

## Comparing CMAQ Forecasts with a Neural Network Forecast Model for PM<sub>2.5</sub> in New York

Samuel Lightstone\*, Barry Gross, and Fred Moshary

Optical Remote Sensing Lab, City College of New York, New York City, NY, USA

### 1. INTRODUCTION

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

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

The U.S. Environmental Protection Agency (EPA) established the National Ambient Air Quality Standards (NAAQS), which regulate levels of pollutants such as fine particulate matter. The New York State Department of Environment Conservation (NYSDEC) operates ground stations for monitoring PM<sub>2.5</sub> and speciation throughout NY State [4]. However, surface sampling is expensive and existing networks are limited and sparse. This results in data gaps that can affect the ability to forecast PM<sub>2.5</sub> over a 24-hour period. The EPA developed the Models-3 Community Multi-scale Air Quality system (CMAQ), to provide 24-48 hour air quality forecasts. CMAQ provides an investigative tool to explore proper emission control strategies. CMAQ has been the standard for modeling air pollution for nearly two decades because of its ability to independently model different pollutants while describing the atmosphere using “first-principles” [5].

In their studies, *McKeen et al* and *Yu et al* evaluate the accuracy of CMAQ forecasts [6-7]. To do so, they use the CMAQ 1200UTC (Version 4.4) forecast model. They observe the midnight-to-midnight local time forecast and compare the hourly and daily average forecasts to the ground monitoring stations. *McKeen et al* [6] observed

minimal diurnal variations of PM<sub>2.5</sub> at urban and suburban monitor locations, with a consistent decrease of PM values between 0100 and 0600 local time. However, the CMAQ model showed significant diurnal variations, leading *McKeen et al* to conclude that aerosol loss during the late night and early morning hours has little effect on PM<sub>2.5</sub> concentrations, while the CMAQ model does not account for this. Therefore, in addition to testing the hourly CMAQ forecast for a 24-hour period, we focus on the daytime window for two reasons: 1) to assess the accuracy of CMAQ when aerosols do not play a reduced roll in forecasting, 2) the forecast should predict the air quality during the time of maximum human exposure.

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

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

In our present assessment of the current operational CMAQ forecast model (Version 4.6), we differ from the regional studies above in the following ways: Firstly, in addition to the 1200UTC forecast, we evaluated the 0600UTC forecast for the same period to determine if release time affects the CMAQ forecast. Second, we focused on specific locations, both rural and urban, to assess

---

\*Corresponding author: Samuel D. Lightstone; e-mail: [slights01@citymail.cuny.edu](mailto:slights01@citymail.cuny.edu)

the potential of CMAQ as a localized forecasting tool. In addition, we revisited the forecast potential of CMAQ for high pollution events, to determine if these events are generally caused by transport, or by local emissions. Finally, we tailor the forecast comparisons to focus on the potential of providing next day forecasts using data prior to 5PM of the previous day, since this is an operational requirement for the state environmental agencies.

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

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

## 2. Datasets

### 2.1 Models

The CMAQ V4.6 (CB05 gas-phase chemistry) with 12km horizontal resolution was used for this paper. The CMAQ product for meteorology predictions used is the North American Model Non-Hydrostatic Multi-Scale Model (NAM-NMMB). This version was made available starting February 2016. The CMAQ data used for this paper is from February 1, 2016 until October 31, 2016. The data can be accessed from reference 11, and the model description can be found in references 12 and 13.

The CMAQ model used has a few different configurations: release times of 0600 UTC and 1200 UTC, and each release time has a standard forecast as well as a bias corrected forecast. The analog ensemble method is used for bias corrections. The idea is to look at similar weather

patterns for the forecast period, and statistically correct the numerical  $PM_{2.5}$  forecast based on historical errors. The analog ensemble method is described in detail in *Huang, et al* [8]. For each release time, CMAQ provides a 48-hour forecast. The release time of 0600 UTC and 1200 UTC (2AM and 8AM EDT) does not give the public enough time to react to the forecast on the same day as the release. Consequently, for the 0600 UTC release time, the forecast hours 22 – 45 were used, and for the release time of 1200 UTC the forecast hours of 16 – 39 were used. This allowed us to construct a complete 24-hour diurnal period for the forecast time window, which facilitated comparison with the field station data.

