**J10.2**

# OPERATIONAL USE OF NUMERICAL AIR QUALITY FORECAST MODEL GUIDANCE: CURRENT PRACTICE AND BENCHMARK SKILL

William F. Ryan
Michelle A. Palmer
The Pennsylvania State University, University Park, Pennsylvania

## 1. INTRODUCTION

Forecast guidance from numerical air quality models has been available for operational use for a number of years (McHenry et al., 2004). Forecast guidance for both $O_3$ and $PM_{2.5}$ are available but, due to the complex nature of particle formation and removal processes, $O_3$ forecast models are currently at a higher level of maturity as reflected in forecast skill. As a result, this paper will focus on the operational use of $O_3$ forecast guidance. Numerical $O_3$ models are subject to errors from a variety of sources (Eder et al., 2009). Of particular interest to operational forecasters are systematic forecast errors. These errors can be driven, to name just a few possibilities, by model responses to topography, boundary layer processes, and variations in emissions of precursors. It may be possible to improve the skill of numerical forecast guidance by post-processing methods that reduce the effect of systematic errors. This paper describes and analyzes a number of post-processing methods that are easily implemented in the current operational forecast setting. Because air quality forecasting in the United States is carried out primarily by state and local air quality agencies, often with limited budgets and time constraints, this paper focuses on methods that are relatively inexpensive to design, install and implement.

## 2. DATA AND METHODS

This paper analyzes forecast guidance from the National Air Quality Forecast Capability (NAQFC) $O_3$ forecast model. The NAQFC is one of many $O_3$ forecast models available to the operational forecaster. A partial list is provided in Appendix A.

The development of a national air quality forecast capability was directed by Congress (Energy Policy Act of 2002). The NAQFC model was developed by the National Oceanic and Atmospheric Administration (NOAA), in association with the United States Environmental Protection Agency (USEPA). Forecast guidance was made available for testing beginning in 2004 and the model became operational in September of 2007 (Otte et al., 2005). The NAQFC runs twice daily, initialized at 0600 and 1200 UTC, and covers the continental United States. This paper analyzes the performance of the 1200 UTC model run.

For this study, $O_3$ forecast guidance is collected from the NAQFC model for the Philadelphia metropolitan area (PHL) during the summer seasons (May-early September) of 2007-2009 (Figure 1). The standard predicand used for $O_3$ forecasting is peak domain-wide 8-hour average $O_3$. This averaging time, and domain-wide spatial measure, is also used to determine attainment with the National Ambient Air Quality Standard (NAAQS) for $O_3$. For operational forecast purposes, this means that the geographic location of peak $O_3$ predicted by the model is of less importance than the magnitude. However, operational forecasters, as shown below, can use the location of the maximum predicted $O_3$ as a way to calibrate the model for known errors.

$O_3$ forecasts are verified with observed $O_3$ from a network of monitors across the forecast area (Figure 1). During the period 2007-2009, there were 15-19 active monitors in the PHL forecast area. The $O_3$ data for 2007-2008 were obtained from the EPA Technology Transfer Network Air Quality System (AQS) archive. Because there is a time lag of months between observations and data availability at AQS, the data for 2009 was downloaded from the near real-time EPA AirNow data system (http://www.airnow.gov). Prior experience has shown that differences between the AQS and AirNow databases are slight for the measures of interest analyzed in this paper.

## 3. LIMITS OF CURRENT FORECAST METHODS

Historically, air quality forecast guidance has been provided by statistical models (Ryan et al.,

* *Corresponding author address*: William F. Ryan, Pennsylvania State University, Department of Meteorology, 401 Walker Building, University Park, PA, 16802, email: wfr1@psu.edu.

2000). A variety of statistical methods have been successfully applied to $O_3$ forecasting including, but not limited to, multiple linear regression (MLR), Classification and Regression Trees (CART) and neural networks (e.g., EPA, 2003; Nebot et al., 2008). In the PHL area, beginning in 1996, many of these methods have been tested and yield similar results. The attraction of statistical models for $O_3$ forecasting is the low cost of development; ease of updating, and, until recently, adequate and reliable performance.

Any statistical model for $O_3$ forecasting is anchored by the strong relationship between peak $O_3$ concentrations and maximum surface temperature ($T_{max}$). For the period 1993-2008, 53% of the variance in peak 8-hour $O_3$ in PHL can be explained by a linear regression model using $T_{max}$ as a predictor. Persistence, or current day $O_3$ forecasted for tomorrow, is another commonly used predictor and, given the long life time of $O_3$ in the troposphere, makes physical sense. A more accurate measure of persistence relates upwind $O_3$ concentrations, as determined by back trajectory models, to next day $O_3$ (Coburn and Hubbard, 1999). For the PHL area, the combination of $T_{max}$ and persistence explains 60% of the variance in peak 8-hour $O_3$.

