State of Delaware
Final Report: Ozone Observations and Forecasts in 2016

A Report Prepared for the Delaware Department of Natural Resources and Environmental Control

Dr. Amy K. Huff
William F. Ryan
Department of Meteorology and Atmospheric Science
The Pennsylvania State University

March 31, 2017

Table of Contents

List of Tables

List of Figures

Executive Summary

- A new, lower 8-hour average ozone (O_3) exceedance threshold of 71 ppbv was in effect for 2016.
- New O_3 Air Quality Index (AQI) breakpoints were also in effect for 2016, with lower thresholds for the key Code Orange (71-85 ppbv) and Code Red (86-105 ppbv) categories.
- In 2016, there were 11 days with observed 8-hour average $O_3 \geq 71$ ppbv, 1 day with observed 8-hour average $O_3 \geq 86$ ppbv, and no days with observed 8-hour average $O_3 \geq 106$ ppbv. This translated into 10 Code Orange days and 1 Code Red day in Delaware on the updated AQI scale.
- The recent substantial downward trend in observed O_3 in Delaware, which began in 2013, continued in 2016. As a result, the 2016 O_3 season was the fourth consecutive historically low O_3 season for the State of Delaware in terms of O_3 exceedances and seasonal mean/median O_3.
- The "step-down" in observed O_3 during 2013-2016 is not unique to Delaware; the same trend is evident across the entire Mid-Atlantic region.
- There is mounting evidence that continued reductions in regional emissions of nitrogen oxides (NO_x), particularly from the upwind Ohio River Valley source region, are the primary cause of the "step-down" in observed O_3 during 2013-2016.
- The argument that changes in weather conditions, related to the cooler and wetter summers of 2013-2014, were primarily responsible for the lower observed O_3 levels since 2013 has been refuted by the warm summers of 2015 and especially 2016.
- The historical strong relationships between O_3 exceedance days and 1) hot weather, 2) persistence O_3, and 3) westerly transport aloft from the Ohio River Valley have all broken down in the 2013-2016 time period.
- Likewise, the traditional conceptual model for high O_3 events in the Mid-Atlantic region, which relies on climatology, synoptic scale transport patterns, temperature, and persistence O_3, has broken down in the 2013-2016 period. As a result, identifying O_3 exceedances in Delaware has become more challenging, especially single-day, isolated events.
- Forecasters are adapting to the "New Normal" O_3 environment, however; forecast skill of O_3 exceedance days improved in 2016. The hit rate was 0.55 and the false alarm rate was 0.33. The false alarm rate was a substantial improvement from 2015 (0.71).
- Overall forecast skill (all days) for 2016 was consistent with 2013-2015, with a median absolute error of 6.0 ppbv and bias of 2.1 ppbv.
- Reflecting the changing air quality conditions, numerical air quality models were remarkably unskillful at predicting the O_3 exceedance days in 2016. All of the models had false alarm rates > 0.50 and hit rates < 0.50.
- Smoke transported from wildfires is becoming an important contributor to local O_3 production in Delaware, with 3 of 11 O_3 exceedances days in 2016 attributable to smoke.

1. Ozone Observations in 2016

For summer 2016, a new 8-hour average ozone (O_3) exceedance threshold of **71 ppbv** was in effect, as a consequence of the new, lower primary O_3 National Ambient Air Quality Standard (NAAQS) of 70 ppbv issued by U.S. EPA on October 1, 2015. New O_3 Air Quality Index (AQI) breakpoints were also in effect for 2016, with lower thresholds for the key Code Orange and Code Red categories (Table 1). In 2016, there were 11 days with 8-hour average $O_3 \geq 71$ ppbv, 1 day with 8-hour average $O_3 \geq 86$ ppbv, and no days with 8-hour average $O_3 \geq 106$ ppbv in the State of Delaware forecast area, which includes Kent, New Castle, and Sussex Counties. This translated into 10 Code Orange days and 1 Code Red day on the updated AQI scale.

Details about the magnitude of the 11 O_3 exceedance days in Delaware during 2016 are listed in Table 2. Of note are two days, July 8 and July 27, when an O_3 exceedance was recorded at only one monitor using a 6-hour running average instead of the standard 8-hour running average. Although these days technically count as exceedances for regulatory purposes, it is possible that had a full 8 hours of observations been available, the monitors would not have recorded an exceedance. This is particularly true for the July 27 case, when the Bellefonte monitor exceeded by only 1 ppbv. As a result, the total of 11 O_3 exceedances recorded for 2016 should be interpreted as more likely having been only 10 or possibly 9 exceedances, based on the actual physical state of the atmosphere.

Figure 1 shows how the 2016 O_3 season (May 1 to September 30) compares to previous years. The recent substantial downward trend in the number of O_3 exceedance days in Delaware (orange bars in Figure 1), which began in 2013, continued in 2016. The average number of O_3 exceedance days was only 8.3 in 2013-2016, compared to 29.3 in 2003-2012 and 49.8 in 1997-2002. As a result, the 2016 O_3 season was the fourth consecutive historically low O_3 season for the State of Delaware. The only exception is the Great Recession year of 2009, which had the lowest number of O_3 exceedances (5) in Delaware since 1997. Note that using an O_3 exceedance threshold of 76 ppbv instead of 71 ppbv changes the magnitude of the number of exceedance days per year, but it does not alter the sharp downward trend in 2013-2016 compared to previous years.

The seasonal (May 1 to September 30) mean and median of the maximum 8-hour observed O_3 have the same downward trend as O_3 exceedance days in Delaware (Figure 2). The 2016 seasonal mean and median were the fourth consecutive historically low values since 1997. For the median, only 2013-2014 had lower values than 2016. Thus, not only have the highest O_3 values dropped abruptly in the period 2013-2016, but the average O_3 values as well.

