1   A Nonlinear Regression Model Estimating Single Source Concentrations of Primary and

2   Secondarily Formed PM2.5

3

4   Kirk R. Baker[1]

5   Kristen M. Foley[2]

6

7   [1]U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards

8   109 TW Alexander Dr., Research Triangle Park, NC 27711 (baker.kirk@epa.gov)

9   [2]U.S. Environmental Protection Agency, Office of Research and Development

10 109 TW Alexander Dr., Research Triangle Park, NC 27711

11

12

13

14

15   **Abstract**

16   Various approaches and tools exist to estimate local and regional PM2.5 impacts

17   from a single emissions source, ranging from simple screening techniques to Gaussian

18   based dispersion models and complex grid-based Eulerian photochemical transport

19   models. These approaches either lack a realistic chemical and physical representation of

20   the atmosphere for secondary PM2.5 formation or, in the case of photochemical models,

21   may be too resource intensive for single source assessments. A simple nonlinear

22   regression model has been developed to estimate annual average downwind primary and

23   secondary PM2.5 nitrate and sulfate from a single emissions source. The statistical model

24   is based on single emissions sources tracked with particulate source apportionment

25   technology in a photochemical transport model. This nonlinear regression model is

26   advantageous in that the underlying data is based on single emissions sources modeled in

27   a realistic chemical and physical environment of a photochemical model, and provides

28   downwind PM2.5 impact information with minimal resource burden. Separate regression

29   models are developed for total primary PM2.5, PM2.5 sulfate ion, and PM2.5 nitrate ion.

30   Regression model inputs include facility emissions rates in tons per year and the distance

31   between the source and receptor. An additional regression model input of receptor

32   ammonia emissions is used to account for the variability in regional ammonia availability

33   that is important for PM2.5 nitrate ion estimates.

34

35   Key words: air quality modeling, reduced form PM2.5 model, single source PM2.5

36   contribution

37

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 (NO_X) emissions may be transformed to PM2.5 nitrate ion, and sulfur dioxide (SO_2) emissions may be converted to PM2.5 sulfate ion. Emissions from stationary point sources account for most of the total SO_2 emission inventory for the eastern United States and more than half of the NO_X 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

61 photochemical models are very useful for a realistic characterization of the chemical and

62 physical processes that transform precursor emissions of sulfur dioxide and nitrogen

63 oxides to secondary particulate matter (Baker and Scheff, 2007). However, the resource

64 intensive nature of photochemical modeling makes quick assessments of single source

65 impacts difficult. These types of evaluations are often useful to focus on a certain group

66 of sources or emissions control options before undertaking a more rigorous assessment of

67 downwind PM2.5 contribution estimates. Many State and Local agencies prioritize point

68 sources for permit review and emission inventory development based on facility

69 emissions totals or emissions divided by distance approaches. Regulatory agencies use

70 these and similar screening approaches that lack physically realistic PM2.5

71 concentrations at monitors of interest to determine which sources require more refined

72 permit modeling.

73   A variety of approaches have been used to quickly translate PM2.5 precursor

74 emissions to estimated downwind PM2.5 concentration or assess strength of source. The

75 Q/D (Q=emissions rate and D=distance from emissions release point) method of dividing

76 emissions by distance has been used as a screening method, but it only provides useful

77 results when comparing the relative influence of a variety of source (U.S. Department of

78 the Interior, 2008; U.S. Environmental Protection Agency, 2005; U.S. Forest Service,

79 2010). This approach does not directly relate emissions to ambient PM2.5 concentration,

80 since non-linear chemical and physical processes important for secondary PM2.5

81 formation are not considered. An additional limitation of this approach is that all PM2.5

82 precursor emissions have an equivalent strength of influence estimate.

83   Another common approach is the use of transfer coefficients which are multipliers

84   used to convert an emissions rate to a concentration. Transfer coefficients have been used

85   for a variety of applications including plume trajectory models and to assess inhalation

86   exposure to PM2.5 (Greco et al., 2007; Tsai and Chen, 2006; Tsuang et al., 2003).

87   Another typical use for transfer coefficients is to efficiently assess the air quality impacts

88   from a variety of pollution control options for numerous sources. These terms are

89   estimated for specific source-receptor combinations and lack generalization to sources

90   and receptors not included in the original study.

91   A simple statistical model has been developed to provide estimates of primary and

92   secondarily formed sulfate and nitrate PM2.5 downwind from stationary point sources

93   based on single sources modeled with a 3-dimensional photochemical transport model.

94   Single source PM2.5 contributions have been tracked with a photochemical transport

95   model enhanced with source apportionment. This type of model extension allows for

96   tracking of precursor emissions from specific sources to PM2.5 transported downwind

97   and formed through secondary chemical reactions (Wagstrom et al., 2008).