Other commonly used predictors include relative humidity, as a proxy for cloud cover and precipitation, wind speed, to account for horizontal ventilation, and temperature advection or stability to account for vertical ventilation. Because $O_3$ has a distinct seasonal cycle, other measures, such as solar zenith angle, day length, or Julian date are also used. The statistical models used for comparison purposes in this study can be found at: http://www.meteo.psu.edu/~wfryan/ams2010/phl-stat1.xls. These models use MLR methods which have the advantage of providing a quantitative (ppbv) forecast in a manner that is transparent to the user.

Statistical $O_3$ models, of any type, have several important limitations. While they can account for meteorological and seasonal factors, they do not fully account for chemically forced effects. With the exception of day of week predictors, statistical models typically assume a static chemical environment. Some changes in emissions, for example, biogenic hydrocarbons, can be partly resolved by temperature predictors but others, often dependent on transport and air mass history, are unresolved. Statistical models, which require a long historical dataset for training, can also be affected by secular trends in emissions. For example, regional reductions in stationary source $NO_x$ emissions following the so-called $NO_x$ SIP Rule in 2002 have had a significant impact on $O_3$ concentrations in the eastern US (Bloomer et al., 2009). The $O_3$ monitor at Big Meadows in Shenandoah National Park, a regional scale observation site, has seen the frequency of days with $O_3$ exceeding 70 parts per billion by volume (ppbv) drop from 38 days per year for the period 1994-2002, to 9 days per year from 2003-2009. Statistical models in PHL, trained on data prior to 2003, show a sustained increase in over-prediction bias, with an attendant loss of skill, since 2003 (Figure 2).

For 2009, a new statistical model, utilizing only post-$NO_x$ SIP Rule data (2003-2007) for training, was implemented in PHL. While this model reduced forecast bias by 28%, compared to the older statistical models, and improved mean absolute error by 18%, it still exhibited an over-prediction bias (6.8 ppbv). This may reflect an ongoing decline in emissions as additional $NO_x$ controls were phased in during the training period, or unusual circumstances during 2009, but it may also be related to more systematic changes in the chemical environment that limit the skill of statistical models. As noted above, the bulk of the skill in statistical models is due to the $T_{max}$-$O_3$ relationship. Since 2002, this relationship appears to have weakened, particularly in the high temperature range (Bloomer, et al., 2009). Peak $O_3$ concentrations in PHL, binned by ranges of $T_{max}$, show a decrease on the order of 10 ppbv in warm (> 82°F) cases. In addition, the variance in $O_3$ explained by a simple one-predictor ($T_{max}$) regression model decreased from 58%, for 1993-2002, to 44% for 2003-2009. This suggests that statistical model skill is unlikely to increase significantly in the coming years and forecasters must increasingly rely on numerical forecast guidance.

## 4. NUMERICAL FORECAST MODEL SKILL

As noted above, the forecast verification metric of interest to operational forecasters is domain-wide peak 8-hour average $O_3$. There are a number of methods to extract peak domain $O_3$ information from the NAQFC but, for this study, peak concentrations were determined from model forecasts at the grid cell closest to the location of existing regulatory $O_3$ monitors. This approach was selected for several reasons: First, the PHL forecast area is characterized by a complex sea-land boundary that includes several large bays. The NAQFC model has a tendency to simulate very high $O_3$ concentrations along land-sea boundaries – especially in prevailing westerly flow conditions. An example of this effect is shown in Figure 3. Monitors in PHL are well away from the immediate coast and forecasts for these locations are less influenced by this effect (see, Figure 1). Second, because the forecast deadline is soon after the availability of 1200 UTC model guidance, point data

is the quickest method to extract quantitative peak $O_3$ forecast information from the NAQFC.