2. The Impact of Continued Regional Reductions in NO_x Emissions in 2016

The recent decrease in observed O_3 in 2013-2016 is not unique to Delaware; the same trend is evident across the entire Mid-Atlantic region. Figure 3 shows the total number of O_3 monitors in the Mid-Atlantic region with maximum 8-hour observed O_3 at or above thresholds of 76 ppbv (orange bars), 96 ppbv (red bars), and 116 ppbv (purple bars) per year for 1994-2016. Three periods are evident from the data in Figure 3: the Pre-NO_x SIP period (prior to 2003; A in Figure 3), which was marked by historically high O_3; the Post-NO_x SIP period (2003-2012; B in Figure

3), when an initial "step-down" in observed O_3 occurred; and the New Normal period (2013-2016; C in Figure 3), characterized by a second "step-down" in observed O_3. A temporary instance of a "step-down" in observed O_3 occurred during the Great Recession year of 2009 (D in Figure 3).

The initial "step-down" in observed O_3 during the Post-NO_x SIP period has been attributed to reductions in O_3 precursors, primarily emissions of nitrogen oxides (NO_x) from energy generating units (EGUs) in the eastern U.S, associated with the so-called "NO_x SIP Rule." There is strong evidence that continued reductions in regional NO_x emissions, particularly from the upwind Ohio River Valley (ORV) source region, are primarily responsible for the second "step-down" in observed O_3 during the New Normal period. Figure 4 shows the annual percent reduction in NO_x emissions from EGUs in local and upwind states for 2006-2016 relative to the base year of 2005 using U.S. EPA's Air Markets Program Data. The local and upwind states include DC, DE, IL, IN, KY, MD, MI, NC, NJ, NY, OH, PA, TN, VA, and WV. The annual percent reductions in NO_x emissions have been on a downward trajectory since 2008, with the largest decreases in 2009 and 2012-2016. The years with > 30% reductions in NO_x emissions, 2009 and 2013-2016, correspond to the historically low observed O_3 years in Delaware and the Mid-Atlantic. This result suggests that 30% reduction in NO_x emissions, relative to 2005 emissions, is the threshold for the sharply lower O_3 levels observed during the Great Recession year and the New Normal period.

Additional evidence that decreases in regional NO_x emissions during the New Normal period are primarily responsible for the recent "step-down" in observed O_3 is provided by the meteorology of summer 2016. The summers of 2009 and 2013-2014 were cooler and wetter than average, and, considered in isolation, the historically low observed O_3 conditions during these years may be attributable to weather conditions. The summer of 2015, however, was near average in terms of temperature, which should have translated into an average O_3 season (relative to the 2003-2012 period). Instead, 2015 was another historically low O_3 season. The summer of 2016 was around normal in terms of precipitation (Figure 5) but much warmer than average (Figure 6). There were 34 days with maximum temperature (T_{max}) ≥ 90 °F recorded at Wilmington, New Castle County Airport (KILG) in June-August 2016 compared to an average of 20. August was especially hot, with 15 days ≥ 90 °F at KILG compared to a mean of 5. The above average warmth in 2016 did not translate into a relatively large number of O_3 exceedance days, as would be expected if meteorology and not NO_x emissions were the primary driver of recent O_3 conditions. Figure 7 shows how the annual number of hot days in Delaware, defined as maximum temperature (T_{max}) ≥ 90 °F at Dover Air Force Base (DOV), during the O_3 season for 1997-2016. The summer of 2016 (36 hot days) was around the same as 2012 (32 hot days) in terms of temperature, but 2012 had 31 O_3 exceedance days compared to only 11 in 2016. O_3 formation during the warm summers of 2015 and especially 2016 strongly suggests that variations in weather are not primarily responsible for the recent "step-down" in observed O_3 during 2013-2016.

Furthermore, results from Figure 8 show that while the number of hot and wet days remained relatively constant in Delaware across the Pre-NO_x SIP, Post-NO_x SIP, and New Normal periods, the number of O_3 exceedance days sharply declined. Figure 8 illustrates that average number of days with T_{max} ≥ 90 °F (21-27 days) and the average number of days with measureable precipitation (41-47 days) at DOV were essentially constant for the three periods. In contrast, the number of O_3 exceedance days dropped from 50 in 1997-2002 to 29 in 2003-2012 to only 8 in 2013-2016. Thus, there were a comparable number of hot and rain-free days in Delaware in the

New Normal period compared to previous years. Based on these weather conditions alone, observed O_3 during 2013-2016 should have been similar to previous years. Since it was not, and regional NO_x emissions have decreased substantially during the same period, it is logical to interpret the recent decreases in observed O_3 as primarily attributable to reductions in regional NO_x emissions.

3. Implications for O_3 Forecasting

The decline in observed O_3 during the New Normal period has implications for air quality forecasting. Historically, most O_3 exceedance days in the Mid-Atlantic region were characterized by a "heat wave" synoptic weather pattern, featuring a "ridge" of high pressure aloft, slowly eastward migrating surface high pressure, westerly transport aloft from the ORV, sunny skies, and hot weather. These synoptic meteorological conditions favored regional, multi-day O_3 exceedance events. As a result, forecasts of T_{max} and persistence O_3 were strong forecast tools that could dependably predict O_3 exceedance days. Forecasters could also rely on O_3 climatology as a foundation for the daily O_3 forecast. The traditional conceptual model for O_3-conducive weather patterns is breaking down in the New Normal period, however, rendering the historical forecast tools unreliable.