98   A nonlinear regression model relating single source precursor emissions, distance

99   between source and receptor, and primary and secondary PM2.5 sulfate and nitrate ion

100   concentration at the receptor location is applied for 99 facilities. This nonlinear

101   regression model has the advantage of being based on stationary point sources modeled

102   in the realistic chemical and physical environment of a photochemical model and

103   provides information with minimal resource burden. Regression models are developed

104   separately for sulfate, nitrate, and primary PM2.5 at each plant to account for differences

105   in chemical formation, deposition, and transport between these species and locations. A

106 hierarchical clustering analysis is used to group similar facilities to provide a form of the

107 nonlinear regression model that would be generalizable to sources not included in the

108 original modeling analysis.

109   Single source assessments often examine contribution from an incremental

110 change in emissions or compare downwind contribution from an existing emissions rate

111 with modified emissions rate (Levy et al., 2002; U.S. Environmental Protection Agency,

112 2005; U.S. Forest Service, 2010). The change in downwind concentration due to change

113 in precursor emission, or relative response, of the nonlinear regression model is compared

114 to changes in photochemical model contribution due to emissions adjustments. This

115 comparison provides additional information about how well the nonlinear regression

116 model captures the relative changes in sulfate, nitrate, and primary PM2.5 and what level

117 of emissions changes are best represented.

118

119 **2. Methods**

120

121 *2.1 Photochemical Model Application*

122   The Comprehensive Air Quality Model with Extensions (CAMx) version 4.5 is a

123 publically available three-dimensional Eularian photochemical transport model that treats

124 the physical and chemical processes that form ozone and PM2.5 (Baker and Scheff, 2007;

125 Nobel et al., 2001; Russell, 2008). CAMx was applied with ISORROPIA inorganic

126 chemistry (Nenes et al., 1999), a semi-volatile equilibrium scheme to partition

127 condensable organic gases between gas and particle phase (Strader et al., 1999), Regional

128 Acid Deposition Model (RADM) aqueous phase chemistry (Chang et al., 1987), and

129 Carbon Bond 05 (CB05) gas-phase chemistry module (ENVIRON, 2008; Gery et al.,

130 1989).

131 Particulate matter source apportionment technology (PSAT) implemented in

132 CAMx estimates the contribution from specific emissions source groups to PM2.5 using

133 reactive tracers (ENVIRON, 2008; Wagstrom et al., 2008). The tracer species are

134 estimated with source apportionment algorithms rather than by the host model routines.

135 PSAT tracks contribution to PM2.5 sulfate, nitrate, ammonium, secondary organic

136 aerosol, and inert primarily emitted species. Non-linear processes like gas and aqueous

137 phase chemistry are solved for bulk species and then apportioned to the tagged species.

138 Emissions of nitrogen oxides are tracked through all intermediate nitrogen species to

139 particulate nitrate ion. Ammonia emissions are tracked to particulate ammonium ion.

140 Even though ammonium nitrate is chemically coupled, the apportionment scheme does

141 not attempt to determine which species is limiting the formation, but directly attributes

142 precursor gases to specific particulate ions. This method only attempts to apportion the

143 emissions and does not provide emissions sensitivity information that would require

144 estimating whether ammonia or nitric acid is limiting PM2.5 ammonium nitrate

145 formation.

146 All photochemical modeling is done with a lambert conformal projection centered

147 at -97 degrees longitude and 40 degrees latitude using square 12 km sized grid cells. The

148 vertical atmosphere up to approximately 15 km above ground level is resolved with 14

149 layers. The layers are smaller inside the planetary boundary layer (mixing layer) to

150 capture the important diurnal variations in the mixing height. Point source emissions of

151 SO_2, NO_X, and primary PM2.5 species from 99 plants in the eastern United States are

152 tracked with PSAT for the entire year of 2005. Specific facilities were selected for source

153 contribution assessment (shown in Figure 1) if total facility emissions of NO_X, SO_X, or

154 primary PM2.5 were among the top 5% of all stationary point sources located east of -97

155 longitude and were located in one of the 3 existing photochemical modeling domains:

156 midwest, mid-Atlantic, and southeast (Figure 1). This criteria resulted in selecting point

157 sources that had emissions of NO_X > 7,000 tpy, SO_X > 20,000 tpy, or PM2.5 > 1,100 tpy

158 based on the 2005 emissions inventory. Each point source has a symbol corresponding to

159 the modeling domain used for tracking that particular source (Figure 1). All point sources

160 were also tracked with emissions reductions of 50% and 90% to assess the relationship

161 between a percentage emissions reduction and the resulting change in source contribution

162 estimates. This sensitivity was performed for the entire 2005 simulation over the

163 southeast and midwest model domains.

