

1 Article

2 Comparing CMAQ Forecasts with a Neural Network 3 Forecast Model for PM_{2.5} in New York

4 Samuel D. Lightstone^{1,*}, Fred Moshary¹ and Barry Gross¹

5 ¹ City College of New York

6 * Correspondence: slights01@citymail.cuny.edu

7 Academic Editor: name

8 Received: date; Accepted: date; Published: date

9 **Abstract:** Human health is strongly affected by the concentration of fine particulate matter (PM_{2.5}).
10 The need to forecast unhealthy conditions has driven the development of Chemical Transport
11 Models such as CMAQ. These models attempt to simulate the complex dynamics of chemical
12 transport by combined meteorology, emission inventories (EI's), and gas/particle chemistry and
13 dynamics. Ultimately, the goal is to establish useful forecasts that could provide vulnerable
14 members of the population with warnings. In the simplest utilization, any forecast should focus on
15 next day pollution levels, and should be provided by the end of the business day (5PM local). This
16 paper explores the potential of different approaches in providing these forecasts. First, we assess
17 the potential of CMAQ forecasts at the single grid cell level (12km), and show that significant
18 variability not encountered in the field measurements occurs. This observation motivates the
19 exploration of other data driven approaches, in particular, a neural network (NN) approach. This
20 approach makes use of meteorology and PM_{2.5} observations as model predictors. We find that this
21 approach generally results in a more accurate prediction of future pollution levels at the 12km
22 spatial resolution scale of CMAQ. **Furthermore, we find that the NN is able to adjust to the sharp
23 transitions encountered in** pollution transported events, such as smoke plumes from forest fires,
24 **more accurately than CMAQ.**

25 **Keywords:** Air quality model; Air Quality System (AQS); Community Multi-Scale Air Quality
26 (CMAQ) model; Fine particulate matter (PM_{2.5}); Aerosol optical depth (AOD)

28 1. Introduction

29 Fine particulate matter air pollution (PM_{2.5}) is an important issue of public health, particularly
30 for the elderly and young children. The study by *Pope et al.* suggests that exposure to high levels of
31 PM_{2.5} is an important risk factor for cardiopulmonary and lung cancer mortality [1-2]. Furthermore,
32 increased risk of asthma, heart attack and heart failure have been linked to exposure to high PM_{2.5}
33 concentrations [3].

34 PM_{2.5} levels are dynamic and can fluctuate dramatically over different time scales. In addition to
35 local emission sources, pollution events can be the result of aerosol plume transport and intrusion
36 into the lower troposphere. When there is a potential high pollution event, the local air quality
37 agencies must alert the public, and advise the population on proper safety measures, as well as direct
38 the reduction of emission producing activities. Therefore, accurately measuring and predicting fine
39 particulate levels is crucial for public safety.

40 The U.S. Environmental Protection Agency (EPA) established the National Ambient Air Quality
41 Standards (NAAQS), which regulate levels of pollutants such as fine particulate matter. The New
42 York State Department of Environment Conservation (NYSDEC) operates ground stations for
43 monitoring PM_{2.5} and speciation throughout NY State [4]. However, surface sampling is expensive
44 and existing networks are limited and sparse. This results in data gaps that can affect the ability to

45 forecast PM_{2.5} over a 24-hour period. The EPA developed the Models-3 Community Multi-scale Air
46 Quality system (CMAQ), to provide 24-48 hour air quality forecasts. CMAQ provides an investigative
47 tool to explore proper emission control strategies. CMAQ has been the standard for modeling air
48 pollution for nearly two decades because of its ability to independently model different pollutants
49 while describing the atmosphere using “first-principles” [5].

50 In their studies, *McKeen et al* and *Yu et al* evaluate the accuracy of CMAQ forecasts [6-7]. To do
51 so, they use the CMAQ 1200UTC (Version 4.4) forecast model. They observe the midnight-to-
52 midnight local time forecast and compare the hourly and daily average forecasts to the ground
53 monitoring stations. *McKeen et al* [6] observed minimal diurnal variations of PM_{2.5} at urban and
54 suburban monitor locations, with a consistent decrease of PM values between 0100 and 0600 local
55 time. However, the CMAQ model showed significant diurnal variations, leading *McKeen et al* to
56 conclude that aerosol loss during the late night and early morning hours has little effect on PM_{2.5}
57 concentrations, while the CMAQ model does not account for this. Therefore, in addition to testing
58 the hourly CMAQ forecast for a 24-hour period, we focus on the daytime window for two reasons:
59 1) to assess the accuracy of CMAQ when aerosols do not play a reduced roll in forecasting, 2) the
60 forecast should predict the air quality during the time of maximum human exposure.

61 While *these studies* make a distinction between rural and urban locations, they take the average
62 results for all rural and urban locations respectively; thereby, their assessment of the CMAQ model
63 was as at a regional scale, rather than a localized one. In addition to regional emissions, these studies
64 also considered extreme pollution events such as the wildfires in western Canada and Alaska, which
65 occurred during the observation period for the studies by *Yu et al* and *McKeen et al*. The results of this
66 assessment concluded that due to insufficient representation of transport pollution associated with
67 the burning of biomass, CMAQ significantly under predicted the PM_{2.5} values for these events.

68 In the study by *Huang et al* [8], the bias corrected CMAQ forecast was assessed for both the 0600
69 and 1200 UTC release times. The study revealed a general improvement of forecasting skill for the
70 CMAQ model. However, it was observed that the bias correction was limited in predicting extreme
71 events, such as wildfires, and new predictors must be included in the bias correction to predict these
72 events. In this study, CMAQ was assessed as a regional forecasting tool, taking 551 sites, and
73 evaluating the average results in six sub-regions.

74 In our present assessment of the current operational CMAQ forecast model (Version 4.6), we
75 differ from the regional studies above in the following ways: Firstly, in addition to the 1200UTC
76 forecast, we evaluated the 0600UTC forecast for the same period to determine if release time affects
77 the CMAQ forecast. Second, we focused on specific locations, both rural and urban, to assess the
78 potential of CMAQ as a localized forecasting tool. In addition, we revisited the forecast potential of
79 CMAQ for high pollution events, to determine if these events are generally caused by transport, or
80 by local emissions. Finally, we tailor the forecast comparisons to focus on the potential of providing
81 next day forecasts using data prior to 5PM of the previous day, since this is an operational
82 requirement for the state environmental agencies.

83 In focusing on both rural and urban areas in New York State, previous studies have shown
84 anomalies in PM_{2.5} from CMAQ forecasts. For example, in [9], using CMAQ (Version 4.5) with various
85 planetary boundary layer (PBL) parameterizations, PM_{2.5} forecasts during the summer pre-dawn and
86 post-sunset periods were often highly overestimated in New York City (NYC). Further analysis of
87 these cases demonstrated that the most significant error was the retrieval of the PBL height, which
88 was often compressed by the CMAQ model, and did not properly take into account the Urban Heat
89 Island mechanisms that expand the PBL layer [10]. This study showed the importance of PBL height
90 dynamics and meteorological factors that motivated the choice of meteorological forecast inputs used
91 during the NN development.

92 The objective of this paper is to determine the best method to forecast PM_{2.5} by direct comparison
93 with CMAQ output products. In particular, using the CMAQ forecast model, as a baseline, we
94 explore the performance of a NN based data driven approach with suitable meteorological and prior
95 PM_{2.5} input factors.

96

97 1.2 Paper Structure

98 Our present paper is organized in the following manner: In section 2, we analyze CMAQ as the
99 baseline forecaster. We briefly describe the CMAQ model and the forecast schedules that are
100 publically available, as well as the relevant ground stations we use for comparison. We then describe
101 and perform a number of statistical tests using both the direct, as well as the bias compensated,
102 CMAQ outputs. In this section, we show the large dispersion in using the direct results without bias
103 correction.

104 In section 3, we present our NN data driven strategy. This includes a description of all the
105 relevant input factors used, including a combination of present and predicted meteorology, as well
106 as diurnal trends of prior PM_{2.5} levels. We present our first statistical results for the comparisons
107 between CMAQ and the NN for a variety of experiments in order to highlight the conditions in which
108 the NN results are generally an improvement. Then we explore the forecast performance for high
109 pollution multiday transport events, which result in the highest surface PM_{2.5} levels during the
110 observed time period. In this comparison, analyzed by combining a sequence of next-day forecasts
111 together, we find that the neural network seems to follow the trends in PM_{2.5} more accurately than
112 the CMAQ model.