### 2.2 Ground-based $PM_{2.5}$ Observations

$PM_{2.5}$  ground data is collected from the EPA's AirNow, which collects NYSDEC monitoring station measurements in real time. The station data used for the forecast experiments in this article are from the New York State stations listed in table A1, from January 1, 2011 until December 31, 2016. To assess the accuracy of CMAQ model forecasts, matching the model to the ground monitoring station is necessary. To do this, we use the ground NYSDEC stations that lay within the CMAQ grid cell only. Ground stations that are not found in a CMAQ grid cell were not used for comparison; therefore, no spatial interpolation was done on the model results while mapping the model or meteorological data to the AirNow ground stations.

### 2.2 Meteorological Observations

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

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

The locational data-points for the NYSDEC stations, the CMAQ PM<sub>2.5</sub> forecasts, and NARR meteorological reanalysis are depicted in Figure 1.

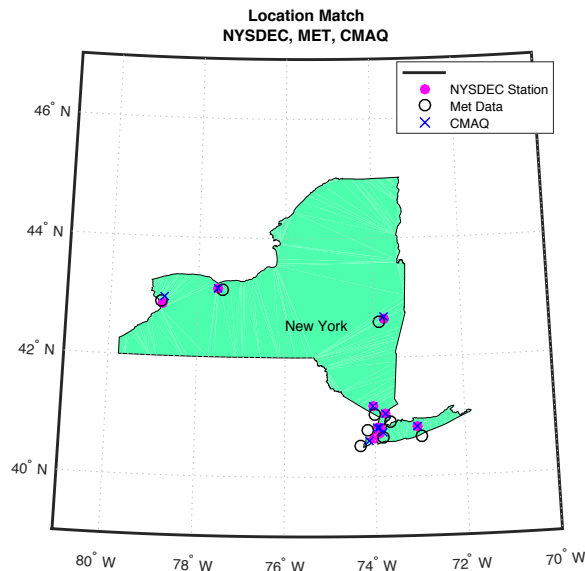


Fig. 1. This map shows the proximity of the ground NYSDEC stations to the NARR meteorological data, and the CMAQ forecast data.

### 3. Development of the Neural Network

As stated above, the accurate prediction of PM<sub>2.5</sub> values is crucial for air quality agencies, so that they could alert the public of the severity and duration of a high pollution event. Therefore, it is imperative that the forecast predictions are released to the public the day before the event. For this paper, we chose 5PM as a target for the forecast release time. Therefore, we ensure that all the methods tested, utilize factors that are available to the state agency prior to 2100 UTC (5PM EDT).

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

In order to optimize the performance of the neural network, preliminary tests were done to determine the optimum utilization of the meteorological input variables. These tests determined that a combination of the “forecast” and the “observed” meteorology should be used as input variables.

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

For the PBLH, the forecast value is always used as the input. For the rest of the meteorological inputs, a combination of the forecast and the observed data was utilized by subtracting the eight observation datasets from the eight forecast datasets. This NN architecture uses meteorological trends as predictors.

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

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

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

#### 3.1 Neural Network and CMAQ Comparison

The R<sup>2</sup> value for CMAQ and the NN, both compared to AirNow observations, is computed for each forecast model and for each location. As a representative example of the overall performance, the R<sup>2</sup> value for NYC, represented by CCNY, is

compared to NYS, represented by Brookside Terrace, a non-NYC, non-urban station, and these results are displayed in Figure 2.