164   The emissions used for the photochemical modeling were based on the 2005

165 National Emission Inventory for stationary point, onroad and nonroad sources. Hour

166 specific continuous emissions monitoring (CEM) data from 2005 are used for emissions

167 of point sources that have CEM data available. Day specific biogenic emissions were

168 estimated using hourly gridded day-specific meteorology for 2005. Other area sources

169 were projected from the 2002 National Emission Inventory to approximate the level of

170 emissions expected in 2005 (Strum, 2008). All emissions were processed using the

171 Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System version 2.5

172 (Houyoux et al., 2000). Speciation profiles allocate selected facility specific emissions

173 based on state and county location and plant identification codes to unique emissions

174 species that are tracked in the source apportionment photochemical model.

175

176   *2.2 Nonlinear Regression Model*

177   Annual average concentrations of PM2.5 sulfate, nitrate, and primary PM2.5

178   attributable to specific facilities output from CAMx/PSAT are used to estimate separate

179   nonlinear regression models (Equations 1 and 2) that relate downwind PM2.5

180   concentration to source emissions and distance between source and receptor. All single

181   source contribution estimates greater than 1,000 km from the source or located in the

182   same grid cell as the source are not included in the data used for estimating the regression

183   model. Equation 1 is used to estimate downwind concentrations of PM2.5 sulfate ion and

184   primary PM2.5 species. Downwind concentrations of PM2.5 nitrate ion are influenced by

185   the presence of ammonia emissions at the receptor location of interest, which is included

186   as an additional term in Equation 2.

187

188     $$Concentration = \frac{\beta_1 * emissions}{1 + (\beta_2 dist)^{\beta_3}}$$ (1)

189

190     $$Concentration = \frac{\beta_1 * emissions + \beta_4 * ammonia}{1 + (\beta_2 * dist)^{\beta_3}}$$ (2)

191

192   Source (facility) and ammonia emissions are expressed in tons per year and distance is

193   expressed in km. Concentration is estimated in units of $\mu g/m^3$ and the coefficients β_1, β_2,

194   β_3, and β_4 are constants estimated using nonlinear least squares. Annual sulfur dioxide

195   emissions are input to Equation 1 to estimate downwind concentrations of PM2.5 sulfate

196   ion. Annual primary PM2.5 emissions are used in Equation 1 to estimate concentrations

197 of downwind primary PM2.5 species. Source specific annual nitrogen oxides emissions

198 are used in Equation 2 with annual emissions of ammonia at the receptor location to

199 estimate downwind concentrations of PM2.5 nitrate ion. Annual average county total

200 ammonia emissions where the receptor location resides are an example of an appropriate

201 input to Equation 2. The coefficients are estimated for each facility for PM2.5 sulfate,

202 nitrate, and primary PM2.5 species using the *nls* function in the R statistical software

203 package (R, 2011).

204   This nonlinear function is selected to match the shape of the underlying source

205 contribution data, which tends to be highest nearest the source at 12 km grid resolution

206 and then steadily decreases with increasing distance. The rate of decrease in source

207 contribution by distance from the source is different for each species shown in Figure 2.

208 In Equation (1), when distance approaches zero the concentration is equal to β_1 *

209 *emissions*. Similar to the use of transfer coefficients, the estimated beta term is used to

210 scale the emissions rate at the source to a concentration value. The additional term in the

211 numerator of Equation (2), β_4 * *ammonia*, is used to account for the spatial variability in

212 regional ammonia that is important for PM2.5 nitrate ion formation. The β_2 and β_3

213 parameters control how quickly concentrations decrease with increasing distance.

214   An important limitation of these empirical equations is that the resolution of the

215 underlying data is 12 km. Application of these equations to distances less than 12 km is

216 not recommended. The source apportionment data used to estimate the beta coefficients

217 extends out to a maximum of 1000 km, which is the furthest distance from the source this

218 equation would potentially be applicable. Beta coefficients are estimated for all facilities

219 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 β_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

243 has no chemical transformations in the atmosphere and sulfur dioxide to PM2.5 sulfate

244 ion transformations are simple and have minimal regional dependence. In contrast,

245 PM2.5 nitrate ion concentrations at individual facilities may have very different regional

246 patterns since NO_X transformation to PM2.5 nitrate is dependent on the availability of

247 ammonia emissions and temperature. Once the clustering analysis has identified groups

248 of similarly behaving facilities, new beta estimates are calculated using all of the

249 available data from these groups.

250

251 **3. Results & Discussion**

252

253 *3.1 Evaluation of Photochemical Model*

254 Photochemical model estimates are matched to observations of chemically

255 speciated PM2.5 in time and space. Speciated PM2.5 data from the EPA Chemical

256 Speciation Network (CSN) and Interagency Monitoring of Protected Visual

257 Environments Network (IMPROVE) are used to estimate operational model performance.

258 Metrics used to describe model performance include mean bias, gross error, fractional

259 bias, and fractional error (Boylan and Russell, 2006). The bias and error metrics describe

260 performance in terms of measured concentration units and the fractional metrics are

261 expressed as a percentage bounded by 200%. The best possible performance is when the

262 metrics approach zero. Model performance metrics are the average of all daily metrics

263 over all three model domains by calendar quarter (Table 1).