113 In section 4, we summarize our results and describe potential improvements.

114 2. CMAQ Local and Regional Assessment

115 2.1. Datasets

116 2.1.1. Models

117 The CMAQ V4.6 (CB05 gas-phase chemistry) with 12km horizontal resolution was used for this
118 paper. The CMAQ product for meteorology predictions used is the North American Model Non-
119 hydrostatic Multi-scale Model (NAM-NMMB). This version was made available starting February
120 2016. The CMAQ data used for this paper is from February 1, 2016 until October 31, 2016. The station
121 names and locations are listed in table A2. The data can be accessed from reference 11, and the model
122 description can be found in references 12 and 13.

123 The CMAQ model used has a few different configurations: release times of 0600 UTC and 1200
124 UTC, and each release time has a standard forecast as well as a bias corrected forecast. The analog
125 ensemble method is used for bias corrections. The idea is to look at similar weather patterns for the
126 forecast period, and statistically correct the numerical PM_{2.5} forecast based on historical errors. The
127 analog ensemble method is described in detail in *Huang, et al* [8]. For each release time, CMAQ
128 provides a 48-hour forecast. The release time of 0600 UTC and 1200 UTC (2AM and 8AM EDT) does
129 not give the public enough time to react to the forecast on the same day as the release. Consequently,
130 for the 0600 UTC release time, the forecast hours 22 – 45 were used, and for the release time of 1200
131 UTC the forecast hours of 16 – 39 were used. This allowed us to construct a complete 24-hour diurnal
132 period for the forecast time window, which facilitated comparison with the field station data.

133 2.1.2. Ground-based Observations

134 PM_{2.5} ground data is collected from the EPA's AirNow, which collects NYSDEC monitoring
135 station measurements in real time. The station data used for the forecast experiments in this article
136 are from the New York State stations listed in table A1, from January 1, 2011 until December 31, 2016.
137 To assess the accuracy of CMAQ model forecasts, matching the model to the ground monitoring
138 station is necessary. To do this, we use the ground NYSDEC stations that lay within the CMAQ grid
139 cell only. Ground stations that are not found in a CMAQ grid cell were not used for comparison;
140 therefore, no spatial interpolation was done on the model results while mapping the model or
141 meteorological data to the AirNow ground stations. This matching method is widely used for
142 comparing the CMAQ model to ground monitoring stations [6,7,14]. The locational data-points are

143 depicted in Figure A1, the NYSDEC station information can be found in appendix A1, and CMAQ
 144 grid cell information can be found in appendix A2.

145 2.2. Methods

146 2.2.1. Assessing Accuracy of CMAQ Forecasting Models

147 The forecasting skill of the different models were evaluated by computing the R^2 and the root
 148 mean square (RMSE) values from a regression analysis comparing the model to the AirNow
 149 observations. High R^2 values and low RMSE values indicated a good match between the prediction
 150 and the observations. Finally, to directly assess potential biases in the regression assessment, residual
 151 plots (see Figure 7a) are provided to show significant concentration bias.

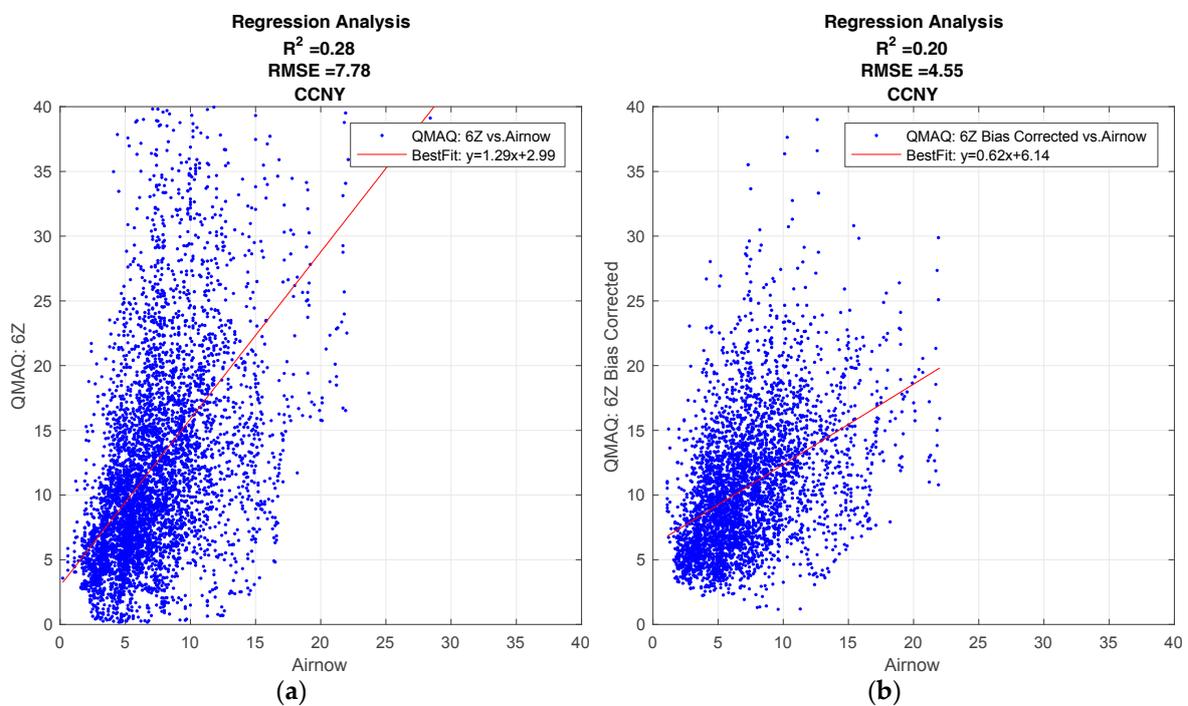
152 2.3. Results

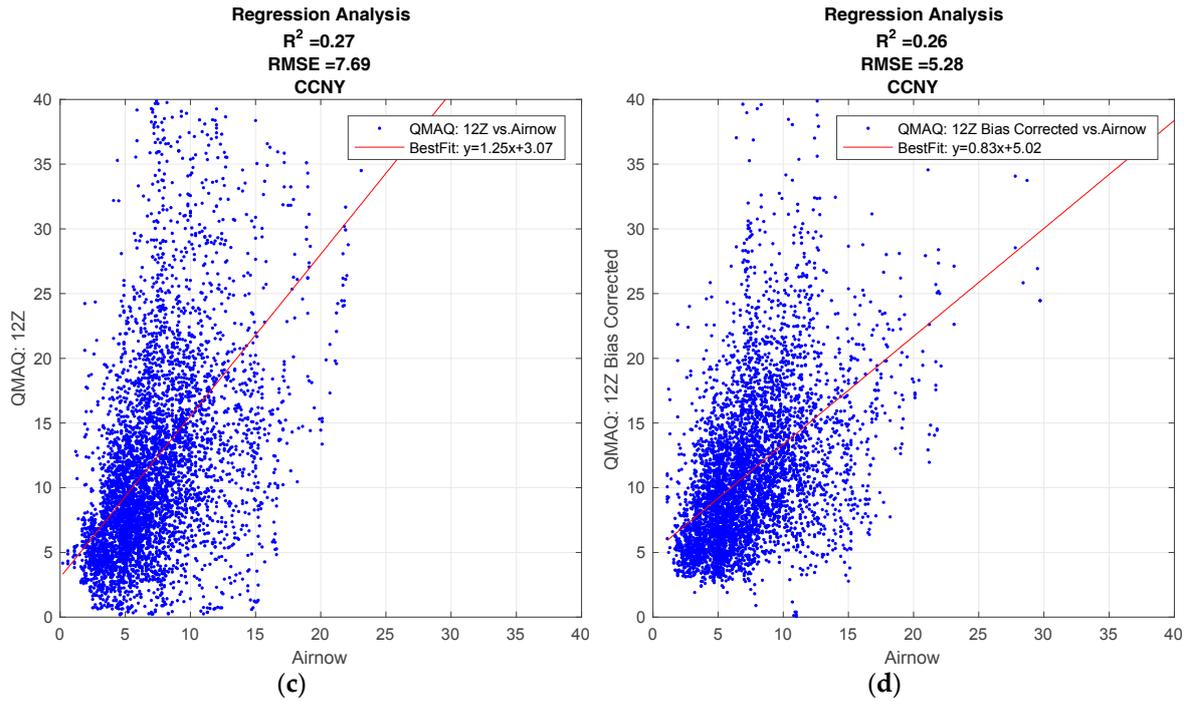
153 2.3.1. Effects of Bias and Release Time

154 Figure 1 shows the regression plots for the hourly CMAQ model output compared to the ground
 155 station data for the City College of New York (CCNY Station) to illustrate the general behavior of the
 156 CMAQ model, and how the forecast is affected by different forecast release times, and by the bias
 157 corrections applied. The results of the R^2 analysis for all ground stations can be found in the
 158 supplementary materials.