3.1 Hot Weather

Historically, there was a strong relationship between high O_3 and hot weather. Figure 9 shows the percent of hot days in Delaware with maximum 8-hour observed $O_3 \geq 71$ ppbv (orange dots/line) and ≥ 86 ppbv (red dots/line) for 1997-2016. During most years of the Pre-NO_x SIP period, almost all hot days (~80-90%) were O_3 exceedance days. That percentage dropped in the Post-NO_x SIP period, but for the most part, a majority of hot days were still O_3 exceedance days. This historically strong relationship between hot weather and O_3 exceedance days was a key forecast tool: a forecast of $T_{max} \geq 90$ °F implied a likely O_3 exceedance day, assuming no optically thick clouds, deep vertical mixing, or strong winds were expected. In the New Normal period, however, only about 10-20% of hot days are O_3 exceedance days. This means that hot weather is still necessary, but no longer sufficient for high O_3. Now, most hot days are *not* O_3 exceedance days in Delaware or the Mid-Atlantic region, and forecasters can no longer rely on forecasts of $T_{max} \geq 90$ °F as a strong O_3 predictor. This trend was clearly illustrated during August 2016, which had only 1 O_3 exceedance day in Delaware (August 27) despite 13 days with $T_{max} \geq 90$ °F at DOV and 15 days with $T_{max} \geq 90$ °F at KILG.

3.2 Persistence O_3

Historically, persistence O_3 was a useful forecast predictor because O_3 has a long enough lifetime in the atmosphere to allow for transport from upwind locations. During a regional O_3 event, synoptic weather conditions would allow for build-up of O_3, making persistence O_3 an informative indicator of "tomorrow's" O_3 levels. In the New Normal period, persistence O_3 has also become an unreliable forecast variable, however. Figures 10-12 illustrate the linear relationship between maximum 8-hour observed O_3 and local 1-day lag (persistence) maximum 8-hour observed O_3 for Delaware during the Pre-NO_x SIP (Figure 10), Post-NO_x SIP (Figure 11), and New Normal (Figure

12) periods. Figure 13 summarizes the downward trend in the coefficient of determination (R^2 value) of the persistence regression equations for the three time periods. During the Pre-NO_x SIP and Post-NO_x SIP periods, persistence O_3 was a useful forecast variable, explaining 39% (Pre-NO_x SIP) and 33% (Post-NO_x SIP) of the total variation in observed O_3. But during the New Normal period, persistence O_3 only explains 25% of the total variation in observed O_3, which limits the effectiveness of persistence as a forecast tool.

Also evident in Figures 10-12 is the shrinking spread in the range of maximum 8-hour observed O_3 on the high end of the distribution. Figure 14 illustrates the 10^{th}, 25^{th}, 75^{th}, and 90^{th} percentile values of maximum 8-hour observed O_3 for the Pre-NO_x SIP, Post-NO_x SIP, and New Normal Periods. The O_3 values at each percentile decreased from the Pre-NO_x SIP period to the Post-NO_x SIP period to the New Normal period, but the decreases are largest at the high end of the distribution. For example, the O_3 values at the 10^{th} percentile dropped by only 3 ppbv from the Pre-NO_x SIP to the New Normal period, but for the 90^{th} percentile, they dropped from 91 ppbv to 66 ppbv. These results show that while the lowest observed O_3 values have remained approximately the same, the highest observed O_3 values have dropped substantially in the New Normal period. As a result, Delaware no longer observes O_3 at the very high end of the distribution, relative to previous years.

The weakening correlation of persistence O_3 and the decrease in the range of observed O_3 on the high end of the distribution means that O_3 exceedance days are occurring less often in multi-day events and more often as single day “spikes.” As shown in Table 2, more than half (7) of the 2016 O_3 exceedance days were single day “spikes.” Single day O_3 exceedances are much more difficult to forecast than regional multi-day O_3 events, as they are often isolated geographically and can depend on mesoscale weather features, which are less accurately predicted by numerical weather models compared to synoptic conditions. In addition, 1 of the 2016 exceedance days (June 11) immediately followed a Code Green (Good) observed O_3 day, signifying a very localized, rapid increase in O_3, which was rare prior to 2013.

3.3 Westerly Transport Aloft

Historically, O_3 exceedance days in Delaware and the Mid-Atlantic region were characterized by a westerly transport pattern, in which air aloft at approximately 500-1500 m above ground level (AGL) flowed from the ORV. This pattern was conducive for high O_3 because the ORV has traditionally been a source region for O_3 and emissions of NO_x due to a large concentration of coal-burning EGUs. With the advent of NO_x emission control technologies, closure of inefficient power plants, and fuel transition from coal to natural gas, NO_x emissions from upwind locations in the ORV have decreased substantially in recent years (Figure 4). As a result, westerly transport aloft is no longer strongly associated with O_3 exceedance days in Delaware.

During the New Normal period, local scale circulation is becoming a more important predictor of high O_3. Figure 15 shows the mean (blue bars) and median (orange bars) lengths of 24-hour backward air mass trajectories ending at Philadelphia International Airport (KPHL) for the 96^{th} percentile maximum 8-hour observed O_3 cases (highest observed O_3 days) during 1999-2001, 2005-2007, 2010-2012, and 2013-2016. Backward air mass trajectories indicate the upwind transport pattern. The years 1999-2001 are representative of the Pre-NO_x SIP period, and 2005-

2007 and 2010-2012 are representative of the Post-NO_x SIP period, avoiding any interference from the Great Recession year of 2009. KPHL is used as a representative location along the urbanized Interstate-95 Corridor in the Mid-Atlantic region. Results from Figure 15 demonstrate that the average length of back trajectories did not vary much in the Pre-NO_x SIP or the Post-NO_x SIP periods (~450-500 km). These trajectories were also predominately westerly from the ORV (not shown). But in the New Normal period, back trajectories shortened to around 250-300 km. These shorter back trajectories correspond to an increase in the number of high O_3 days when local circulation dominated. In contrast, there were no high O_3 days during 2005-2007 with an upwind source region within 250 km of KPHL. Days when winds stagnate and local emissions dominate are now more likely to be high O_3 days in the Mid-Atlantic compared to the period prior to 2013, when westerly transport was dominant.