264 Average model-observation bias is less than 0.5 $\mu g/m^3$ for PM2.5 sulfate and

265 nitrate ion. Under-estimation of organic carbon is related to under-estimated secondary

266   organic aerosol formation which is typically not directly related to large stationary point

267   source emissions. Better model representation of organic carbon formation could provide

268   some indirect impact on modeled sulfate and nitrate but would unlikely make a large

269   change to sulfate and nitrate contribution estimates. Model performance for speciated

270   PM2.5 is similar to other published photochemical modeling applications (Doraiswamy,

271   2010; Tesche et al., 2006). With the exception of summer organic carbon and summer

272   nitrate ion, fractional bias and fractional error metrics for other species aggregated by

273   quarter meet recommended model performance goals (Boylan and Russell, 2006). PM2.5

274   nitrate ion ambient concentrations and model predictions are very small in the summer

275   and the model appropriately predicts the elevated ambient PM2.5 nitrate ion

276   concentrations typical in the cooler months making model performance acceptable for

277   this application.

278

279   *3.2 Nonlinear Regression Model*

280   Nonlinear least squares regression is used to estimate the beta coefficients relating

281   to emissions (β_1 and β_4) and dispersion over distance (β_2 and β_3) (Neter, 1996). The β_1, β_2,

282   and β_3 coefficients are estimated for PM2.5 sulfate, nitrate, and primary PM2.5 while β_4

283   is only estimated for nitrate. Data for all sources (N=99) in all model domains are used to

284   estimate beta coefficients for each facility, which represent different dominant

285   meteorological regimes, chemical environments, and topography. The coefficients

286   estimated using annual average source contribution information are shown in Figure 3.

287   The sulfate and primary PM2.5 coefficients are estimated using Equation (1) and the

288   nitrate coefficients are estimated using Equation (2).

289   The estimated β_1 coefficients relating to emissions are very similar across all

290   facilities for the sulfate and primary PM2.5 regression models (see far left panel of Figure

291   3). There is greater variability in the estimated emissions related beta coefficients (β_1 and

292   β_4) for the nitrate regression model due to differences in downwind ammonia availability

293   and regional climate having more of an impact on nitrate. A notable geographic pattern is

294   seen in the coefficient term relating to NO_X emissions (β_1). The nitrate β_1 coefficient for

295   sources in the midwest region and eastern Pennsylvania are generally higher than the

296   southeast and Appalachian mountain region. The midwest and mid-Atlantic regions are

297   more conducive for PM2.5 nitrate formation due to ambient ammonia availability and the

298   higher frequency of cold temperatures than the southeast (Baker and Scheff, 2007). The

299   estimated dispersion related coefficient, β_2, tends to be smaller in the sulfate and nitrate

300   models compared to the estimates for the primary PM2.5 model (second panel of Figure

301   3). A smaller β_2 value in Equation (1) will result in higher predicted concentrations at

302   greater distances. This is consistent with Figure 2 which shows that the primary PM2.5

303   concentration values decreasing sharply within 200 km of the source and sulfate

304   concentrations having a larger regional impact. PM2.5 nitrate dispersal is difficult to

305   compare with the other species due to the influence of receptor ammonia on

306   concentration estimates.

307   The coefficient of determination (R^2) and root mean square error (RMSE) are

308   estimated using the source contribution data compared to the nonlinear regression model

309   fitted concentrations. Metrics are based on contribution data for distances less than 500

310   km to focus the evaluation on the region nearer the source with the largest pollutant

311   concentrations. Figure 4 shows both metrics for sulfate and primary PM2.5 using

312 Equation (1), nitrate using Equation (2), and all species using a simpler form of the

313 nonlinear regression model where the denominator exponent coefficient is removed. In

314 addition, a model using nitrate and Equation (1) is presented to compare the impact of

315 including receptor ammonia emissions in the nonlinear regression model. Performance

316 metrics improve when using Equations (1) and (2) compared to the simpler forms.

317 Nonlinear regression model estimates of PM2.5 nitrate ion using Equation (2) have less

318 error and higher correlation with the source contribution data compared to using Equation

319 (1). This supports the inclusion of receptor ammonia emissions in the nonlinear

320 regression model to appropriately estimate regional PM2.5 nitrate contributions from

321 large sources of nitrogen oxides.