Two immediate issues are raised for forecast verification and post-processing during the period 2007-2009. First, the NAQFC, and its sub-models, underwent continual revisions, including yearly emissions updates, during this period (see, http://www.emc.ncep.noaa.gov/mmb/aq/AQChangelog.html). Several of the post-processing methods to be discussed below utilize 2007-2008 results to inform forecasts issued in 2009. Because the model is not "frozen", results from 2007-2008 may not be stable enough to support their use as predictors. However, in the PHL area, overall NAQFS performance was similar for both 2007 and 2008 (Figure 4). Categorical forecasts of high $O_3$ (Code Orange or $[O_3] \geq 76$ ppbv) were similar, although slightly better in 2007, with a "hit" rate of 75%, compared to 69% in 2008, and a "false alarm" rate of 25% compared to 46% in 2008. Second, 2009 was a uniquely low $O_3$ season. In PHL, for the first time in the modern monitoring network era, there were no instances of Code Red (≥ 96 ppbv) $O_3$. While the 2009 summer season (June-August) was cool and wet both in PHL and across the region, it was similar to the cool and wet summers of 2000 and 2004. However, the frequency of Code Orange $O_3$ cases was 78% lower in 2009 (7 cases in 2009 compared to 35 in 2000 and 30 in 2004). At the regional scale, the monitor at Big Meadows in Shenandoah National Park did not exceed 70 ppbv (8-hour average) during the summer compared to an average of 10 cases per year for 2003-2008. 2009 was also a remarkably low summer for $PM_{2.5}$ concentrations. $PM_{2.5}$ is generally less sensitive to temperature than $O_3$. In 2009, $PM_{2.5}$ concentrations in PHL did not exceed 30 $\mu g/m^3$ (24-hour average), compared to 16.4 such days for 2004-2008. In addition, the frequency of days at SNP exceeding 20 $\mu g/m^3$ decreased from 23% (2004-2008) to 2% in 2009. The low frequency of enhanced $PM_{2.5}$ concentrations cases suggests that factors in addition to weather, perhaps reduced industrial activity due to the economic recession, also played a role.

Overall results for the NAQFC model in 2009, compared to persistence and the post-processing methods to be described below, are given in Table 1. Because forecasts are used to initiate Air Quality Action Day alerts, skill in the upper end of the $O_3$ distribution (Code Orange or higher) are of most importance. Table 1 provides skill score measures based on a 2 x 2 contingency table using Code Orange as the threshold (Stephenson, 2000, Wilks, 1995). The overall error statistics for 2009 are similar to 2007-2008 but with a slightly higher over-prediction bias. The critical problem for the NAQFC model in 2009 is the high frequency of "false alarms" – Code Orange forecast but not observed. The false alarm rate for the NAQFC in 2009 was 74% compared to an average of 34% in 2007-2008. This may be due to the unusual weather patterns during 2009 and may also reflect changes in the emissions base. In any case, the over prediction bias and false alarm rate highlights the need for a robust post-processing technique.

## 5. POST-PROCESSING OF NAQFC FORECAST GUIDANCE

A number of post-processing methods were tested using the 2007-2009 forecast dataset. The methods were selected, in part, based on ease of application. Operational air quality forecasting in the US is carried out almost exclusively at the state and local government level where time and resources are limited. As a result, adoption of numerical model forecast guidance as part of the usual forecast "rote" requires post-processing methods that are inexpensive to install and operate, locally focused and, because forecast deadlines follow immediately after the NAQFC 1200 UTC run completes its cycle, timely.

Four methods were tested initially: (1) binned bias correction; (2) running bias correction; (3) "trend" correction; and, (4) a simple ensemble. The methods are described in more detail below with results from the 2009 summer season provided in Table 1. The *binned bias correction* method used forecast guidance from 2007-2008 placed in bins of 10 ppbv size with the forecast bias computed for each bin and used to correct the operational model forecast. In general, the NAQFC model tended to under-predict in low $O_3$ forecasts and over-predict in high $O_3$ forecasts. The length of the *running bias correction*, based on results from earlier tests, was set at two days. That is, mean error over the current and previous day was added (or subtracted) to the next day model forecast. The *trend correction* method compares the current day NAQFC forecast, determined from the previous day's 1200 UTC model run, to the next day forecast and adjusts the current day *observed* $O_3$ concentrations to account for the forecast trend. For example, if 60 ppbv was forecast for the current day and 70 ppbv for the next day, then 10 ppbv is added to today's observed concentrations. In operational use, as opposed to this test, the last two methods are limited by incomplete knowledge of current day peak 8-hour $O_3$. That is, the forecast is issued during the early afternoon hours before peak 8-hour average concentrations are typically reached. The *simple ensemble* method is a so-called "poor man's ensemble" that combines the NAQFC forecast with output from statistical forecast models. In this case, the weighting is 50% NAQFC model, 30% new statistical model (trained on post-$NO_x$ SIP Rule data) and 20% old statistical model (trained on pre-$NO_x$ SIP Rule data).