159 All forecasts from the CMAQ model over CCNY have a positive correlation to the ground data.
 160 The effect on the forecast for different release times, if any, is minimal.

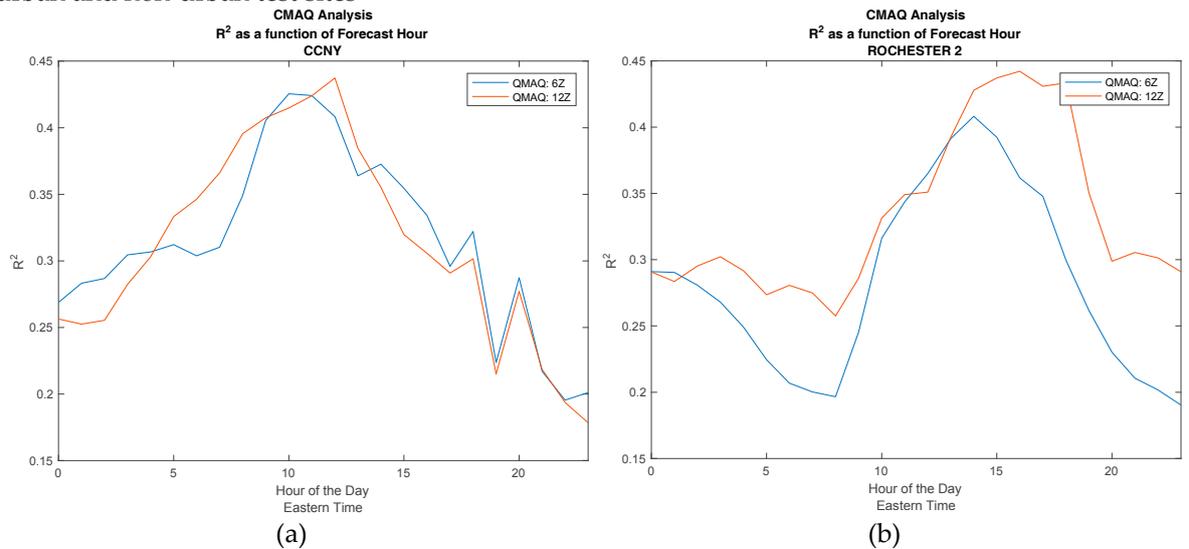
161 As seen in the Figures 1(a) and 1(c), the standard model generally overestimates the ground.
 162 While the bias correction improves the over-prediction, the results are more dispersed. This can be
 163 verified from the fact that the bias correction decreases the root mean square error (RMSE), but it also
 164 decreases the R^2 value for both release times.
 165

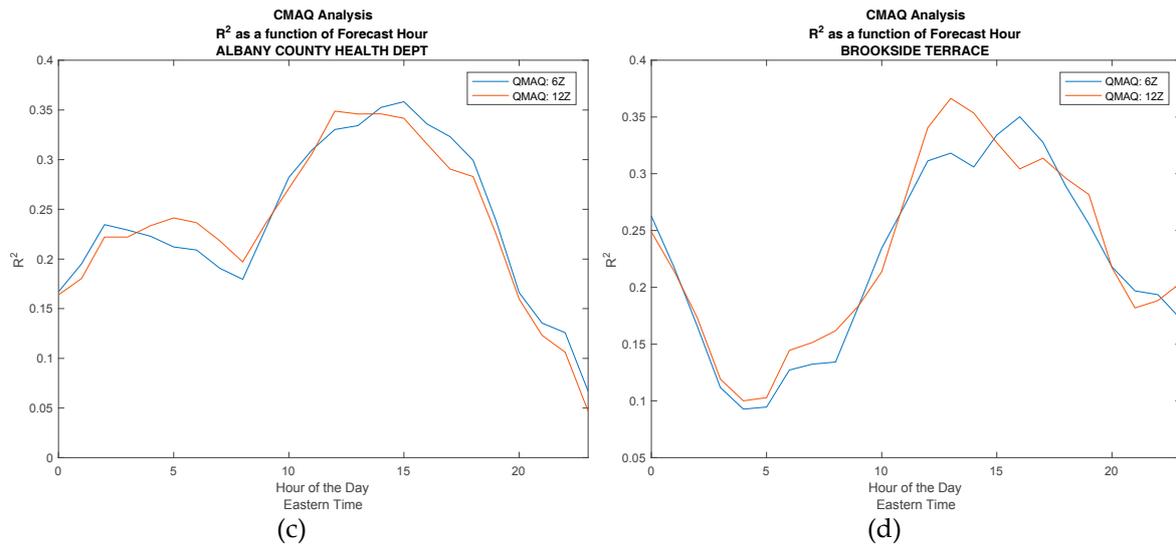




166 **Figure 1.** CMAQ regression analysis. (a) Standard, 06Z release time; (b) Bias Corrected, 06Z
 167 release time; (c) Standard, 12Z release time; (d) Bias Corrected, 12Z release time

168 In Figure 1 we assess the overall skill for a 24-hr CMAQ forecast. In Figure 2, we determine if
 169 the CMAQ model could be improved by simply moving the forecast release time to a later point in
 170 the day, thereby including the most up-to-date inputs in the model. To do this, we make a direct
 171 comparison between CMAQ forecasts with different release times. In Figure 2, the R^2 value is
 172 computed for each hour of the day. The release time of 0600 UTC, with forecast hours of 22 – 45,
 173 is compared to the 1200 UTC release time, with forecast hours 16 – 39, to determine if the lower
 174 number of forecast hours yields more accurate predictions. It is clear from Figure 2 that the later
 175 release time does not lead to a significant improvement in the accuracy of the forecast, and this is true for both
 176 urban and non-urban test sites





177 **Figure 2.** Comparing the effect of different release times for CMAQ by plotting the R^2 value as a
 178 function of time of day. (a) CCNY; (b) Rochester; (c) Albany; (d) Brookside Terrace

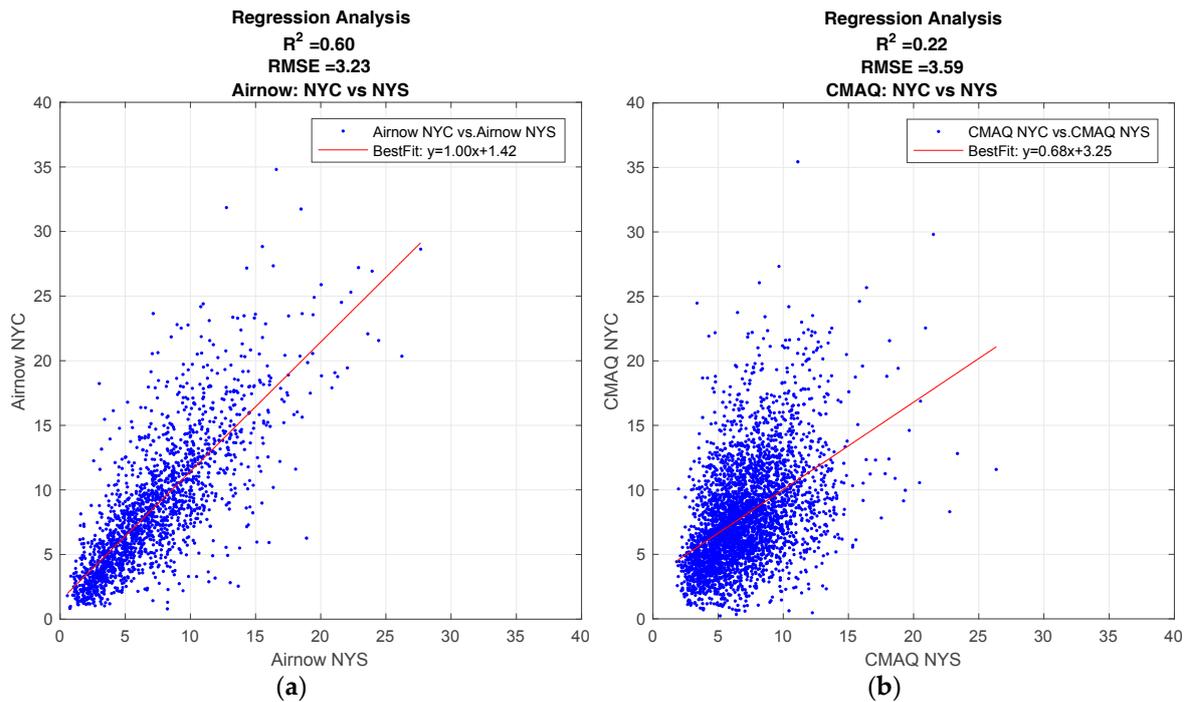
179 It can be seen from this analysis that the CMAQ model performs best for midday hours, which
 180 is reasonable, since this is the period when convective mixing is most dominant. As discussed in
 181 reference [9], PBL modeling is very complex during the predawn / post-sunset period and errors in
 182 the PBL height clearly are a significant concern for further model development.

