The CAMx PSAT contribution results for PM2.5 nitrate ion in Figure 6 clearly show the importance of considering regional (receptor) ammonia to generate an accurate spatial characterization of concentration gradients. When only emissions of NOX are considered (far right panel of Figure 6), the spatial pattern of PM2.5 nitrate ion peaks near the source and does not match the spatial peaks seen in the PSAT data (far left panel of Figure 6) in areas near the NOX source that have elevated ammonia. The single source PM2.5 nitrate contributions estimated by CAMx PSAT show that regional ammonia needs to be well
characterized by any modeling system to appropriately estimate regional PM2.5 nitrate
centrations (Levy et al., 2002). By capturing the magnitude of the concentrations and
the spatial gradient away from the source, the proposed nonlinear regression models are
well suited to be used for screening a group of sources or emission control options before
using more rigorous modeling of downwind PM2.5 contribution estimates.

3.3 Hierarchical Clustering Analysis

Sources are grouped using complete linkage hierarchical cluster analysis for
sulfate, primary PM2.5, and nitrate to identify sources that reflect micro-scale impacts
from meteorology, orography, or other plant specific characteristics that may not be
obvious in a large scale analysis. A dendrogram diagram of the source clusters is
qualitatively evaluated to isolate and remove small groups of sources that are most
dissimilar from the largest cluster group(s). A total of 17 sources are removed from the
primary PM2.5 dataset and 23 are removed from the sulfate data. The removed groups
represent sources that may be influenced by unique circumstances and are less likely to
represent a typical spatial pattern of source contribution. For example, the mean
emissions and mean stack height of the primary and sulfate sources that are removed are
significantly lower than the average emissions and stack height of the other facilities in
the domain. The remaining sources (over 75% of the original data set) are found to be
more similar in terms of the estimated regression models which reflects the fairly simple
nature of the dispersal and chemical transformation of sulfate and primary PM2.5. Beta
coefficients estimated using data from the remaining sources for sulfate and primary
PM2.5 are shown in Table 2.
The cluster analysis for the nitrate ion contribution sources resulted in 3 cluster
groups. One group was made up of a single source with a comparatively small stack
height. The remaining two groups are made up of a similar number of sources (see
Groups 1 and 2 in Figure 7). These two groups reflect known chemical and physical
relationships in that one group tends to be closer to ammonia rich environments (Midwest
and eastern Pennsylvania) and the other in less ammonia rich areas and warmer climate,
both of which are less conducive for PM2.5 nitrate formation. The cluster groups tend to
overlap along the boundary between these two regions. Estimated coefficients for these 2
groups are also presented in Table 2. Information about local and regional (or receptor)
ammonia emissions is necessary to select which group of nitrate ion coefficients might be
more applicable for large point sources not included in this analysis. PM2.5 nitrate group
1 tends to represent sources in areas favorable to PM2.5 nitrate formation: abundant
ammonia emissions and cooler temperatures. Cluster group 2 for nitrate represents plants
located in less favorable conditions for nitrate formation. Evaluation metrics (again
estimated for distances less than 500 km) for the new nonlinear regression models based
on the clustered data are very similar to the $R^2$ and RMSE values from the source specific
regression models (Table 3).

A cross validation analysis is performed to evaluate how generalizable the final
beta estimates are to other sources within the domain not included in this analysis. For
each cluster, 15% of the sources are sampled at random to be withheld for cross
validation and new beta coefficients are estimated based on the remaining data. These
new beta estimates are used to predict the concentrations at the cross validation sites and
new evaluation metrics are calculated. This process is repeated a total of ten times and
the average evaluation statistics across the ten replicates are reported in Table 4. For
sulfate ion and primary PM2.5 the $R^2$ is .89 and .85 respectively, and the mean fractional
bias is less than 2% for both species, indicating the estimated nonlinear regression model
is well suited for predicting source contributions in this domain. The performance of the
nitrate model is weaker, with mean fractional bias of 30% and 22% and $R^2$ values of .67
and .51 for the two clusters. Model performance metrics and cross validation results
suggest betas presented in Table 2 would be generalizable. PM2.5 nitrate ion formation is
much more complex than the other species considered and that is reflected in the
comparatively degraded performance. However, predicted spatial patterns of PM2.5
nitrate ion contribution compare well to CAMx/PSAT estimates indicating the nonlinear
regression model is capable of estimating downwind single source PM2.5 nitrate ion.
Information about local and regional (or receptor) ammonia emissions is necessary to
select which group of nitrate ion coefficients might be more applicable for large point
sources not included in this analysis.