In 2016, only 1 O_3 exceedance day (June 20) in Delaware was associated with westerly transport aloft. In contrast, a subset of 3 of the single exceedance day "spikes" were isolated exceedances (July 8 and 27, August 27), meaning that only 1-2 monitors in Delaware exceeded (with no exceedances in adjacent forecast areas); these days were influenced by stagnating and recirculating winds.

3.4 Climatology

Historically, maximum 8-hour observed O_3 followed a seasonal cycle in Delaware and the Mid-Atlantic region, with a peak from late June to late July. This seasonal climatology provided a foundation that forecasters could use to anchor the daily O_3 forecast, using the 25th percentile, median, 75th percentile, and 95th percentile values as benchmarks. For example, on an expected high O_3 day, the forecaster could use the 95th percentile O_3 value as an upper limit for the forecast.

In the New Normal period, however, the historical seasonal O_3 cycle has virtually disappeared. Figure 16 shows the daily time series of maximum 8-hour observed O_3 in Delaware during the summer of 2016 (red dots/line) compared to the 2003-2012 average (blue dots/line), overlaid with the best polynomial fit to the 2003-2012 average (black line). The trend line illustrates the historical climatological O_3 cycle, featuring a ramp up in May, the seasonal peak in June-July, a downward trend in August, and a sharp drop off in September. The 2016 daily observed O_3 values do not follow this seasonal cycle. Instead, the 2016 data roughly delineate a straight line, with periods of peak O_3 from late May to late September. In this way, the O_3 climatology is now spread out across the entire O_3 summer season, with maximum O_3 likely at almost any time. This change makes O_3 climatology a much weaker foundation for the daily O_3 forecast in the New Normal period compared to previous years.

4. Skill of Ozone Forecasts in 2016

Despite the challenges presented by a new O_3 exceedance threshold and breakdowns in historical forecast predictors, forecasts of O_3 exceedance days in 2016 were skillful. A time series of forecasts and observations for 2016 is given in Figure 17, and Table 2 highlights (grey shading) the 7 O_3 exceedance days which were accurately forecasted, with health alerts issued to the public. Figure 18 shows the false alarm rate (red bars) and hit rate (blue bars) for O_3 exceedance day forecasts for 2011-2016. Note that prior to 2016, the exceedance threshold was 76 ppbv. Details on the calculation of threshold skill scores are given in Appendix A. Forecasts of O_3 exceedance days in 2016 were slightly less skillful than during the Post-NO_x SIP period (2011-2012), but they continued the positive trend begun in 2015, with the hit rate higher than the false alarm rate. The false alarm rate in 2016 was 0.33, which is a substantial improvement on 2015 (0.71). The hit rate of 0.55 in 2016 falls short of the 2015 hit rate of 1.0, but the sample size of hits in 2015 was only 2 (Figure 19). The continued improvement in false alarm rate and hit rate in 2016, the fourth year of the current New Normal period, demonstrates that forecasters have adjusted to the new observed O_3 environment and the new, lower O_3 exceedance threshold.

Overall forecast skill (all days) for 2016 was consistent with 2013-2015, as shown in Figure 20, with a median absolute error of 6.0 ppbv (same as 2013-2015). Forecasts in 2016 were slightly less biased, with a bias of 2.1 ppbv compared to ~3.0 ppbv in 2013-2015.

5. Performance of Ozone Numerical Forecast Models in 2016

In contrast to the skill of expert forecasts of O_3 exceedance days in 2016, numerical air quality models were remarkably unskillful. Figure 21 shows the false alarm rates (red bars) and hit rates (blue bars) of the numerical model guidance compared to the expert forecasts for Delaware in 2016. The O_3 numerical air quality models utilized in 2016 included the NOAA-EPA model (NAQFC), the North Carolina Division of Air Quality model (NCDAQ), and two versions of the Baron Advanced Meteorological Services model (BAMS), the CMAQ and RT. All of the models had false alarm rates > 0.50 and hit rates < 0.50 in 2016, making the expert forecasts far superior to numerical guidance.

Also shown in Figure 21 are the skill scores for a new statistical model developed for Delaware. This model was trained on data from the period 2013-2015 to reflect the recent changes in observed O_3. Statistical guidance was discontinued in Delaware in 2011 because changes in emissions of O_3 precursors associated with the NO_x SIP Rule rendered models unskillful that had been trained on data prior to 2003. Statistical models for the Philadelphia metropolitan area trained on the periods 2004-2013 and 2007-2013 were tested in 2015 and found to have poor skill, presumably because of the recent "step-down" in observed O_3 that began in 2013. It was hoped that the new Delaware statistical model, representing trends during the New Normal period, would make O_3 exceedance days easier to identify. Unfortunately, data in Figure 21 show that the statistical model had a high false alarm rate (0.50) and a very low hit rate (0.18) in 2016, indicating that the model was only able to identify 2 of the 11 O_3 exceedance days. Given the decoupling of O_3 exceedance days from hot weather and persistence O_3 in the New Normal period, statistical models seem unlikely to provide much forecast skill moving forward.

The poor skill of the NAQFC model in particular for the 2016 O_3 exceedance days was surprising, given the model's generally skillful performance historically in Delaware. Figure 22 shows the false alarm rate (red bars) and hit rate (blue bars) for the NAQFC in Delaware for 2011-2016. The skill of the model has suffered in the New Normal period, particularly in 2015 (false alarm rate of 1.00) and 2016 (hit rate of only 0.36). The model seems to have had trouble identifying hits in the New Normal period (Figure 23), which is driving down its hit rate, even for 2016 when the total number of O_3 exceedance days was higher. The performance of the NAQFC model for all days (Figure 24) has remained consistent in the New Normal period, however, with a median absolute effort or ~6.0 ppbv and a bias of ~2.0 ppbv, analogous to the overall skill of the expert forecasts. This suggests that the NAQFC model provides useful forecast guidance in general, but it cannot be relied upon to differentiate between high Code Yellow and Code Orange O_3 days.