183 2.3.2. Differences Between Urban and Non-Urban Locations

184 To get a better understanding of the spatial performance of the model, a multi-year time-series
 185 of daily averaged $PM_{2.5}$ observations from ground monitoring is used to compare the relationship
 186 between $PM_{2.5}$ values in New York City to the rest of New York State. Figure 3(a) is the regression
 187 analysis for this time period, and shows how the $PM_{2.5}$ values for NYC are strongly correlated to non-
 188 NYC areas, $R^2 \sim 0.6$. This indicates that while $PM_{2.5}$ values in NYC are generally higher than the rest of
 189 the state, the $PM_{2.5}$ level in NYC are still correlated to the levels in the rest of the state.

190 The same analysis comparing NYC to the rest of NYS was done with CMAQ forecast values as
 191 seen in Figure 3(b). In this case, the correlation between NYC and NYS is not so strong, $R^2 \sim 0.2$. From
 192 this analysis alone, we can only speculate the reason for a low correlation between CMAQ forecasts
 193 for NYC and the rest of NYS is due to strong spatial differences in the National Emission Inventory
 194 (NEI) entries. However, the strong correlation in ground observations between NYC and NYS shows
 195 that while urban source emission may be a significant cause for somewhat higher levels of $PM_{2.5}$,
 196 there is still a strong correlation between NYC and NYS, and an accurate forecasting model must take
 197 this into account.

198



199 **Figure 3.** Regression analysis comparing PM_{2.5} levels between NYC and the rest of NYS (non-
 200 NYC sites). (a); Multi-year day-averaged PM_{2.5} analysis from NYSDEC ground observations; (b)
 201 CMAQ model comparison between NYC and NYS.

202 The limitations of CMAQ forecasting on a local pixel level indicate that other approaches should
 203 be explored. In particular, we explore the potential of data-driven models for localized forecasting in
 204 the next section.

205 3. Data Driven (Neural Network) Development.

206 3.1. Datasets

207 3.1.1. Ground-based observations

208 PM_{2.5} data collected from NYSDEC ground-monitoring stations is used for inputs in the neural
 209 network. These are the same ground stations listed above, in section 2.1.2.

210 3.1.2. Models

211 The meteorological data was collected from the National Centers for Environmental Prediction
 212 (NCEP) North American Regional Reanalysis (NARR). NARR has high-resolution reanalysis of the
 213 North American region, 0.3 degrees (32km) at the lowest latitude, including assimilated precipitation.
 214 The NARR makes available 8-times-daily and monthly means respectively. The data collected for this
 215 paper is the 8-times-daily means for the duration January 1, 2011 until December 31, 2016. Figure A1
 216 shows the proximity of the meteorological data and the CMAQ model outputs to the ground stations.

217 The NN network was created and tested using historical data. In this paper, meteorology
 218 “forecast” data refers to NARR data that was observed the day of the PM_{2.5} forecast. “Observed” or
 219 “measured” meteorology refers to NARR data that was observed before the forecast release time.

220 3.2. Methods

221 3.2.1. Development of the Neural Network

222 As stated above, the accurate prediction of PM_{2.5} values is crucial for air quality agencies, so that
 223 they could alert the public of the severity and duration of a high pollution event. Therefore, it is
 224 imperative that the forecast predictions are released to the public the day before the event. For this

225 paper, we chose 5PM as a target for the forecast release time. Therefore, we ensure that all the
 226 methods tested, utilize factors that are available to the state agency prior to 2100 UTC (5PM EDT).

227 3.2.1.1 Input Selection Scenarios

228 The NN input includes the following NARR meteorological data: surface air temperature,
 229 surface pressure, planetary boundary layer height (PBLH), relative humidity, and horizontal wind
 230 (10m). To account for the seasonal variations, the month is also used as an input in the neural
 231 network. The PM input variables for the NN are the PM_{2.5} measurements averaged over a three-hour
 232 frequency to match the meteorological dataset. The NN output is the next day PM_{2.5} values.

233 In order to optimize the performance of the neural network, preliminary tests were done to
 234 determine the optimum utilization of the meteorological input variables. These test were done to
 235 determine if the “forecast” or the “observed” meteorology, or a combination of the two, should be
 236 used as input variables.

237 The forecast time window is midnight-to-midnight EDT for the forecast day, while the time
 238 window with the observed data is midnight to 5PM EDT the day the forecast is released.

239 For the PBLH, the forecast value is always used as the input. One NN design employed only the
 240 forecast meteorological values as inputs. The second design utilized a combination of the forecast
 241 and the observed data, by subtracting the eight observation datasets from the eight forecast datasets.
 242 This first NN architecture uses the meteorological values as predictors, while the second design uses
 243 metrological trends as predictors. We note that this comparison does not affect the number of inputs
 244 used, allowing for a direct comparison of information content.

245 In scenario 1, where only the MET forecasts are used, we use the following inputs, where *i*
 246 represents the indices for time windows for the observation day, and *j* represents the indices for time
 247 windows for the forecast day (from the NARR forecasts), the NN inputs design is:

248

$$\begin{array}{llll}
 PM_{2.5}(i) & i = 1:5 & time\ window\ (i) = (i - 1) \times 3 : i \times 3 & \text{(Field measurements)} \\
 MET_{forecast}(j) & j = 1:8 & time\ window\ (j) = (j - 1) \times 3 : j \times 3 & \text{(NARR Forecasts)} \\
 PBLH(j) & j = 1:8 & time\ window\ (j) = (j - 1) \times 3 : j \times 3 & \text{(NARR Forecasts)}
 \end{array}$$

249 In scenario 2, where the differential between the observation day and forecast day of the MET
 250 variables are used, the architecture for the NN inputs is:

251

$$\begin{array}{llll}
 PM_{2.5}(i) & i = 1:5 & time\ window\ (i) = (i - 1) \times 3 : i \times 3 & \text{(Field measurements)} \\
 MET_{forecast}(j) - & j = 1:8 & time\ window\ (j) = (j - 1) \times 3 & \text{(NARR Forecasts)} \\
 MET_{observed}(i) & i = 1:8 & : j \times 3 & \text{(NARR Observations)} \\
 & & time\ window\ (i) = (i - 1) \times 3 : i \times 3 & \\
 PBLH(j) & j = 1:8 & time\ window\ (j) = (j - 1) \times 3 : j \times 3 & \text{(NARR Forecasts)}
 \end{array}$$

252 To show the robustness of the NN, the data used for training the neural networks came from
 253 2011-2015 alone, while the network was tested with data from 2016. In both scenarios, the targets for
 254 the NN were taken to be the complete set of PM_{2.5} over all time windows of the forecast day:
 255

256

$$\begin{array}{llll}
 \text{Targets: } PM_{2.5}(j) & j = 1:8 & time\ window\ (j) = (j - 1) \times 3 : j \times 3 & \text{(Field measurements)}
 \end{array}$$