## 6. RESULTS AND DISCUSSION

The bias correction methods (binned and two-day running bias) provided the best results overall. As expected, they reduced forecast bias but also reduced mean and median absolute error in the range of 4-9%. The binned bias method also reduced rms error by 15% suggesting that it has fewer large error cases. The large error cases typically involve abrupt air mass changes and/or precipitation. None of the methods solved the problem of false alarms although all improved on the Threat Score of the NAQFC. Although a variety of skill measures were calculated from the contingency table, only the Threat Score is given in Table 1 as it is the most meaningful measure in situations where the threshold is only rarely exceeded – in this case less than 10% of the time (Stephenson, 2000). The other measures corroborated the results given by the Threat Score. The 2-day running bias correction provided the most improvement at the Code Orange threshold, reducing the false alarm rate at only a slight cost to the hit rate.

Further analysis of the false alarm cases in 2009 showed that the NAQFC tended to over-predict in warm temperature cases and in cases where previous day upwind convection affected regional scale $O_3$ concentrations. The temperature affect is shown in Figure 5. The upwind effects are not easily accounted for but can be approached using persistence $O_3$ as a predictor. With this in mind, an additional post-processing method was tested with a simple multiple linear regression model (MLR) using NAQFC forecast $O_3$, $T_{max}$ and persistence $O_3$ as predictors. The details of the model are provided in Table 2 and the results in Table 3. Overall, the MLR model provided the best forecast, rivaling the public (expert analysis) forecast. More importantly, the MLR model was able to reduce the false alarm rate without adversely affecting the hit rate – a rare occurrence. In absolute terms, the number of false alarms was reduced from 14 (NAQFC model) to 5 by the MLR model and, of these five, three were cases of "near misses" - observed peak $O_3 \geq 71$ ppbv. Other skill measures corroborate the results shown in Table 3. For example, Brier scores, a measure of both forecast reliability and resolution, show an increase in skill for the MLR model (Table 4). Table 4 also shows the effect of changes in emissions due to the $NO_x$-SIP Rule. The "old" statistical model, trained on data prior to the $NO_x$ SIP Rule and which provided adequate forecasts prior to 2003, shows no skill as applied in 2009. The "new" statistical model shows some skill but less than the NAQFC and considerably less than the post-processed NAQFC.

## 7. CONCLUSIONS

Numerical air quality forecast models now routinely provide forecast guidance to operational air quality forecasters. Numerical models for $O_3$ are more advanced, in terms of skill, than $PM_{2.5}$ models but their adoption as a key tool in operational forecast preparation has been slow. Adoption of model guidance as part of the routine forecast "rote" will be accelerated if skill overall and in the high end of the $O_3$ distribution can be shown ($\geq$ 76 ppbv). The standard metric for forecast verification, local peak 8-hour $O_3$ concentrations, is a difficult measure for numerical forecast models to resolve. Model performance may be improved by the use of post-processing methods. Because air quality forecasting is carried out at the local air quality agency level, often under significant time and resource constraints, post-processing methods must be inexpensive to design and operate, locally focused and timely.

A number of post-processing methods, adaptable to local conditions, were tested in the PHL area using data from the summer of 2009. Of the various methods tested, the most effective were bias correction, with the two-day running bias correction showing the largest improvement in skill. A recurring problem in 2009 was the frequency of "false alarm" forecasts of Code Orange conditions. This was likely due both to weather effects (2009 was a cool and wet summer) and to changes in emissions (e.g., impacts of the severe economic recession). A simple MLR model was able to reduce the rate of false alarms significantly, with no loss of detection skill. The unusual circumstances in 2009 suggest that further research along these lines will be necessary.

## ACKNOWLEDGEMENTS

The authors acknowledge the support of the Delaware Valley Regional Planning Commission (Sean Greene) and the States of Delaware and New Jersey, as well as the Commonwealth of Pennsylvania, for this research as well as operational air quality forecasting in Philadelphia. The support of the Air and Radiation Management Administration of the Department of the Environment of the State of Maryland (David Krask and Michael Woodman) is also gratefully acknowledged.

**REFERENCES**


Bloomer, B. J., et al., 2009, Observed relationships of ozone air pollution with temperature and emissions, *Geophys. Res. Letters*, **36**, L09893.

Cobourn W. G., and M. C. Hubbard, 1999: An enhanced ozone forecasting model using air mass trajectory analysis, *Atmos. Environ*., **33**, 4663-4674.