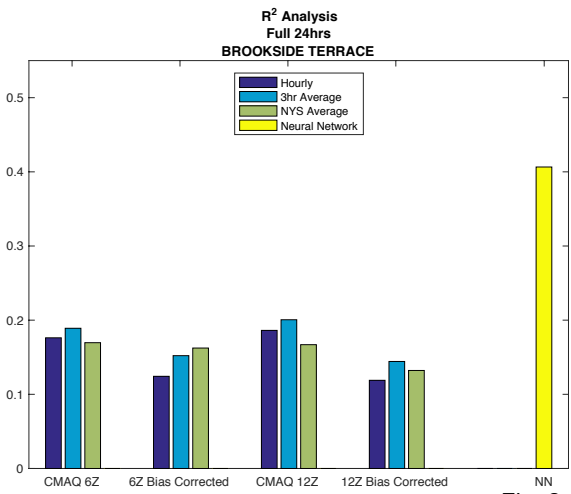


Fig. 2a

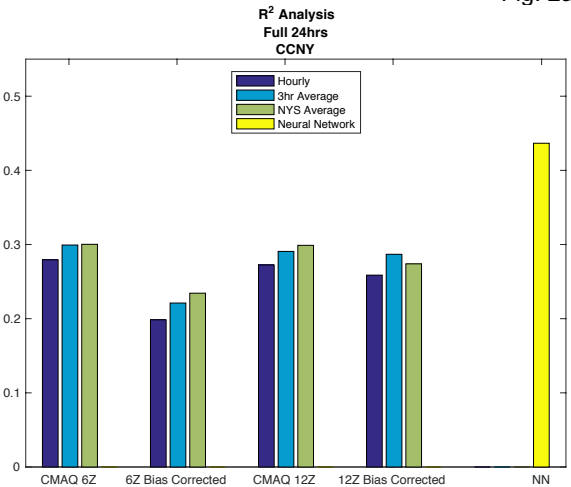


Fig 2b

Fig. 2. Regression analysis is computed for the comparison between AirNow observations and the various prediction models, for the complete months of February to October 2016. The  $R^2$  value for each model is plotted in the figure above to compare CMAQ to the NN. The CMAQ model includes the different release times as well as bias compensated vs. uncompensated runs. In addition, different time and spatial averaging of CMAQ is considered at each location. (a) Brookside Terrace, representative of non-NYC; (b) CCNY, representative of NYC.

From Figure 2 above, it can be seen that the most accurate forecast model is the neural network for both NYS and NYC over any of the CMAQ forecasts studied. Regarding CMAQ, we note better performance for NYC than for non-urban areas. This is in contrast to the neural network, where there is very little variation in the results for locations that are urban versus non-urban,

indicating that locational inputs in the model, such as the surface pressure, improves forecasting skill.

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

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

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

### 3.2 Heavy Pollution Transport Events

Because the neural network is data-driven, the network performs better when the most up-to-date inputs are used. In the current design of the neural network, we only used five  $PM_{2.5}$  inputs, instead of maximum possible in a 24-hour period, eight. In the training of the NN, there were very few extreme event cases,  $PM_{2.5} > 25\mu g/m^3$ . The lack of suitable training statistics for these events causes the NN approach to have difficulty in adjusting to the sharp contrast with the onset of the event.

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

The differences between these neural network designs can be seen in Figure 3.

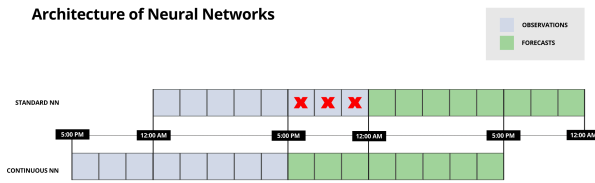


Fig. 3. Architecture of the two Neural Networks utilized highlight the difference in input values and time-period.