257 3.2.1.2 Neural Network Training Approach

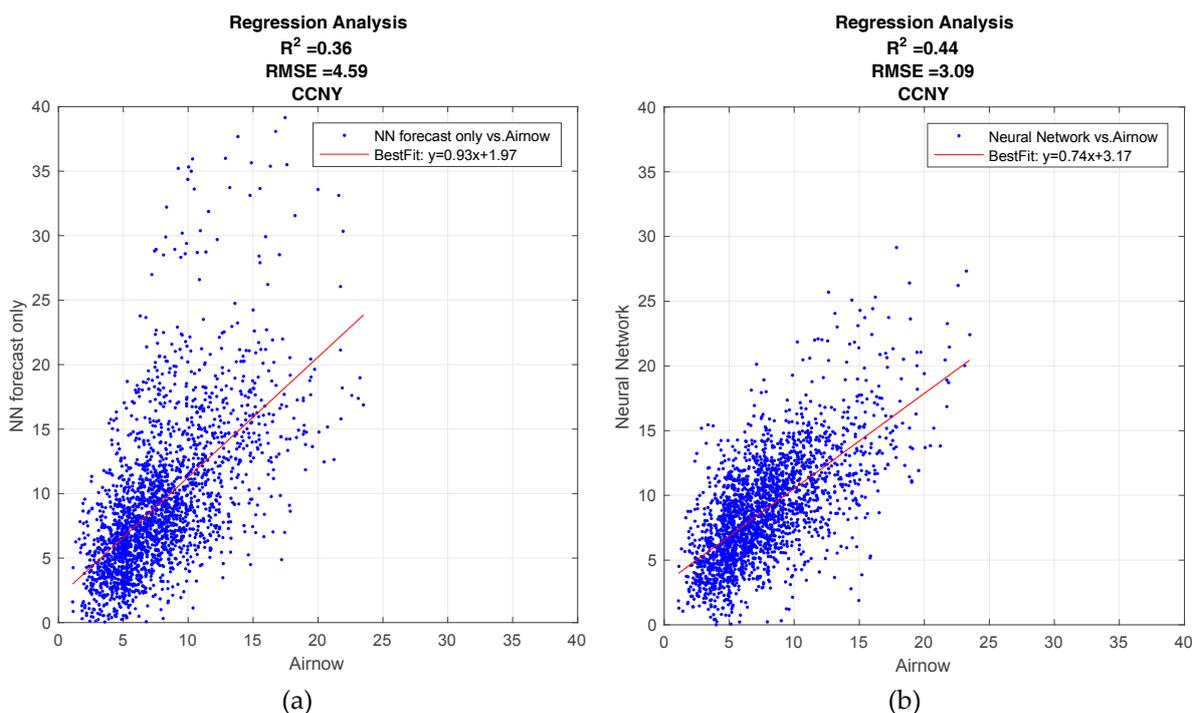
258 In developing a NN PM_{2.5} forecast for all of New York State (NYS), we needed to take into
 259 account the very different emission sources, and to a lesser extent the meteorological conditions,
 260 between New York City (NYC) and the other sites in NYS. We found that the best solution is to design
 261 two different neural networks. The first is trained only over NYC sites, while the second is trained
 262 for the rest of NYS. It is important to note that we do not try to build a unique NN for every station,
 263 since this is not a useful approach for local agencies. PM and Meteorological data from 2011-2015,
 264 were used for training.

265 For NYC, since the stations are very close to each other, the NN was trained with spatial mean
 266 values of the ground PM monitors and NARR meteorological datasets. For NYS, all the PM and
 267 meteorological data from each site outside of NYC were used. Some site-specific information was
 268 implicitly included by using the surface pressure as inputs, which provides some indicator of surface
 269 elevation.

270 The neural network was developed using the MATLAB Neural Network Toolbox [14]. The
 271 Levenberg-Marquardt network was deployed using 10 hidden nodes. The break down for the NN
 272 input data is: 70% training, 15% validation, and 15% testing. Because the sample set of training,
 273 validation, and testing is divided randomly over the entire dataset, accuracy of the NN was
 274 determined by testing each network over 2016 data only, a time window that was not included in
 275 training. Once the NN function was created, the 2016 meteorological and PM data was passed
 276 through the network, and the outputs were stored with the date-time and station location as indices.

277 3.2.1.3 Neural Network Scenario Results

278 Figure 4(a) shows the performance of the NN using the forecast metrological data as inputs,
 279 while figure 4(b) shows the performance of the NN using the difference between the forecast and the
 280 current days measurements. The NN utilizing the difference configuration is clearly better, with a
 281 higher R² value, 0.44 compared to 0.36, and a lower root-mean-square value, 3.09 compared to 4.59.
 282 In addition, there are substantially less anomalous high PM_{2.5} forecasts. Since this improvement was
 283 seen in all test cases, we only used scenario 2, (differential meteorology) NN configuration. From
 284 these results, we see that meteorological trends are better indicators of PM_{2.5} than meteorology alone.
 285 This appears to us to be a reasonable result since the meteorology trend better isolates particular
 286 mesoscale conditions, which is known to be a significant factor in boundary layer dynamics.
 287

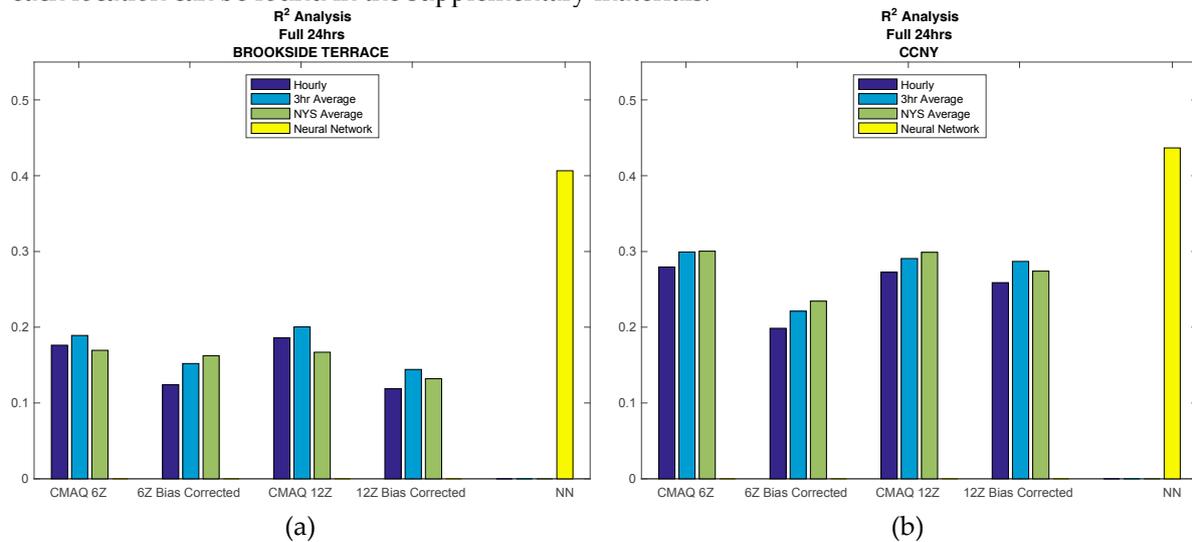


288 **Figure 4.** Results from the regression analysis to maximize Neural Network performance for the
 289 different scenarios. (a) NN designed with the forecast meteorological data (Scenario 1); (b) NN
 290 designed by taking the difference between the forecast and the current days measurements (Scenario
 291 2)

292 3.3. Results

293 3.3.1. Neural Network and CMAQ Comparison

294 The R^2 value for CMAQ and the NN, both compared to AirNow observations, is computed for
 295 each forecast model and for each location. As a representative example of the overall performance,
 296 the R^2 value for NYC, represented by CCNY, is compared to NYS, represented by Brookside Terrace,
 297 a non-NYC, non-urban station, and these results are displayed in Figure 5. The individual results for
 298 each location can be found in the supplementary materials.



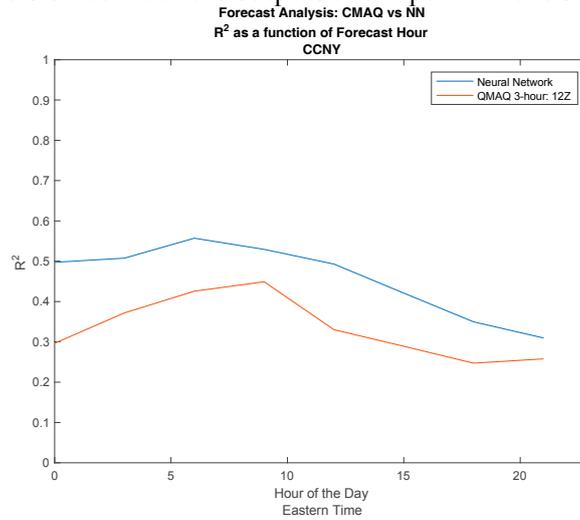
299 **Figure 5.** Regression analysis is computed for the comparison between AirNow observations
 300 and the various prediction models. The R^2 value for each model is plotted in the figure above to
 301 compare CMAQ to the NN. The CMAQ model includes the different release times as well as bias
 302 compensated vs. uncompensated runs. In addition, different time and spatial averaging of CMAQ is
 303 considered at each location. (a) Brookside Terrace, representative of non-NYC; (b) CCNY,
 304 representative of NYC.