Eder, B. et al., 2009, A performance evaluation of the National Air Quality Forecast Capability for the summer of 2007, *Atmos. Environ*., **43**, 2312-2320.

EPA, 2003, *Guidelines for Developing an Air Quality (Ozone and $PM_{2.5}$) Forecasting Program*, U.S. Environment Protection Agency, Office of Air Quality Planning and Standards, EPA-456/R-03-002.

McHenry, J. N., et al., 2004, A real-time Eulerian photochemical model forecast system: Overview and initial ozone forecast performance in the Northeast U.S. Corridor, *Bull. Amer. Meteor. Soc*., **85**, 525-548.

Nebot, A. et al., 2008, Ozone prediction based on meteorological variables: A fuzzy inductive reasoning approach, *Atmos. Chem. Phys. Discuss*., **8**, 12343-12370.

Otte, T. L., et al., 2005, Linking the Eta model with the Community Multiscale Air Quality (CMAQ) modeling system to build a national air quality forecasting system, *Wea. Forecasting*, **20**, 367-384.

Ryan, W. F., C. A. Piety and E. D. Luebehusen, 2000, Air quality forecasting in the mid-Atlantic region: Current practice and benchmark skill, *Wea. Forecasting*, **15**, 46-60.

Stephenson, D. B., 2000: Use of the “odds ratio” for diagnosing forecast skill, *Wea. Forecasting*, **15**, 221-232.

Wilks, D. S., *Statistical Methods in the Atmospheric Sciences*, Academic Press, 467pp., 1995.

Table 1. Skill measures for PHL $O_3$ forecast methods in 2009. All measures are given in ppbv for peak domain-wide 8-hour average $O_3$ concentrations. The skill score measures are based on a 2 x 2 contingency table using 76 ppbv (Code Orange) as the threshold (Stephenson, 2000). "AE" refers to absolute, or unsigned, error. The "Threat" score is also known as the Critical Success Index (CSI). A summary of skill score measures can be found at: http://www.meteo.psu.edu/~wfryan/ams2010/skill-contingency-2009.docx

| **Philadelphia Ozone Forecasts (2009)<br>Error and Skill Measures** | | | | | | |
|---|---|---|---|---|---|---|
| | **NAQFC** | **Binned Bias** | **Running Bias** | **Trend** | **Ensemble** | **Persistence** |
| **Bias** | 4.3 | 2.8 | -0.2 | -0.3 | 6.0 | 0.0 |
| **Mean AE** | 7.5 | 6.8 | 7.2 | 7.9 | 7.6 | 10.6 |
| **Median AE** | 7.0 | 6.7 | 6.5 | 7.0 | 6.4 | 10.0 |
| **rms** | 8.9 | 7.6 | 8.8 | 9.5 | 9.2 | 12.8 |
| | **Skill Score Measures** | | | | | |
| **Hit** | 0.71 | 0.57 | 0.67 | 0.67 | 0.71 | 0.14 |
| **False Alarm** | 0.74 | 0.67 | 0.56 | 0.64 | 0.71 | 0.86 |
| **Threat** | 0.24 | 0.27 | 0.36 | 0.31 | 0.26 | 0.08 |

Table 2. A summary of information describing the multiple linear regression model (MLR) used to post-process NAQFC forecasts for Philadelphia in 2009. The model was trained using data from 2007-2008 (N = 269).

| **Multiple Linear Regression Model for<br>Post-Processing NAQFC Forecasts in Philadelphia** | | | |
|---|---|---|---|
| $\mathbf{[O_3]_{obs} = 0.58*[\ O_3]_{NAQFC} + 0.44*T_{max} + 0.17*[O_3]_{lag} - 21.3}$<br>($r = 0.79$; $r^2 = 0.65$) | | | |
| | **Standard Coefficient ("Beta" Weight)** | **Tolerance** | **t** |
| NAQFC | 0.512 | 0.397 | 8.58 |
| $T_{max}$ | 0.209 | 0.411 | 3.56 |
| Persistence $O_3$ | 0.168 | 0.683 | 3.69 |

Table 3. Forecast results and skill scores, as in Table 1, for the NAQFC model, two post-processing methods described in the text (MLR Model and 2-Day Running Bias Correction), the forecast issued to the public after expert analysis (Public Forecast), and the reference forecast (Persistence).