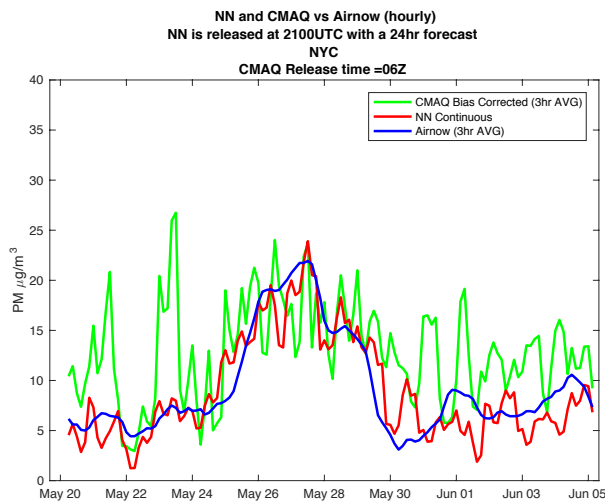


Fig. 4a

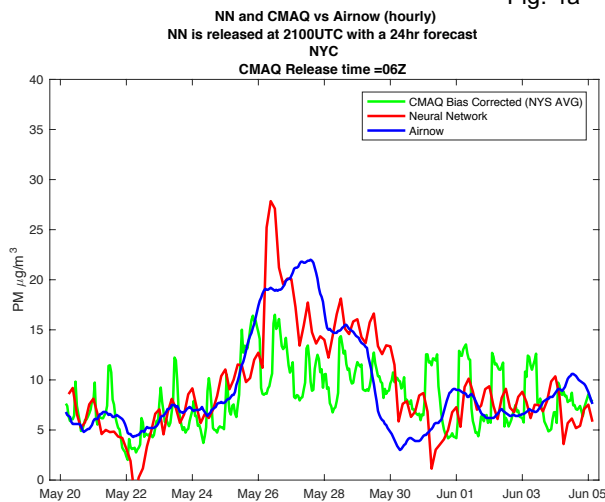


Fig. 4b

Fig 4. NYC surface PM<sub>2.5</sub> levels affected from the wildfires of Alberta Canada 2016. The plots focus on the aloft plumes mixing down into the PBL on May 25. The plots show different models vs. AirNow observation (a) CMAQ Biased 3-hour average, NN continuous; (b) CMAQ Biased state average, standard Neural Network

To explore the behavior of the different models under high pollution transport conditions, the forecasts coinciding with the wildfires of Fort McMurray in Alberta, Canada were analyzed. The wildfire started on May 1, 2016, and was declared under control on July 5, 2016. Although the wildfire lasted for over two months, evidence of increased PM<sub>2.5</sub> surface levels in NYC resulting from the wildfire were detected on May 9, and on May 25. On these dates, instances of aloft plume intrusions and the mixing down into the planetary-boundary layer were observed by a ceilometer and a Raman-Mie lidar [15]. In Figure 4, we plot the CMAQ and NN model forecasts, focusing on the transport intrusions into NYC on May 25.

The first thing to notice in Figure 4(a), is the oscillations in the CMAQ model even after time averaging over a three-hour window, and to notice how these oscillations smooth out in 4(b) when the New York State spatial average is tested. It is logical that for heavy transport cases, domain averaging helps decrease oscillations; however, we still see significant underestimation of the event.

This is the first case where we analyze the behavior of the continuous neural network. Looking at Figures 4(a) and 4(b), it is clear that the continuous neural network is able to respond to the trend of the high pollution event faster, and more accurately, than the standard neural network.

#### 4. Conclusions

Prior assessments of the CMAQ forecasting model has found significant dispersion as well as a tendency for the model to overestimate the ground truth field measurements. Even in the bias corrected case, the residuals error in the model was found to have significant bias patterns, indicating that there are predictors not included in the model that could significantly improve the results.

These results motivated the development of data driven approaches such as a NN. In developing a data driven NN next day forecast model, we found a general improvement of performance when using prior PM<sub>2.5</sub> inputs together with the difference between present and next day meteorological parameter forecasts. This “differential NN” approach indicates that meteorological pattern trends are important indicators.

Using this NN architecture, we then made extensive regression based comparisons between CMAQ next day forecast models and regionally trained NN next day forecasts for the NYS and NYC regions. In general, we found that the NN results are a significant improvement over the CMAQ forecasts in all cases. These comparisons were

made to be consistent with state agencies where forecasts should be available by 5PM.

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

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

In this application, we found the neural network approach provides a reasonably smooth forecast, although the transition from a clean state to a polluted state is very poor. Nevertheless, the standard NN performed better than CMAQ in this scenario. Further improved results for the NN were obtained in the transition period when the forecast time of the NN was reduced (NN continuous), making the transition from training to testing continuous. More details can be found in ref [16].

## 5. Future Work

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

**6. Acknowledgement:** S. Lightstone would like to acknowledge the support of this work through a NYSERDA EMEP Fellowship.