305 From Figure 5 above, it can be seen that the most accurate forecast model is the neural network
 306 for both NYS and NYC over any of the CMAQ forecasts studied. Regarding CMAQ, we note better
 307 performance for NYC than for non-urban areas. This is in contrast to the neural network, where there
 308 is very little variation in the results for locations that are urban versus non-urban, indicating that
 309 locational inputs in the model, such as the surface pressure, improves forecasting skill.

310 In addition, for all cases, it can be seen that taking the time average improves the CMAQ results.
 311 Furthermore, the spatial averaging over NYS (with 1-hour time sampling) shows more improvement
 312 in most NYC cases and some non-NYC cases as well. These results indicate the possibility that the
 313 best use for CMAQ forecasting is on a regional level. **This is supported from the 12km grid cell
 314 resolution for CMAQ, a cell size typical for regional analysis.**

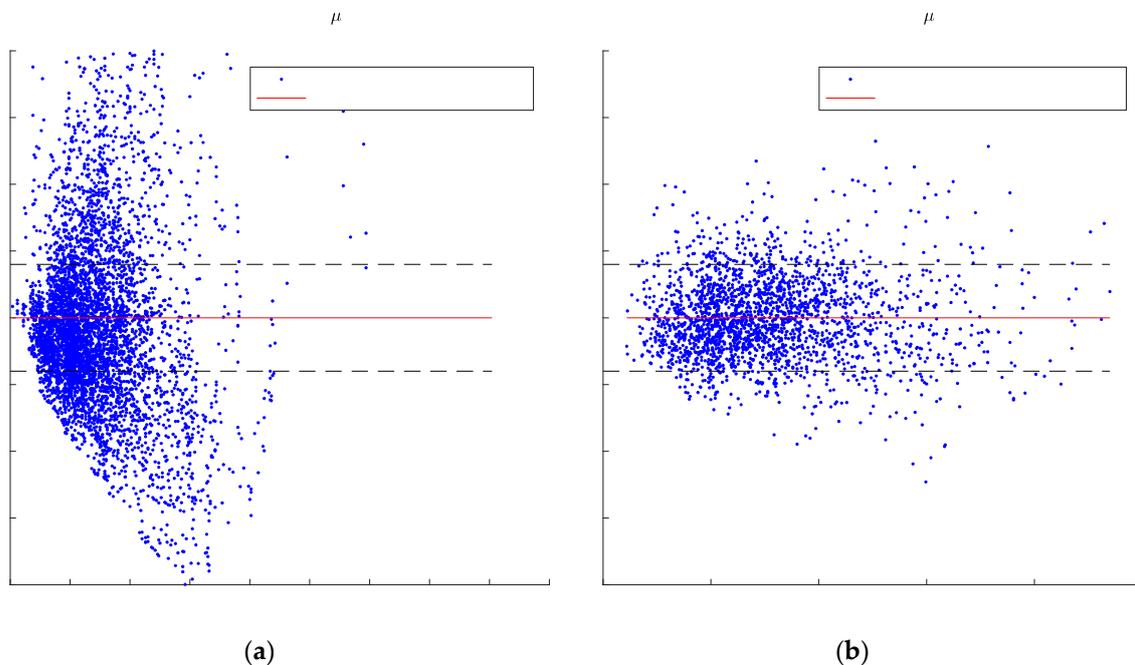
315 We note again that the different release times for CMAQ has almost no effect on the forecast
 316 accuracy. In Figure 6, we compared the diurnal performance of the NN to the CMAQ model. The
 317 most apparent result is the dramatic improvement of the NN during the night and morning hours,
 318 where the CMAQ model has the most difficulty. This is clearly due to the machine learning approach
 319 where the time differences, the inputs, and forecast periods have a dramatic effect on output
 320 performance.

321 This also explains the general downward trend, where performance tails off in the late afternoon
 322 and becomes closer to the CMAQ performance. This can be expected, because larger time delays
 323 should lead to more dispersion between the outputs and input PM levels.



324 **Figure 6.** Comparing the effect of different release times for the NN in comparison to CMAQ by
 325 plotting the R² value as a function of time of day

326 Figure 7 below shows the residual results for CMAQ in comparison to the neural network. For
 327 CMAQ, as noted above, there seems to exist a non-random bias pattern, where CMAQ generally over
 328 predicts for low and high PM values, and under predicts for medium values. This pattern seems to
 329 indicate that the CMAQ model may not capture all of the underlying variability factors. On the other
 330 hand, for the neural network, the behavior of the residuals is clearly stochastic in nature.
 331



332 **Figure 7.** Residual analysis. The standard deviation of PM_{2.5} from the AirNow ground
 333 monitoring sites was calculated to be 4μg/m³, therefore, +/-4μg/m³ was used for the error bounds (a)
 334 CMAQ; (b) Neural Network

335 **We find that an optimized NN approach generally results in a more accurate prediction of future**
 336 **pollution levels, as compared to CMAQ, for a single grid cell (resolution 12km).**

337 3.3.1. Heavy Pollution Transport Events

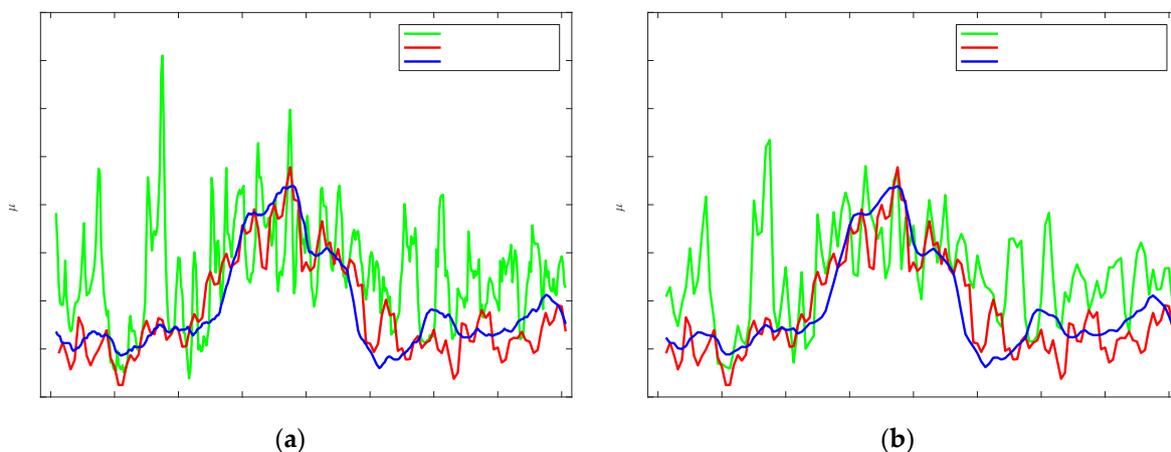
338 Because the neural network is data-driven, the network performs better when the most up-to-
339 date inputs are used. This explains the degradation of performance with time, as seen in Figure 6. In
340 the current design of the neural network, we only used five $PM_{2.5}$ inputs, instead of maximum
341 possible in a 24-hour period, eight. In the training of the NN, there were very few extreme event
342 cases, $PM_{2.5} > \mu g/m^3$. The lack of suitable training statistics for these events causes the NN approach
343 to have difficulty in adjusting to the sharp contrast with the onset of the event.