6. Growing Importance of Transported Wildfire Smoke

Smoke transported from wildfires is becoming a major factor for local O_3 production in Delaware. Smoke contains high concentrations of NO_x and reactive hydrocarbons, which are precursors for O_3 production. When smoky air is transported into Delaware, the presence of additional NO_x and hydrocarbons can amplify local O_3 production substantially. In the New Normal period, where anthropogenic NO_x emissions have been significantly reduced, the most widespread and highest magnitude O_3 exceedances appear to occur on days when transported smoke is present.

In 2016, 3 of the 11 O_3 exceedance days (May 25-26 and July 22) were impacted by smoke, including the Code Red day on May 25 (Table 2). These days include the two highest magnitude O_3 exceedances of the entire 2016 season, May 25 (86 ppbv) and July 22 (84 ppbv). The May event (Figure 25) was particularly widespread, with Code Orange O_3 observed across the Mid-Atlantic region, New York, and southern New England, with embedded pockets of Code Red O_3 along the Interstate-95 Corridor, including northern Delaware. This event was driven by dilute smoke transported from the Fort McMurray fire in Alberta, Canada, which began on May 1.

It can be difficult to know when dilute smoke is present in an air mass. Satellite aerosol imagery can identify smoke if it is optically thick enough, but often by the time smoke from fires in the western U.S., Canada, or Alaska (the most common fire source regions during the summertime) reaches the Mid-Atlantic region, it is not evident in satellite imagery. This was the case for the May 25-26 event, which was not immediately recognizable as smoke-influenced. Forecasters thought the air mass moving into the Mid-Atlantic region on May 25 was a typical of a historical regional high O_3 event, which led to under-forecasting on May 25 (Figure 17), when the NO_x-enhanced smoky air was directly over the Mid-Atlantic, and over-forecasting on May 26 (Figure 17), when the bulk of the smoky air had shifted eastward into New England. In addition, current numerical air quality models do not include smoke in their boundary conditions, so they are not useful for predicting smoke-influenced O_3 events. Looking ahead, aerosol products from the new GOES-16 satellite may help indicate smoke transport. GOES-16, launched in November 2016, will routinely provide aerosol observations every 15 minutes over the full disk of the Earth, covering North and South America, and every 5 minutes over the continental U.S. These high temporal resolution aerosol observations may be available in time for testing during the end of the 2017 O_3 season.

7. Outlook for 2017

In 2017, a continuation of the New Normal period is expected, with 5-12 O_3 exceedance days likely for Delaware. The recent trends toward more isolated, single-day O_3 exceedance “spikes” and the growing importance of transported wildfire smoke are expected to continue. Identifying the exceedance days will remain difficult, given the breakdown in the historical forecast predictors of temperature, persistence, and climatology, as well as the rising influence of transported smoke. NOAA has implemented improvements to the NAQFC for the 2017 O_3 season, including a new version of the NAM meteorological model, which will hopefully improve the NAQFC's skill in identifying O_3 exceedances. The inconsistent performance of the numerical air quality models in the New Normal period has added to the forecast challenges. Nevertheless, the recent improvements in hit rate and false alarm rate for the expert forecasts are expected to continue, as forecasters become acclimated to the New Normal conditions.

8. Summary

For summer 2016, a new 8-hour average O_3 exceedance threshold of 71 ppbv was in effect. In 2016, there were 11 days with observed 8-hour average $O_3 \geq 71$ ppbv, 1 day with observed 8-hour average $O_3 \geq 86$ ppbv, and no days with observed 8-hour average $O_3 \geq 106$ ppbv. This translated into 10 Code Orange days and 1 Code Red day on the updated AQI scale.

The recent substantial downward trend in observed O_3 in Delaware, which began in 2013, continued in 2016. As a result, the 2016 O_3 season was the fourth consecutive historically low O_3 season for Delaware in terms of O_3 exceedances and seasonal mean/median O_3. Three distinct periods are now evident in historical observed O_3: the Pre-NO_x SIP period (prior to 2003), the Post-NO_x SIP period (2003-2012), and the New Normal period (2013-2016). The Pre-NO_x SIP period had historically high O_3 (average of 49.8 exceedance days). An initial "step-down" in observed O_3 occurred during the Post-NO_x SIP period (average of 29.3 exceedance days), which has been attributed to regional reductions in NO_x emissions from EGUs in the eastern U.S. associated with the "NO_x SIP Rule." Another, temporary instance of a "step-down" in O_3 precursor emissions and concentrations occurred during the Great Recession in 2009, when only 5 O_3 exceedance days were observed in Delaware. The New Normal period is characterized by a second "step-down" in observed O_3 (average of only 8.3 exceedance days). This 2013-2016 "step-down" is not unique to Delaware; the same trend is evident across the Mid-Atlantic region.

There is mounting evidence that continued reductions in regional NO_x emissions, particularly from the upwind ORV source region, are the primary cause of the recent "step-down" in observed O_3 during the New Normal period. No other explanation captures the daily and seasonal trends in observed O_3 in the 2013-2016 period. The argument that changes in weather conditions, related to the cooler and wetter summers of 2013-2014, were primarily responsible for the lower observed O_3 levels since 2013 has been refuted by the warm summers of 2015 and especially 2016.

During the New Normal period, the traditional conceptual model for O_3-conducive weather patterns, driven by synoptic scale weather and westerly transport effects, is breaking down. As a result, the historical forecast tools of temperature, persistence O_3, and climatology are no longer reliable for predicting O_3 exceedance days. Mesoscale features, such as convection, local emissions, and recirculation, are becoming greater contributing factors for observed O_3. These phenomena all occur on the local scale, which means they are more difficult for numerical weather models to forecast, and thus, add uncertainty to the air quality forecast.