| Philadelphia Ozone Forecasts (2009)<br>Error and Skill Measures | | | | | |
|---|---|---|---|---|---|
| | **NAQFC** | **MLR Model** | **Public Forecast** | **Running Bias** | **Persistence** |
| **Bias** | 4.3 | 3.0 | 3.4 | -0.2 | 0.0 |
| **Mean AE** | 7.5 | 6.4 | 6.4 | 7.2 | 10.6 |
| **Median AE** | 7.0 | 5.9 | 5.0 | 6.5 | 10.0 |
| **rms** | 8.9 | 8.2 | 8.0 | 8.8 | 12.8 |
| | **Skill Scores** | | | | |
| **Hit** | 0.71 | 0.71 | 0.57 | 0.67 | 0.14 |
| **False Alarm** | 0.74 | 0.50 | 0.56 | 0.56 | 0.86 |
| **Threat** | 0.24 | 0.42 | 0.33 | 0.36 | 0.08 |

Table 4. Brier Score and Brier Skill Score for PHL $O_3$ forecasts in 2009. Persistence is used as the reference forecast for the calculation of the Brier Skill Score. Three categories, based on the lowest 25$^{th}$ percentile of observed $O_3$, the inter-quartile range and the highest 25$^{th}$ percentile, were used to calculate the Brier scores. The Old Statistical Model is trained on pre-$NO_x$ SIP Rule data and the New Statistical Model is trained on post-$NO_x$ SIP Rule data.

| Brier Score Measures for 2009<br>Philadelphia Forecast Area | | |
|---|---|---|
| | **Brier Score** | **Brier Skill Score** |
| **NAQFC** | 0.30 | 0.23 |
| **MLR Model** | 0.20 | 0.50 |
| **Running Bias** | 0.26 | 0.34 |
| **Old Statistical Model** | 0.43 | -0.09 |
| **New Statistical Model** | 0.34 | 0.14 |
| **Persistence** | 0.39 | - |

Figure 1. The Philadelphia metropolitan air quality forecast area (blue box) with location of surface $O_3$ measurement monitors given by black triangles. Figure courtesy of USEPA AirNow Tech website (http://www.airnowtech.org/).

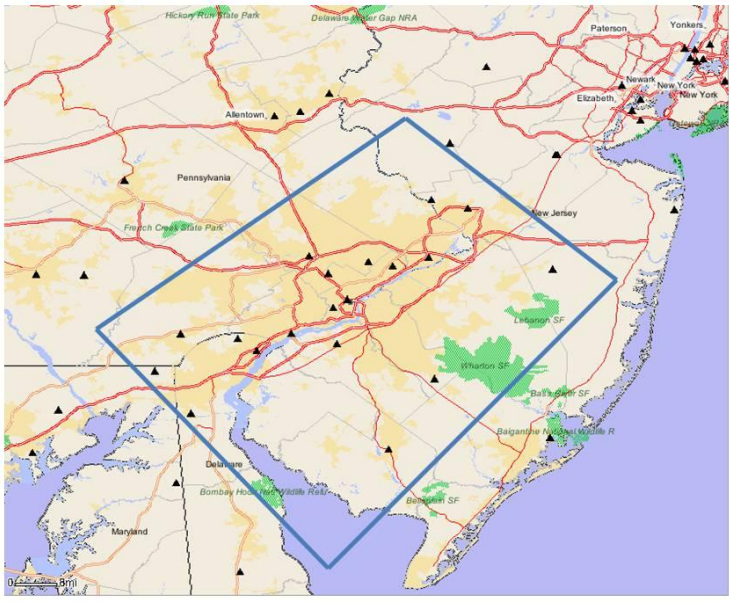



Figure 2. Forecast bias for peak domain-wide 8-hour $O_3$ for two statistical models used in the Philadelphia metropolitan area for the summer (early May-early September) seasons of 2003-2009. Both models used MLR regression techniques and were trained on data preceding the regional $NO_x$ emission controls in 2002.