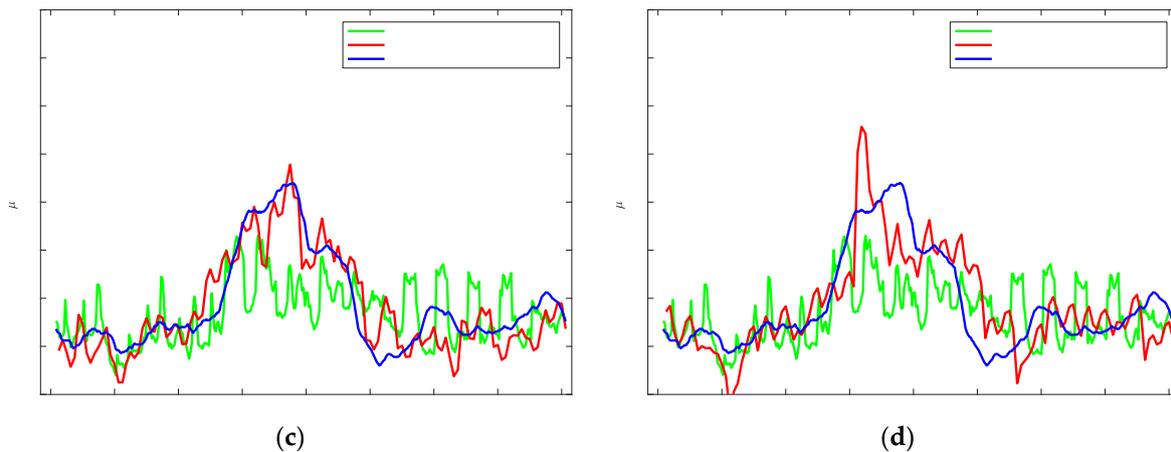
344 Therefore, a second neural network was trained with the same design as the neural network
345 illustrated above; however, this neural network produces a 24-hour forecast at 5PM for the time
346 period, 5PM – 5PM (instead of a next day 24-hour midnight-to-midnight forecast). This neural
347 network uses all eight PM measurements, because there is no lag time between the release time and
348 the first forecast hour. This neural network, referred to as NN Continuous, was not used in the
349 statistical analysis for the different forecast models (because the 24-hour forecast period is different
350 than the forecast analysis above), but is being explored in the extreme event cases. The reason for
351 developing this continuous neural network is to determine if the continuous nature of the network
352 produces better results in extreme pollution events.

353 To explore the behavior of the different models under high pollution transport conditions, the
354 forecasts coinciding with the wildfires of Fort McMurray in Alberta, Canada were analyzed. **The
355 wildfire started on May 1, 2016, and was declared under control on July 5, 2016. Although the wildfire
356 lasted for over two months, evidence of increased $PM_{2.5}$ surface levels in NYC resulting from the
357 wildfire were detected on May 9, and on May 25. On these dates, instances of aloft plume intrusions
358 and the mixing down into the planetary-boundary layer were observed by a ceilometer and a Raman-
359 Mie lidar [16]. In Figure 8, we plot the CMAQ and NN model forecasts, focusing on the transport
360 intrusions into NYC on May 25.**

361 The first thing to notice in Figure 8(a), is the oscillations in the CMAQ model, and to notice how
362 these oscillations smooth out in 8(b) and 8(c), where the three-hour time average and the New York
363 State spatial average are tested respectively. It is logical that for heavy transport cases, domain
364 averaging helps decrease oscillations; however, we still see significant underestimation of the event.

365 This is the first case where we analyze the behavior of the continuous neural network. Looking
366 at Figures 8(c) and 8(d), it is clear that the continuous neural network is able to respond to the trend
367 of the high pollution event faster, and more accurately, than the standard neural network.
368





369 **Figure 8.** NYC surface PM_{2.5} levels affected from the wildfires of Alberta Canada 2016. The plots
 370 focus on the aloft plumes mixing down into the PBL on May 25. The plots show different models vs.
 371 AirNow observation (a) CMAQ Biased hourly, NN continuous; (b) CMAQ Biased 3-hour average,
 372 NN continuous; (c) CMAQ Biased state average, NN continuous; (d) CMAQ Biased state average,
 373 standard Neural Network

374 4. Conclusions

375 In this paper, we first made a baseline assessment of the V4.6 CMAQ forecasts, and found
 376 significant dispersion as well as a tendency for the model to over estimate the ground truth field
 377 measurements. Even in the bias corrected case, the residuals error in the model was found to have
 378 significant bias patterns, indicating that there are predictors not included in the model that could
 379 significantly improve the results.

380 These results motivated the development of data driven approaches such as a NN. In developing
 381 a data driven NN next day forecast model, we found a general improvement of performance when
 382 using prior PM_{2.5} inputs together with the difference between present and next day meteorological
 383 parameter forecasts. This “differential NN” approach performed significantly better than if we used
 384 only the future forecast variables, indicating that meteorological pattern trends are important
 385 indicators.

386 Using this NN architecture, we then made extensive regression based comparisons between
 387 CMAQ next day forecast models and regionally trained NN next day forecasts for the NYS and NYC
 388 regions. In general, we found that the NN results are a significant improvement over the CMAQ
 389 forecasts in all cases. These comparisons were made to be consistent with state agencies where
 390 forecasts should be available by 5PM. In addition, we also made a diurnal comparison, which
 391 illustrated that; the NN approach had superior forecasting skills during the early part of the day but
 392 degraded smoothly as the forecast time increased. By mid-day, the differences between the two
 393 approaches was much closer

394 To improve the CMAQ forecasts, we found limited improvement when spatial averaging is
 395 extended beyond the single pixel 12km resolution to all of New York State. Even in this case, the NN
 396 results were generally more accurate.

397 Finally, we focused on forecast performance for transported high pollution events such as
 398 Canadian wildfires. In these cases, we found that the CMAQ forecasts had large temporal
 399 fluctuations, which could hide most of the event. In this case, significant improvement was obtained
 400 when using state averaged bias corrected outputs; however, in general, the smoothed results
 401 underestimate the local PM_{2.5} measurements.

402 In this application, we found the neural network approach provides a reasonably smooth
 403 forecast, although the transition from a clean state to a polluted state is very poor. **Nevertheless, the
 404 standard NN performed better than CMAQ in this scenario. Further improved results for the NN**

405 were obtained in the transition period when the forecast time of the NN was reduced (NN
406 continuous), making the transition from training to testing continuous.

407 4.1 Future Work

408 While the continuous NN does adjust quickly to the sharp contrast in transport events, this
409 design limits the scope of the forecast period. Clearly, local data alone is not ideal for this application.
410 Non-local data that can identify high pollution events and assesses their potential mixing with our
411 region is needed. As a preliminary analysis, we explored the use of a combination of HYSPLIT Air
412 Parcel Trajectories with GOES satellite Aerosol Optical Depth (AOD) retrievals to improve the NN.
413 In particular, we analyzed the use of these tools to quantify the relative AOD levels for all air parcels
414 that reach our target area. We found that by properly counting the trajectories weighted by the AOD,
415 a good correlation was seen between the relative AOD and the PM_{2.5} levels. Therefore, we believe
416 that using the relative AOD metric as an additional input factor can make improvements in the NN
417 approach. When GOES-R AOD retrievals, with high data latency and multispectral inversion
418 capabilities [17-18], become available, we plan to incorporate these AOD metrics as predictors in the
419 NN.

420 **Supplementary Materials:** The following are available online at www.mdpi.com/link, Figure S1: Regression
421 Analysis

422 **Acknowledgments:** The authors gratefully acknowledge the NOAA Air Resources Laboratory (ARL) for the
423 provision of the HYSPLIT transport and dispersion model used in this publication. We further gratefully
424 acknowledge Mr. Jeff McQueen for providing the V4.6 CMAQ forecasts. Finally, S. Lightstone would like to
425 acknowledge partial support of this project through a NYSERDA EMEP Fellowship award.

426 **Author Contributions:**

427 Mr. Samuel Lightstone was responsible for the NN experiments and the CMAQ assessment; Mr. Lightstone also
428 wrote the paper. Dr. Barry Gross was responsible for design of the different experiments as well as conceiving
429 the use of the GOES AOD and HYSPLIT trajectories to estimate transported AOD as a predictor of PM_{2.5}. Dr.
430 Fred Moshary provided significant critical assessment and suggestions of all results

431 **Conflicts of Interest:** The authors declare no conflict of interest.