322   The spatial extent and gradient estimated by the regression model for a single

323 source is illustrated for sulfate and primary PM2.5 in Figure 5 and nitrate in Figure 6. The

324 magnitude and spatial pattern estimated by the nonlinear regression model compare well

325 with the PSAT annual average contribution estimates. The spatial pattern of downwind

326 PM2.5 nitrate ion concentrations estimated with Equation (2) show more directional

327 variability due to the inclusion of receptor ammonia emissions in the regression model.

328 The CAMx PSAT contribution results for PM2.5 nitrate ion in Figure 6 clearly show the

329 importance of considering regional (receptor) ammonia to generate an accurate spatial

330 characterization of concentration gradients. When only emissions of NO_X are considered

331 (far right panel of Figure 6), the spatial pattern of PM2.5 nitrate ion peaks near the source

332 and does not match the spatial peaks seen in the PSAT data (far left panel of Figure 6) in

333 areas near the NO_X source that have elevated ammonia. The single source PM2.5 nitrate

334 contributions estimated by CAMx PSAT show that regional ammonia needs to be well

335 characterized by any modeling system to appropriately estimate regional PM2.5 nitrate

336 concentrations (Levy et al., 2002). By capturing the magnitude of the concentrations and

337 the spatial gradient away from the source, the proposed nonlinear regression models are

338 well suited to be used for screening a group of sources or emission control options before

339 using more rigorous modeling of downwind PM2.5 contribution estimates.

340

341 *3.3 Hierarchical Clustering Analysis*

342   Sources are grouped using complete linkage hierarchical cluster analysis for

343 sulfate, primary PM2.5, and nitrate to identify sources that reflect micro-scale impacts

344 from meteorology, orography, or other plant specific characteristics that may not be

345 obvious in a large scale analysis. A dendrogram diagram of the source clusters is

346 qualitatively evaluated to isolate and remove small groups of sources that are most

347 dissimilar from the largest cluster group(s). A total of 17 sources are removed from the

348 primary PM2.5 dataset and 23 are removed from the sulfate data. The removed groups

349 represent sources that may be influenced by unique circumstances and are less likely to

350 represent a typical spatial pattern of source contribution. For example, the mean

351 emissions and mean stack height of the primary and sulfate sources that are removed are

352 significantly lower than the average emissions and stack height of the other facilities in

353 the domain. The remaining sources (over 75% of the original data set) are found to be

354 more similar in terms of the estimated regression models which reflects the fairly simple

355 nature of the dispersal and chemical transformation of sulfate and primary PM2.5. Beta

356 coefficients estimated using data from the remaining sources for sulfate and primary

357 PM2.5 are shown in Table 2.

358   The cluster analysis for the nitrate ion contribution sources resulted in 3 cluster

359   groups. One group was made up of a single source with a comparatively small stack

360   height. The remaining two groups are made up of a similar number of sources (see

361   Groups 1 and 2 in Figure 7). These two groups reflect known chemical and physical

362   relationships in that one group tends to be closer to ammonia rich environments (Midwest

363   and eastern Pennsylvania) and the other in less ammonia rich areas and warmer climate,

364   both of which are less conducive for PM2.5 nitrate formation. The cluster groups tend to

365   overlap along the boundary between these two regions. Estimated coefficients for these 2

366   groups are also presented in Table 2. Information about local and regional (or receptor)

367   ammonia emissions is necessary to select which group of nitrate ion coefficients might be

368   more applicable for large point sources not included in this analysis. PM2.5 nitrate group

369   1 tends to represent sources in areas favorable to PM2.5 nitrate formation: abundant

370   ammonia emissions and cooler temperatures. Cluster group 2 for nitrate represents plants

371   located in less favorable conditions for nitrate formation. Evaluation metrics (again

372   estimated for distances less than 500 km) for the new nonlinear regression models based

373   on the clustered data are very similar to the R^2 and RMSE values from the source specific

374   regression models (Table 3).

375   A cross validation analysis is performed to evaluate how generalizable the final

376   beta estimates are to other sources within the domain not included in this analysis. For

377   each cluster, 15% of the sources are sampled at random to be withheld for cross

378   validation and new beta coefficients are estimated based on the remaining data. These

379   new beta estimates are used to predict the concentrations at the cross validation sites and

380   new evaluation metrics are calculated. This process is repeated a total of ten times and

381 the average evaluation statistics across the ten replicates are reported in Table 4. For

382 sulfate ion and primary PM2.5 the R^2 is .89 and .85 respectively, and the mean fractional

383 bias is less than 2% for both species, indicating the estimated nonlinear regression model

384 is well suited for predicting source contributions in this domain. The performance of the

385 nitrate model is weaker, with mean fractional bias of 30% and 22% and R^2 values of .67

386 and .51 for the two clusters. Model performance metrics and cross validation results

387 suggest betas presented in Table 2 would be generalizable. PM2.5 nitrate ion formation is

388 much more complex than the other species considered and that is reflected in the

389 comparatively degraded performance. However, predicted spatial patterns of PM2.5

390 nitrate ion contribution compare well to CAMx/PSAT estimates indicating the nonlinear

391 regression model is capable of estimating downwind single source PM2.5 nitrate ion.