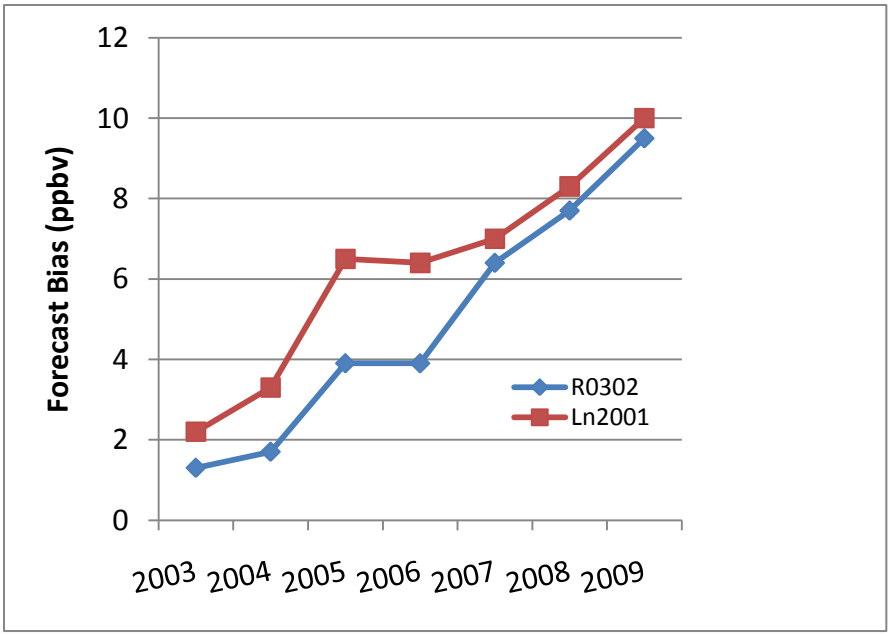


Figure 3. Peak 8-hour average $O_3$ forecast (in ppbv) for August 5, 2009 from the NAQFC model initialized at 1200 UTC on August 4. Figure courtesy of NOAA EMC Mesoscale Modeling Branch (http://www.emc.ncep.noaa.gov/mmb/aq/).

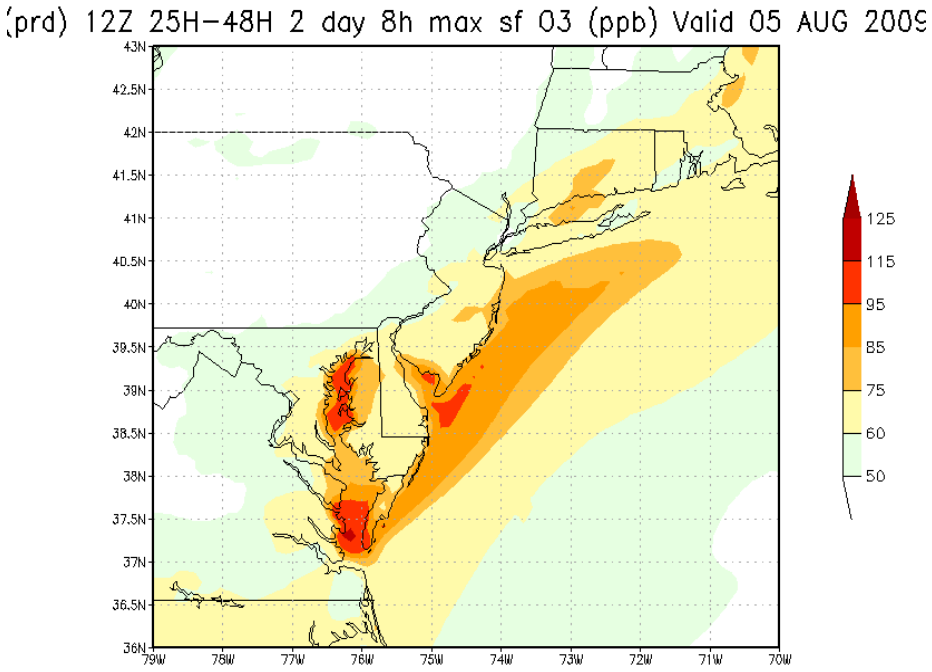



Figure 4. Selected skill measures for peak 8-hour average $O_3$ for the NAQFC model in the PHL area for 2007-2008. Bias, mean and median absolute error (in ppbv) are at the top left. Pearson's correlation coeffient (r) and explained variance ($r^2$) for a simple linear regression model, using NAQFC forecast $O_3$ as a predictor, are given in the bottom left. A histogram of signed forecast error is given on the right.