432

433 **Appendix A: Datasets**

434

Table A1. NYSDEC Station Information

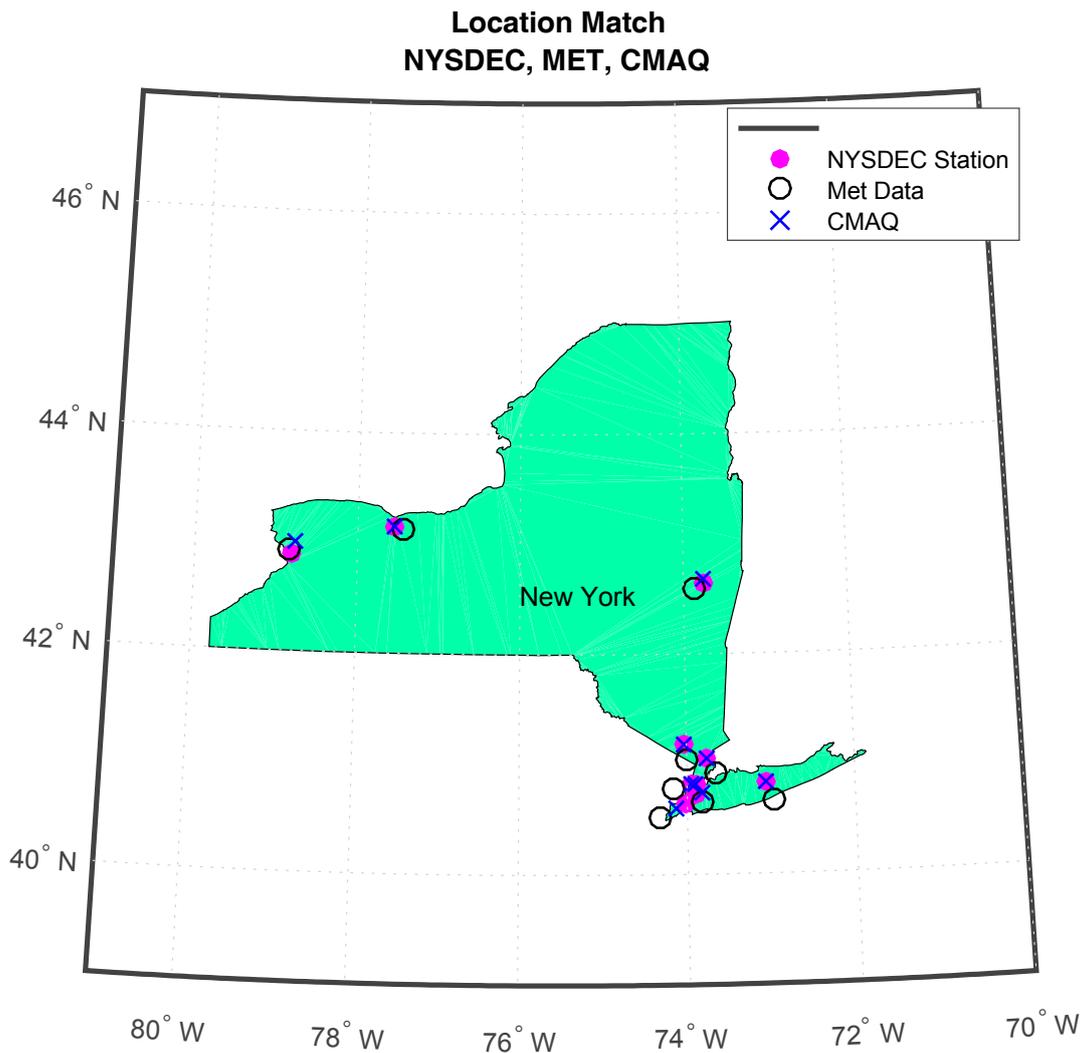
NYSDEC ID	Station Name	Latitude	Longitude	Land Type
360010005	Albany County Health Dept	42.6423	-73.7546	Urban
360050112	IS 74	40.8155	-73.8855	Suburban
360291014	Brookside Terrace	42.9211	-78.7653	Suburban
360551007	Rochester 2	43.1462	-77.5482	Urban
360610135	CCNY	40.8198	-73.9483	Urban
360810120	Maspeth Library	40.7270	-73.8931	Suburban
360850055	Freshkills West	40.5802	-74.1983	Suburban
360870005	Rockland County	41.1821	-74.0282	Rural
361030009	Holtsville	40.8280	-73.0575	Suburban
361192004	White Plains	41.0519	-73.7637	Suburban

435

Table A2. CMAQ Grid Cell Information

Name	Abbreviation	Latitude	Longitude	Land Type
Amherst	AMHT	42.99	-78.77	Suburban
CCNY	CCNY	40.82	-73.95	Urban
Holtsville	HOLT	40.83	-73.06	Suburban
IS 52	IS52	40.82	-73.90	Suburban
Loudonville	LOUD	42.68	-73.76	Urban
Queens College 2	QC2	40.74	-73.82	Suburban
Rochester Pri 2	RCH2	43.15	-77.55	Urban
Rockland County	RCKL	41.18	-74.03	Rural
S. Wagner HS	WGHS	40.60	-74.13	Urban
White Plains	WHPL	41.05	-73.76	Suburban

436



437
438

Figure A1. This map shows the proximity of the ground NYSDEC stations to the NARR meteorological data, and the CMAQ forecast data.

439 References

- 440 1. Pope III, C. Arden, and Douglas W. Dockery. "Health effects of fine particulate air pollution: lines that
441 connect." *Journal of the air & waste management association* 56.6 (2006): 709-742.
- 442 2. Pope III, C. Arden, et al. "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine
443 particulate air pollution." *Jama* 287.9 (2002): 1132-1141.
- 444 3. Weber, Stephanie A., et al. "Assessing the impact of fine particulate matter (PM 2.5) on respiratory-
445 cardiovascular chronic diseases in the New York City Metropolitan area using Hierarchical Bayesian Model
446 estimates." *Environmental Research* 151 (2016): 399-409.
- 447 4. Rattigan, Oliver V., et al. "Multi-year hourly PM 2.5 carbon measurements in New York: diurnal, day of
448 week and seasonal patterns." *Atmospheric Environment* 44.16 (2010): 2043-2053.
- 449 5. Byun, Daewon, and Kenneth L. Schere. "Review of the governing equations, computational algorithms,
450 and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling
451 system." *Applied Mechanics Reviews* 59.2 (2006): 51-77.
- 452 6. McKeen, S., et al. "Evaluation of several PM2.5 forecast models using data collected during the
453 ICARTT/NEAQS 2004 field study." *Journal of Geophysical Research: Atmospheres* 112.D10 (2007).
- 454 7. Yu, Shaocai, et al. "Evaluation of real - time PM2.5 forecasts and process analysis for PM2.5 formation
455 over the eastern United States using the Eta - CMAQ forecast model during the 2004 ICARTT
456 study." *Journal of Geophysical Research: Atmospheres* 113.D6 (2008).
- 457 8. Huang, Jianping, et al. "Improving NOAA NAQFC PM2.5 Predictions with a Bias Correction
458 Approach." *Weather and Forecasting* 32.2 (2017): 407-421.

- 459 9. Doraiswamy, Prakash, et al. "A retrospective comparison of model-based forecasted PM2.5 concentrations
460 with measurements." *Journal of the Air & Waste Management Association* 60.11 (2010): 1293-1308.
- 461 10. Gan, Chuen-Meei, et al. "Application of active optical sensors to probe the vertical structure of the urban
462 boundary layer and assess anomalies in air quality model PM 2.5 forecasts." *Atmospheric environment* 45.37
463 (2011): 6613-6621.
- 464 11. <http://www.emc.ncep.noaa.gov/mmb/aq/sv/grib/>
- 465 12. <http://www.emc.ncep.noaa.gov/mmb/aq/AQChangelog.html>
- 466 13. Lee, Pius, et al. "NAQFC developmental forecast guidance for fine particulate matter (PM2.5)." *Weather and
467 Forecasting* 32.1 (2017): 343-360.
- 468 14. McKeen, S., et al. "An evaluation of real - time air quality forecasts and their urban emissions over eastern
469 Texas during the summer of 2006 Second Texas Air Quality Study field study." *Journal of Geophysical
470 Research: Atmospheres* 114.D7 (2009).
- 471 15. MATLAB and Neural Network Toolbox Release 2016a, The MathWorks, Inc., Natick, Massachusetts,
472 United States
- 473 16. Wu, Yonghua, et al. "Wildfire Smoke Transport and Impact on Air Quality Observed by a Multi-
474 Wavelength Elastic-Raman Lidar and Ceilometer in New York City," to be published in the Proceedings of
475 the 28th International Laser Radar Conference, Bucharest, Romania, June 2017.
- 476 17. Anderson, William, et al. "The Geostationary Operational Satellite R Series SpaceWire Based Data System."
477 (2016).
- 478 18. Lebar, William, et al. "Post launch calibration and testing of the Advanced Baseline Imager on the GOES-
479 R satellite." (2016).



© 2017 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>)