392 Information about local and regional (or receptor) ammonia emissions is necessary to

393 select which group of nitrate ion coefficients might be more applicable for large point

394 sources not included in this analysis.

395

396 *3.4 Source Contribution Sensitivity*

397   PM2.5 precursor emissions are reduced by 50% and 90% for all plants tracked

398 with source apportionment in the southeast and midwest model domains. This type of

399 evaluation is intended to provide information about how well the nonlinear model

400 replicates the changes in PM2.5 concentration due to emissions adjustments and whether

401 some level of emissions adjustments are not well characterized. Figure 8 shows the model

402 response for annual average PM2.5 sulfate, nitrate, and primary PM2.5 contributions

403 estimated by CAMx/PSAT as distance from the source increases. Model response is

404 shown in the form of a relative response ratio. The relative response ratio is the source

405 contribution concentration in the sensitivity simulation divided by the source contribution

406 concentration estimated in the original model application. As expected, sulfate and

407 primary PM2.5 changes in concentration match the percent emissions change most

408 closely. The most variability in relative response ratio estimated by CAMx/PSAT is for

409 changes in nitrate concentration due to differences in downwind meteorological regimes,

410 non-linear chemistry, and ammonia availability.

411   The average relative response ratio estimated by the photochemical model for

412 PM2.5 sulfate ion after emissions reductions of 50% and 90% are 0.53 and 0.11

413 respectively (N=56 sources). Average photochemical model estimated primary PM2.5

414 relative response rations are 0.49 and 0.09 for 50% and 90% emissions reductions. The

415 ratios estimated using the nonlinear regression model for these same emissions reductions

416 are 0.50 and 0.10 due to the direct proportionality between emissions and concentration

417 in Equation (1). Photochemical model predicted average relative response ratio for

418 PM2.5 nitrate ion is 0.55 for a 50% reduction of nitrogen oxides emissions and 0.12 for a

419 90% emissions reduction. The nonlinear regression model produced an average relative

420 response ratio of 0.61 for a 50% emissions reduction and a ratio of 0.30 for a 90%

421 emissions reduction. The change in sulfate and primary PM2.5 relative response is

422 consistent with the percent emissions reduction suggesting the nonlinear regression

423 model estimates would provide utility in providing a preliminary estimate of the change

424 in concentration that might result from a change in emissions. The relative response of

425 PM2.5 nitrate ion is fairly comparable between nonlinear regression model estimates and

426   CAMx PSAT estimates at a 50% emissions reduction but less representative at a large

427   emissions change of 90%.

428

429   **4. CONCLUSIONS**

430   Nonlinear regression models relating facility emissions and distance between

431   source and receptor to PM2.5 concentration are developed using annual average

432   distributions of single source contributions as estimated by a photochemical grid model.

433   These regression models are most relevant to applications for annual average single

434   source PM2.5 contribution in the eastern United States. The nonlinear regression model

435   does well predicting single source downwind PM2.5 sulfate, PM2.5 nitrate, and primarily

436   emitted PM2.5 species compared to PSAT estimates. Some incommensurability is

437   inevitable when estimating contributions from stationary point sources using a grid-based

438   photochemical model since emissions are diffused throughout the volume of the grid cell

439   where emissions occur. This would need most consideration when interpreting the results

440   nearest the stack, which is in the 0 to 24 km range since a 12 km grid is used for the

441   photochemical modeling. A fine scale dispersion model or sub-grid plume treatment may

442   be more appropriate to estimate contribution from emissions when considering distances

443   less than the grid resolution of this model application. Other studies suggest single source

444   contribution estimates of a secondarily formed pollutant (ozone) are appropriate using

445   grid cell sizes of 12 km (Cohan et al., 2006).

446   Past methods estimated transfer coefficients for a specific source and receptor

447   combination which precludes generalized application. Often PM2.5 contributions are

448   needed for sources or receptors that were not part of a retrospective analysis which makes

449   a generalized equation for planning purposes useful. This nonlinear regression model

450   presents an alternative to traditional transfer coefficient terms and has the advantage of

451   including the source-receptor distance relationship as part of the downwind concentration

452   estimate. This approach is also favorable to the Q/D method, since downwind PM2.5

453   estimates are given in physically meaningful units and different precursor emissions have

454   different downwind source strength.