Despite the challenges presented by a new O_3 exceedance threshold and breakdowns in historical forecast predictors, forecasters are adjusting to the New Normal period. In 2016, the hit rate for forecasts of O_3 exceedance days was 0.55 and the false alarm rate was 0.33. The false alarm rate was a substantial improvement from 2015 (0.71). Overall forecast skill for all days in 2016 was consistent with 2013-2015, with a median absolute error of 6.0 ppbv (same as 2013-2015). Forecasts in 2016 were slightly less biased than recent years, with a bias of 2.1 ppbv compared to ~3.0 ppbv in 2013-2015.

In contrast to the skill of expert forecasts of O_3 exceedance days in 2016, numerical air quality models were very unskillful. All of the models used as guidance in 2016 had false alarm rates > 0.50 and hit rates < 0.50, making the expert forecasts far superior to numerical guidance and an updated statistical model for Delaware.

Smoke transported from wildfires is becoming a major factor for local O_3 production in Delaware, with 3 of 11 O_3 exceedance days in 2016 attributable to smoke. In the New Normal period, the most widespread O_3 exceedances appear to occur on days when transported smoke is present. Identifying these days will be a challenge moving forward, as there are no dependable forecast tools for smoke, although new high temporal resolution GOES-16 aerosol products have promise.

In 2017, a continuation of the New Normal period is expected, with 5-12 O_3 exceedance days likely for Delaware. Identifying the exceedance days will remain difficult, given the breakdown in the historical forecast predictors of temperature, persistence, and climatology, as well as the recent poor performance of the numerical air quality models and the rising influence of transported smoke.

Tables and Figures

Table 1. Air Quality Index (AQI) color codes and breakpoints for O_3 in 2008-2015 and 2016.

Color	Category	2008-2015 O_3 Breakpoints (ppbv)	2016 O_3 Breakpoints (ppbv)
Green	Good	0-59	0-54
Yellow	Moderate	60-75	55-70
Orange	Unhealthy for Sensitive Groups	76-95	71-85
Red	Unhealthy	96-115	86-105
Purple	Very Unhealthy	116-374	106-200

Table 2. Details regarding observed O_3 exceedance days in Delaware in 2016. Grey shading indicates days that O_3 exceedances were correctly forecasted, with health alerts issued to the public. T_{max} indicates maximum temperature at Dover Air Force Base (DOV). * indicates the exceedance was based a 6-hour O_3 running average instead of the standard 8-hour running average.

Date	Day of Week	T_{max} DOV (°F)	Maximum 8-Hr O_3 (ppbv)	Number of Monitors	Conditions
5/25	Wed	86	86	6	Transported smoke
5/26	Thu	89	75	5	
6/11	Sat	90	74	4	Upwind transport; hot; previous day Good
6/20	Mon	90	73	2	Westerly transport aloft; hot
7/8	Fri	94	77*	1	Sea breeze at Lewes monitor
7/22	Fri	91	84	4	Transported smoke
7/27	Wed	93	71*	1	Recirculating surface winds; hot
8/27	Sat	90	73	2	Recirculating surface winds; hot; mid-level ridge overhead
9/14	Wed	90	73	2	Pre-frontal; hot; light southwesterly surface winds
9/22	Thu	82	78	3	Suspected local power plants running without seasonal NOx emission controls
9/23	Fri	86	82	3	