3.4 Source Contribution Sensitivity

PM2.5 precursor emissions are reduced by 50% and 90% for all plants tracked
with source apportionment in the southeast and midwest model domains. This type of
evaluation is intended to provide information about how well the nonlinear model
replicates the changes in PM2.5 concentration due to emissions adjustments and whether
some level of emissions adjustments are not well characterized. Figure 8 shows the model
response for annual average PM2.5 sulfate, nitrate, and primary PM2.5 contributions
estimated by CAMx/PSAT as distance from the source increases. Model response is
shown in the form of a relative response ratio. The relative response ratio is the source contribution concentration in the sensitivity simulation divided by the source concentration estimated in the original model application. As expected, sulfate and primary PM2.5 changes in concentration match the percent emissions change most closely. The most variability in relative response ratio estimated by CAMx/PSAT is for changes in nitrate concentration due to differences in downwind meteorological regimes, non-linear chemistry, and ammonia availability.

The average relative response ratio estimated by the photochemical model for PM2.5 sulfate ion after emissions reductions of 50% and 90% are 0.53 and 0.11 respectively (N=56 sources). Average photochemical model estimated primary PM2.5 relative response ratios are 0.49 and 0.09 for 50% and 90% emissions reductions. The ratios estimated using the nonlinear regression model for these same emissions reductions are 0.50 and 0.10 due to the direct proportionality between emissions and concentration in Equation (1). Photochemical model predicted average relative response ratio for PM2.5 nitrate ion is 0.55 for a 50% reduction of nitrogen oxides emissions and 0.12 for a 90% emissions reduction. The nonlinear regression model produced an average relative response ratio of 0.61 for a 50% emissions reduction and a ratio of 0.30 for a 90% emissions reduction. The change in sulfate and primary PM2.5 relative response is consistent with the percent emissions reduction suggesting the nonlinear regression model estimates would provide utility in providing a preliminary estimate of the change in concentration that might result from a change in emissions. The relative response of PM2.5 nitrate ion is fairly comparable between nonlinear regression model estimates and
CAMx PSAT estimates at a 50% emissions reduction but less representative at a large emissions change of 90%.

4. CONCLUSIONS

Nonlinear regression models relating facility emissions and distance between source and receptor to PM2.5 concentration are developed using annual average distributions of single source contributions as estimated by a photochemical grid model. These regression models are most relevant to applications for annual average single source PM2.5 contribution in the eastern United States. The nonlinear regression model does well predicting single source downwind PM2.5 sulfate, PM2.5 nitrate, and primarily emitted PM2.5 species compared to PSAT estimates. Some incommensurability is inevitable when estimating contributions from stationary point sources using a grid-based photochemical model since emissions are diffused throughout the volume of the grid cell where emissions occur. This would need most consideration when interpreting the results nearest the stack, which is in the 0 to 24 km range since a 12 km grid is used for the photochemical modeling. A fine scale dispersion model or sub-grid plume treatment may be more appropriate to estimate contribution from emissions when considering distances less than the grid resolution of this model application. Other studies suggest single source contribution estimates of a secondarily formed pollutant (ozone) are appropriate using grid cell sizes of 12 km (Cohan et al., 2006).

Past methods estimated transfer coefficients for a specific source and receptor combination which precludes generalized application. Often PM2.5 contributions are needed for sources or receptors that were not part of a retrospective analysis which makes
a generalized equation for planning purposes useful. This nonlinear regression model presents an alternative to traditional transfer coefficient terms and has the advantage of including the source-receptor distance relationship as part of the downwind concentration estimate. This approach is also favorable to the Q/D method, since downwind PM2.5 estimates are given in physically meaningful units and different precursor emissions have different downwind source strength.