455   The PM2.5 concentrations estimated by this nonlinear regression model do not

456   replace the need for refined modeling to support regulatory decisions, but may inform

457   planning efforts and provide information when resources or time do not allow for

458   rigorous modeling applications. Like any reduced form model, this model is strongest at

459   replicating simpler systems like primary PM2.5 dispersion compared to very complex

460   systems like PM2.5 nitrate ion formation. This nonlinear regression model could be

461   expanded to be applicable to other types of emissions sources, such as low level

462   stationary point sources and near-ground level emissions sources such as mobile and area

463   source emissions. Additionally, coefficients could be estimated by season to better

464   replicate the intra-annual variability in PM2.5 formation. These augmentations would

465   involve doing additional PSAT simulations to develop new underlying data to estimate

466   additional regression coefficients.

467

468   **Disclaimer**

469   Although this manuscript has been reviewed and approved for publication, it does not

470   necessarily reflect the policy or views of the US Environmental Protection Agency.

471

479   **Literature cited**

480

481   Baker, K., Scheff, P., 2007. Photochemical model performance for PM2.5 sulfate, nitrate,
482   ammonium, and precursor species SO2, HNO3, and NH3 at background monitor
483   locations in the central and eastern United States. Atmospheric Environment 41, 6185-
484   6195.

485   Boylan, J.W., Russell, A.G., 2006. PM and light extinction model performance metrics,
486   goals, and criteria for three-dimensional air quality models. Atmospheric Environment
487   40, 4946-4959.

488   Chang, J.S., Brost, R.A., Isaksen, I.S.A., Madronich, S., Middleton, P., Stockwell, W.R.,
489   Walcek, C.J., 1987. A 3-Dimensional Eulerian Acid Deposition Model - Physical
490   Concepts and Formulation. J. Geophys. Res.-Atmos. 92, 14681-14700.

491   Cohan, D.S., Hu, Y.T., Russell, A.G., 2006. Dependence of ozone sensitivity analysis on
492   grid resolution. Atmospheric Environment 40, 126-135.

493   Doraiswamy, P., Hogrefe, C., Hao, W., Civerolo, K., Ku, J., Sistla, G., 2010. A
494   Retrospective Comparison of Model-Based Forecasted PM2.5 Concentrations with
495   Measurements. Journal of Air & Waste Management Association 60, 1293-1308.

496   ENVIRON, 2008. User's Guide Comprehensive Air Quality Model with Extensions.
497   ENVIRON International Corporation, Novato.

498   Gery, M.W., Whitten, G.Z., Killus, J.P., Dodge, M.C., 1989. A Photochemical Kinetics
499   Mechanism for Urban and Regional Scale Computer Modeling. J. Geophys. Res.-Atmos.
500   94, 12925-12956.

501   Greco, S.L., Wilson, A.M., Spengler, J.D., Levy, J.I., 2007. Spatial patterns of mobile
502   source particulate matter emissions-to-exposure relationships across the United States.
503   Atmospheric Environment 41, 1011-1025.

504   Holmes, N.S., Morawska, L., 2006. A review of dispersion modelling and its application
505   to the dispersion of particles: An overview of different dispersion models available.
506   Atmospheric Environment 40, 5902-5928.

507   Houyoux, M.R., Vukovich, J.M., Coats, C.J., Wheeler, N.J.M., Kasibhatla, P.S., 2000.
508   Emission inventory development and processing for the Seasonal Model for Regional Air
509   Quality (SMRAQ) project. J. Geophys. Res.-Atmos. 105, 9079-9090.

510   Johnson, R.A., Wichern, D.W., 2002. Applied Multivariate Statistical Analysis. Prentice-
511   Hall Inc., NJ.

512 Levy, J.I., Spengler, J.D., Hlinka, D., Sullivan, D., Moon, D., 2002. Using CALPUFF to
513 evaluate the impacts of power plant emissions in Illinois: model sensitivity and
514 implications. Atmospheric Environment 36, 1063-1075.

515 Nenes, A., Pandis, S.N., Pilinis, C., 1999. Continued development and testing of a new
516 thermodynamic aerosol module for urban and regional air quality models. Atmospheric
517 Environment 33, 1553-1560.

518 Neter, J., Kutner, M. H., Wasserman, W., 1996. Applied Linear Regression Models, 3 ed.
519 Irwin/McGraw-Hill.

520 Nobel, C.E., McDonald-Buller, E.C., Kimura, Y., Allen, D.T., 2001. Accounting for
521 spatial variation of ozone productivity in NOx emission trading. Environmental Science
522 & Technology 35, 4397-4407.

523 Pope, C.A., 2000. Review: Epidemiological basis for particulate air pollution health
524 standards. Aerosol Science and Technology 32, 4-14.

525 R, 2011. R: A language and environment for statistical computing, R Development Team.
526 R Development Team, R Foundation for Statistical Computing.