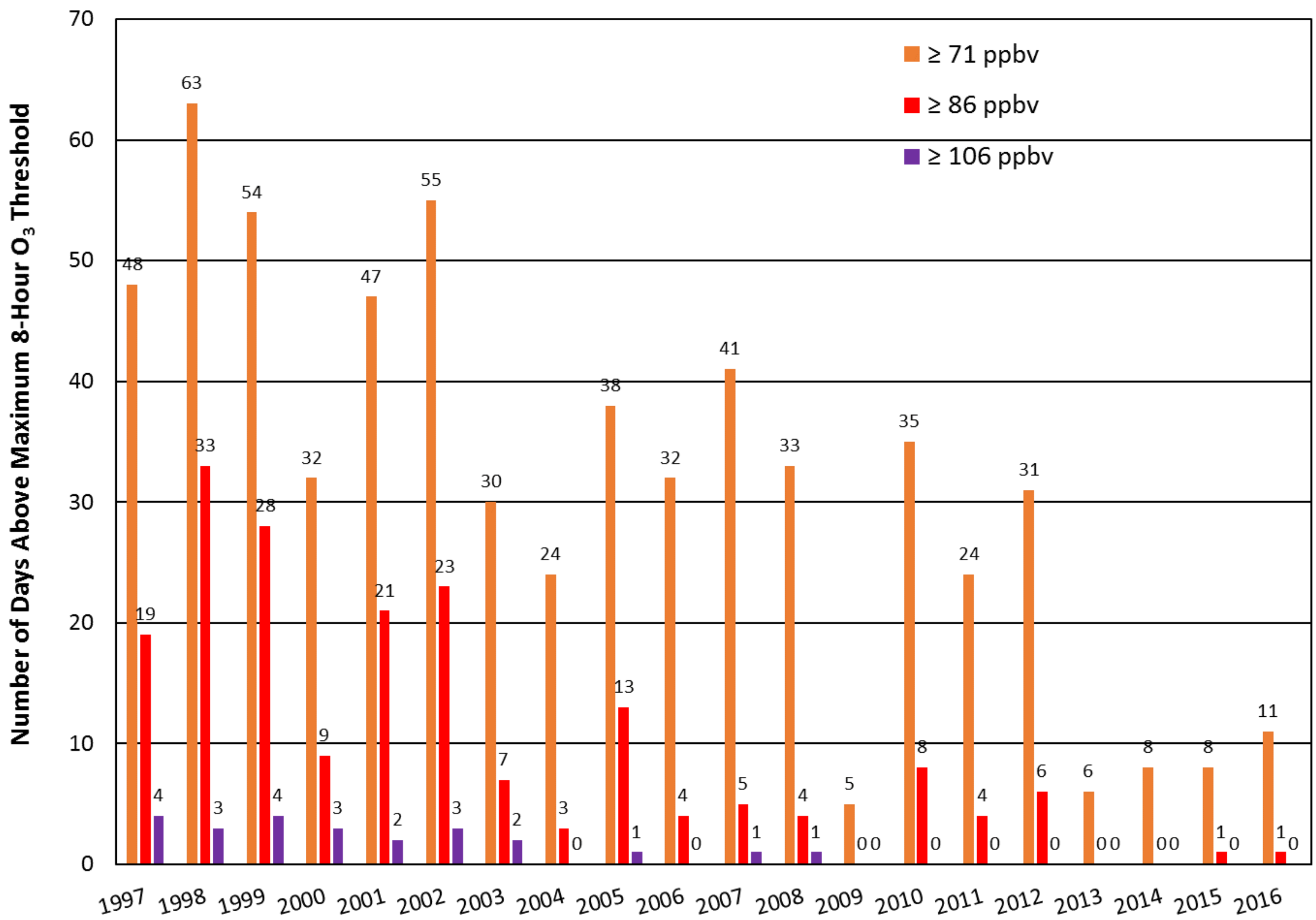

Figure 1. Number of days with maximum 8-hour observed O_3 at or above thresholds of 71 ppbv (orange bars), 86 ppbv (red bars), and 106 ppbv (purple bars) in Delaware for 1997-2016.

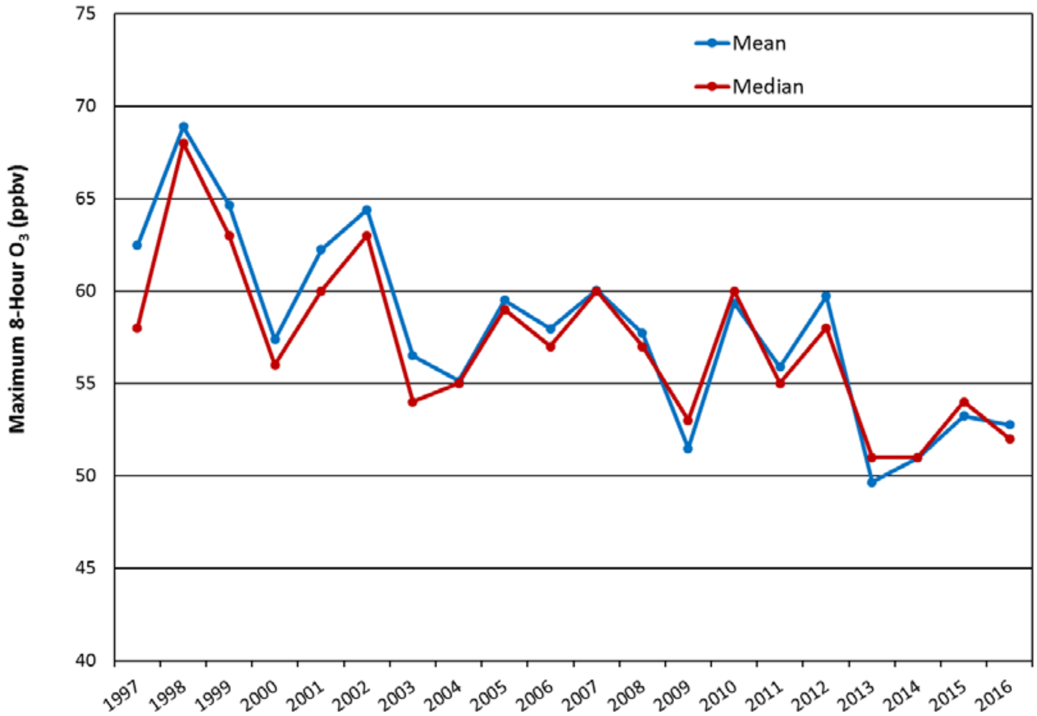

Figure 2. Seasonal (May 1 to September 30) mean and median maximum observed 8-hour O_3 in Delaware for 1997-2016.