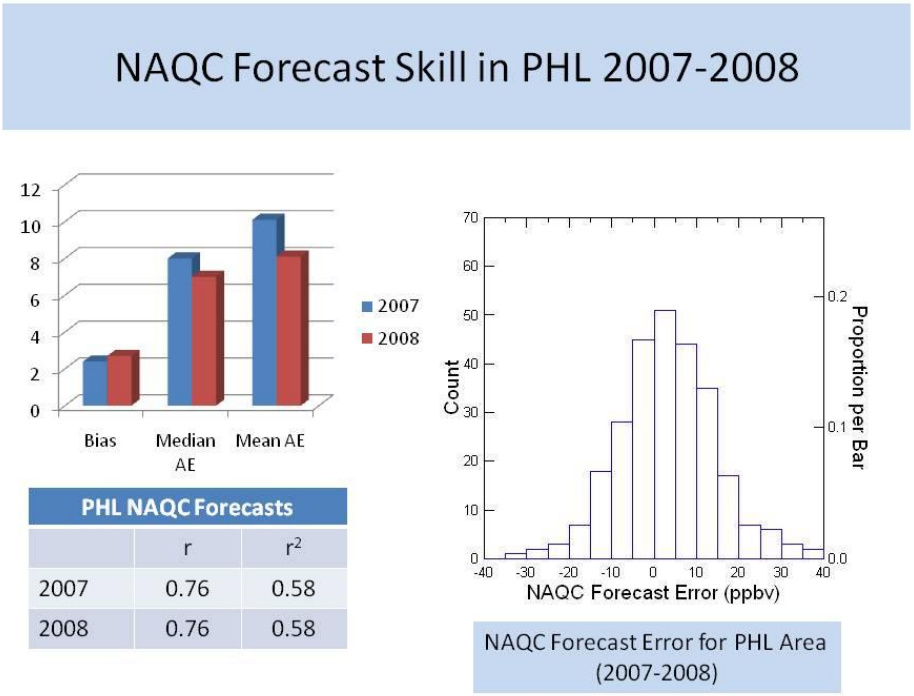
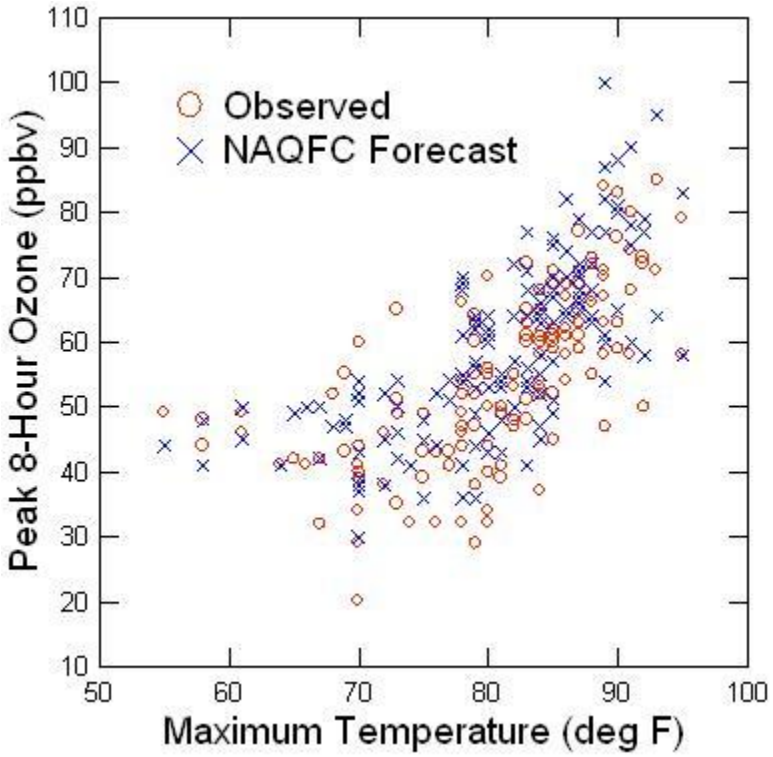


Figure 5. Maximum temperature at PHL and peak 8-hour $O_3$ concentrations observed and forecast for the summer season of 2009.




**Appendix A.** A Partial List of Operational Numerical $O_3$ Forecast Models

National Air Quality Forecast Capability Model (NOAA/EPA):
http://www.weather.gov/aq/

Environment Canada:
http://www.weatheroffice.gc.ca/chronos/index_e.html#o3_10

Barons Advanced Meteorological Systems:
http://www.baronams.com/

University of Houston, Institute for Multi-Dimensional Air Quality Studies:
http://www.imaqs.uh.edu/ozone_forecast.htm

North Carolina Department of Environment and Natural Resources, Division of Air Quality:
http://daq.state.nc.us/airaware/forecast/model/

SUNY-Albany, Atmospheric Sciences Research Center:
http://asrc.albany.edu/research/aqf/aqvis/tomorrowforecast_maps.htm

Washington State University, Air-Quality Forecasting for the Pacific Northwest (AIRPACT):
http://lar.wsu.edu/airpact-3/