527 Russell, A.G., 2008. EPA Supersites Program-related emissions-based particulate matter
528 modeling: Initial applications and advances. J. Air Waste Manage. Assoc. 58, 289-302.

529 Strum, M., Houyoux, M., Mason, R., 2008. Technical Support Document: Preparation of
530 Emissions Inventories For the 2002-based Platform, Version 3, Criteria Air Pollutants, in:
531 Agency, U.S.E.P. (Ed.), Research Triangle Park, NC.

532 Tesche, T.W., Morris, R., Tonnesen, G., McNally, D., Boylan, J., Brewer, P., 2006.
533 CMAQ/CAMx annual 2002 performance evaluation over the eastern US. Atmospheric
534 Environment 40, 4906-4919.

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

537 Tsuang, B.J., Chen, C.L., Lin, C.H., Cheng, M.T., Tsai, Y.I., Chio, C.P., Pan, R.C., Kuo,
538 P.H., 2003. Quantification on the source/receptor relationship of primary pollutants and
539 secondary aerosols by a Gaussian plume trajectory model: Part II. Case study.
540 Atmospheric Environment 37, 3993-4006.

541 U.S. Department of the Interior, 2008. Federal Land Managers' Air Quality Related
542 Values Work Group (FLAG), Vol. 73, No. 131, ed. Federal Register.

543 U.S. Environmental Protection Agency, 2005. 40 CFR Part 51 Regional Haze
544 Regulations and Guidelines for Best Available Retrofit Technology (BART)
545 Determinations; Final Rule. Federal Register.

546   U.S. Environmental Protection Agency, 2010. Our Nation's Air Status and Trends
547   Through 2008, EPA-454/R-09-002 ed. Office of Air Quality Planning and Standards,
548   Research Triangle Park.

549   U.S. Forest Service, N.P.S., U.S. Fish and Wildlife Service, 2010. Federl Land Managers'
550   Air Quality Related Values Work Group (FLAG) Phase I Report--Revised. U.S.
551   Department of the Interior, Denver, Colorado.

552   Wagstrom, K.M., Pandis, S.N., Yarwood, G., Wilson, G.M., Morris, R.E., 2008.
553   Development and application of a computationally efficient particulate matter
554   apportionment algorithm in a three-dimensional chemical transport model. Atmospheric
555   Environment 42, 5650-5659.
556
557

558   **List of Tables**

559

560   Table 1. Photochemical model performance metrics for PM2.5 species by quarter.

561   Table 2. Estimated regression model coefficients for PM2.5 sulfate, primary PM2.5, and

562   PM2.5 nitrate.

563   Table 3. Nonlinear regression model performance metrics for PM2.5 sulfate, primary

564   PM2.5 and PM2.5 nitrate. Metrics are based on source contributions at distances less than

565   500km. Metrics include coefficient of determination (R^2), root mean square error (RMSE,

566     $\mu g/m^3$) and mean fractional bias (MFB, %).

567   Table 4. Cross validation (CV) results for PM2.5 sulfate, primary PM2.5 and PM2.5

568   nitrate. Cross validation performance statistics are reported as the average metric across

569   the ten replicates.

570

571

572   **List of Figures**

573

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

575   nonlinear regression model.

576   Figure 2. CAMx/PSAT estimates of single source concentrations of (top) PM2.5 sulfate

577   ion, (middle) PM2.5 nitrate ion, and (bottom) primary PM2.5 by distance from the

578   source. Lower and upper edges of the box represent the 25th and 75th percentiles.

579   Figure 3. Regression coefficients estimated for PM2.5 sulfate, PM2.5 nitrate, and primary

580   PM2.5 at all single sources. Lower and upper edges of the box represent the 25th and 75th

581   percentiles.

582   Figure 4. Model performance metrics R^2 and root mean square error for all species using

583   final and alternative nonlinear regression models. Metrics are based on contribution data

584   for distances less than 500 km. Lower and upper edges of the box represent the 25th and

585   75th percentiles.

586   Figure 5. Single source annual average estimated PM2.5 sulfate by (top left)

587   CAMx/PSAT and (top right) nonlinear regression model; single source annual average

588   estimated primary PM2.5 by (bottom left) CAMx/PSAT and (bottom right) nonlinear

589   regression model.

590   Figure 6. Single source annual average estimated PM2.5 nitrate by (left) CAMx/PSAT,

591   (middle) alternative nonlinear regression model excluding ammonia emissions, and

592   (right) final nonlinear regression model.

593   Figure 7. Single emissions sources grouped using cluster approach based on PM2.5

594   nitrate regression model.

595   Figure 8. Annual average relative response ratios averaged by distance for 50% and 90%

596   emissions reductions shown for (left) PM2.5 sulfate, (middle) primary PM2.5, and (right)

597   PM2.5 nitrate. Lower and upper edges of the box represent the 25th and 75th percentiles.

598