## 7. References

1. Pope, C.A., III; Dockery, D.W. Health effects of fine particulate air pollution: Lines that connect. *J. Air Waste Manag. Assoc.* 2006, 56, 709–742.
2. Pope, C.A., III; Burnett, R.T.; Thun, M.J.; Calle, E.E.; Krewski, D.; Ito, K.; Thurston, G.D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 2002, 287, 1132–1141.
3. Weber, S.A.; Insaif, T.Z.; Hall, E.S.; Talbot, T.O.; Huff, A.K. Assessing the impact of fine particulate matter (PM<sub>2.5</sub>) on respiratory-cardiovascular chronic diseases in the New York City Metropolitan area using Hierarchical Bayesian Model estimates. *Environ. Res.* 2016, 151, 399–409.
4. Rattigan, O.V.; Felton, H.D.; Bae, M.; Schwab, J.J.; Demerjian, K.L. Multi-year hourly PM<sub>2.5</sub> carbon measurements in New York: Diurnal, day of week and seasonal patterns. *Atmos. Environ.* 2010, 44, 2043–2053.
5. Byun, D.; Schere, K.L. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.* 2006, 59, 51–77.
6. McKeen, S.; Chung, S.H.; Wilczak, J.; Grell, G.; Djalalova, I.; Peckham, S.; Gong, W.; Bouchet, V.; Moffet, R.; Tang, Y.; et al. Evaluation of several PM<sub>2.5</sub> forecast models using data collected during the ICARTT/NEAQS 2004 field study. *J. Geophys. Res. Atmos.* 2007, 112, doi:10.1029/2006JD007608.
7. Yu, S.; Mathur, R.; Schere, K.; Kang, D.; Pleim, J.; Young, J.; Tong, D.; Pouliot, G.; McKeen, S.A.; Rao, S.T. Evaluation of real-time PM<sub>2.5</sub> forecasts and process analysis for PM<sub>2.5</sub> formation over the eastern United States using the Eta-CMAQ forecast model during the 2004 ICARTT study. *J. Geophys. Res. Atmos.* 2008, 113, doi:10.1029/2007JD009226.
8. Huang, J.; McQueen, J.; Wilczak, J.; Djalalova, I.; Stajner, I.; Shafran, P.; Allured, D.; Lee, P.; Pan, L.; Tong, D.; et al. Improving NOAA NAQFC PM<sub>2.5</sub> Predictions with a Bias Correction Approach. *Weather Forecast.* 2017, 32, 407–421.
9. Doraiswamy, P.; Hogrefe, C.; Hao, W.; Civerolo, K.; Ku, J.Y.; Sistla, G. A retrospective comparison of model-based forecasted PM<sub>2.5</sub> concentrations with measurements. *J. Air Waste Manag. Assoc.* 2010, 60, 1293–1308.
10. Gan, C.-M.; Wu, Y.; Madhavan, B.L.; Gross, B.; Moshary, F. Application of active optical sensors to probe the vertical structure of the urban boundary layer and assess anomalies in air quality model PM<sub>2.5</sub> forecasts. *Atmos. Environ.* 2011, 45, 6613–6621.
11. Files in /mmb/qaq/sv/grib. Available online: <http://www.emc.ncep.noaa.gov/mmb/qaq/sv/grib/> (accessed on 01/12/2016).
12. NCEP Operational Air Quality Forecast Change Log. Available online: <http://www.emc.ncep.noaa.gov/mmb/qaq/AQChangelog.html> (accessed on 01/05/2017).
13. Lee, P.; McQueen, J.; Stajner, I.; Huang, J.; Pan, L.; Tong, D.; Kim, H.; Tang, Y.; Kondragunta, S.; Ruminski, M.; et al. NAQFC developmental forecast guidance for fine particulate matter (PM<sub>2.5</sub>). *Weather Forecast.* 2017, 32, 343–360.
14. Demuth, H; Beale, M. *MATLAB and Neural Network Toolbox Release 2016a*; The MathWorks, Inc.: Natick, MA, USA, 1998.
15. Wu, Y.; Pena, W.; Diaz, A.; Gross, B.; Moshary, F. Wildfire Smoke Transport and Impact on Air Quality Observed by a Multi-Wavelength Elastic-Raman Lidar and Ceilometer in New York City. In *Proceedings of the 28th International Laser Radar Conference*, Bucharest, Romania, 25–30 June 2017.
16. Lightstone, S.D., Moshary, F. and Gross, B., 2017. Comparing CMAQ Forecasts with a Neural Network Forecast Model for PM<sub>2.5</sub> in New York. Special Issue "Air Quality Monitoring and Forecasting Atmosphere, 8(9), p.161
17. Anderson, W.; Krimchansky, A.; Birmingham, M.; Lombardi, M. The Geostationary Operational Satellite R Series SpaceWire Based Data System. In *Proceedings of the 2016 International SpaceWire Conference (SpaceWire)*, Yokohama, Japan, 25–27 October 2016.
18. Lehair, W.; Rollins, C.; Kline, J.; Todirita, M.; Kronenwetter, J. Post launch calibration and testing of the Advanced Baseline Imager on the GOES-R satellite. In *Proceedings of the SPIE Asia-Pacific Remote Sensing*, New Delhi, India, 3–7 April 2016.