The PM2.5 concentrations estimated by this nonlinear regression model do not replace the need for refined modeling to support regulatory decisions, but may inform planning efforts and provide information when resources or time do not allow for rigorous modeling applications. Like any reduced form model, this model is strongest at replicating simpler systems like primary PM2.5 dispersion compared to very complex systems like PM2.5 nitrate ion formation. This nonlinear regression model could be expanded to be applicable to other types of emissions sources, such as low level stationary point sources and near-ground level emissions sources such as mobile and area source emissions. Additionally, coefficients could be estimated by season to better replicate the intra-annual variability in PM2.5 formation. These augmentations would involve doing additional PSAT simulations to develop new underlying data to estimate additional regression coefficients.

**Disclaimer**

Although this manuscript has been reviewed and approved for publication, it does not necessarily reflect the policy or views of the US Environmental Protection Agency.
Acknowledgements

The authors would like to recognize the contribution of Margaret McCourtney of Minnesota Pollution Control Agency, who provided insight, testing, and comment on the regression model approach. Additional support was provided by Jenise Swall, James Crooks, Ann Marie Carlton, Pat Dolwick and Karen Wesson of the U.S. Environmental Protection Agency.
Literature cited


Tsai, Y.I., Chen, C.L., 2006. Atmospheric aerosol composition and source apportionments to aerosol in southern Taiwan. Atmospheric Environment 40, 4751-4763.


List of Tables

Table 1. Photochemical model performance metrics for PM2.5 species by quarter.
Table 2. Estimated regression model coefficients for PM2.5 sulfate, primary PM2.5, and PM2.5 nitrate.
Table 3. Nonlinear regression model performance metrics for PM2.5 sulfate, primary PM2.5 and PM2.5 nitrate. Metrics are based on source contributions at distances less than 500km. Metrics include coefficient of determination ($R^2$), root mean square error (RMSE, $\mu g/m^3$) and mean fractional bias (MFB, %).
Table 4. Cross validation (CV) results for PM2.5 sulfate, primary PM2.5 and PM2.5 nitrate. Cross validation performance statistics are reported as the average metric across the ten replicates.
List of Figures

Figure 1. Single emissions sources tracked with CAMx/PSAT and used to generate the nonlinear regression model.

Figure 2. CAMx/PSAT estimates of single source concentrations of (top) PM2.5 sulfate ion, (middle) PM2.5 nitrate ion, and (bottom) primary PM2.5 by distance from the source. Lower and upper edges of the box represent the 25th and 75th percentiles.

Figure 3. Regression coefficients estimated for PM2.5 sulfate, PM2.5 nitrate, and primary PM2.5 at all single sources. Lower and upper edges of the box represent the 25th and 75th percentiles.

Figure 4. Model performance metrics $R^2$ and root mean square error for all species using final and alternative nonlinear regression models. Metrics are based on contribution data for distances less than 500 km. Lower and upper edges of the box represent the 25th and 75th percentiles.

Figure 5. Single source annual average estimated PM2.5 sulfate by (top left) CAMx/PSAT and (top right) nonlinear regression model; single source annual average estimated primary PM2.5 by (bottom left) CAMx/PSAT and (bottom right) nonlinear regression model.

Figure 6. Single source annual average estimated PM2.5 nitrate by (left) CAMx/PSAT, (middle) alternative nonlinear regression model excluding ammonia emissions, and (right) final nonlinear regression model.

Figure 7. Single emissions sources grouped using cluster approach based on PM2.5 nitrate regression model.
Figure 8. Annual average relative response ratios averaged by distance for 50% and 90% emissions reductions shown for (left) PM2.5 sulfate, (middle) primary PM2.5, and (right) PM2.5 nitrate. Lower and upper edges of the box represent the 25\textsuperscript{th} and 75\textsuperscript{th} percentiles.