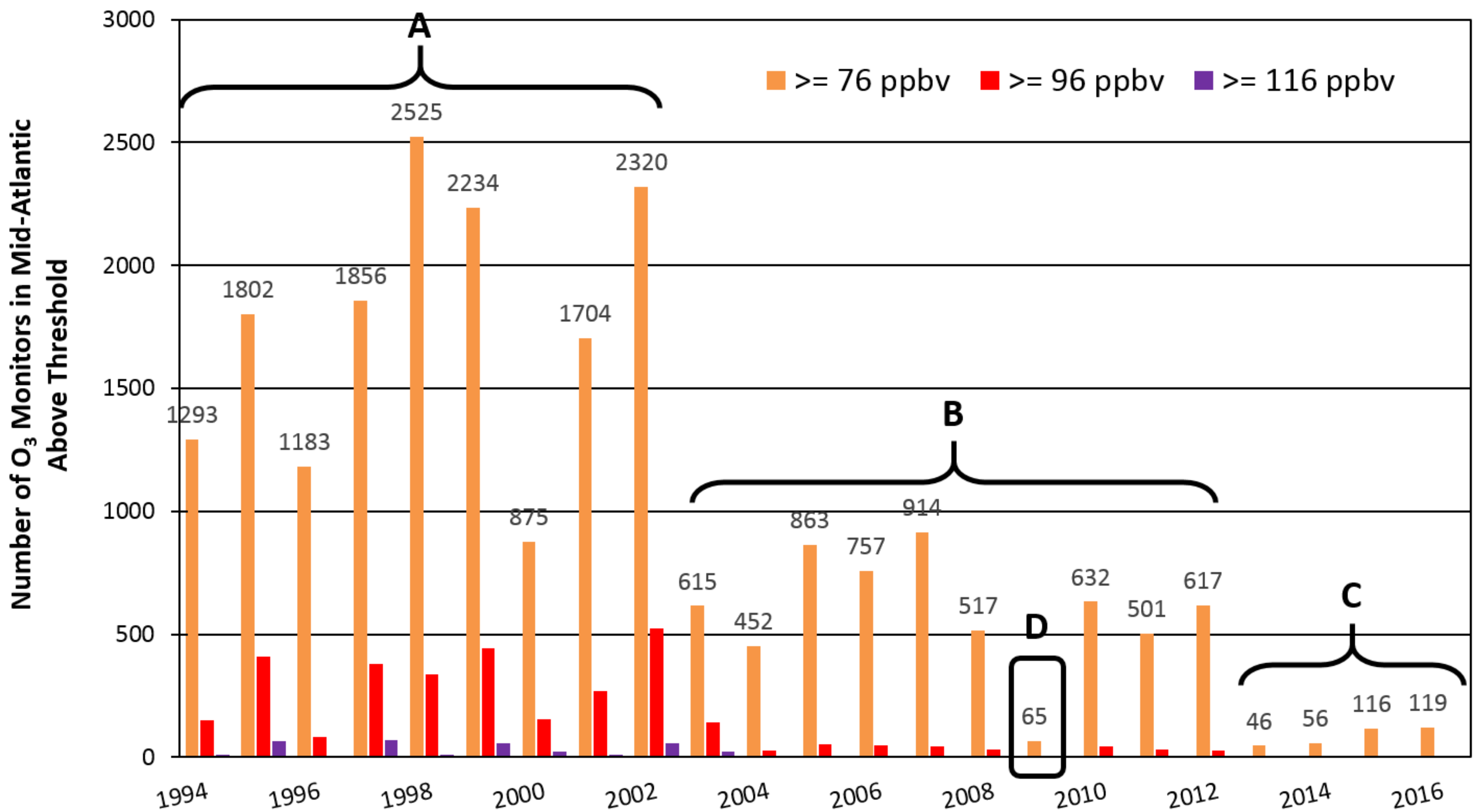

Figure 3. The number of monitors in the Mid-Atlantic region with maximum 8-hour observed O_3 above given thresholds for 1994-2016. The numbers above the orange bars indicate the number of monitors with $O_3 \geq 76$ ppbv. The Pre-NO_x SIP period (A; years up to 2002), the Post NO_x SIP period (B; 2003-2012), and the New Normal period (C; 2013-2016) are indicated by brackets. The Great Recession year of 2009 (D) is circled.

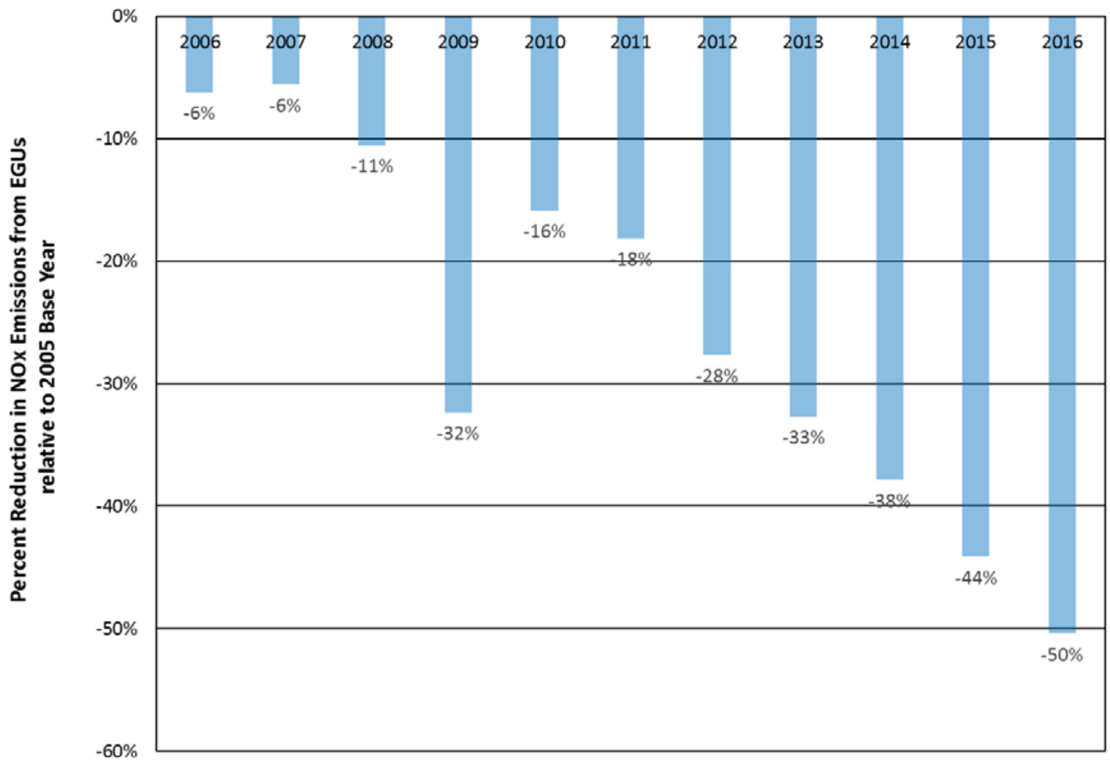

Figure 4. Annual percent reduction in emissions of nitrogen oxides (NO_x) from energy generating units (EGUs) in local and upwind states for 2006-2016 relative to base year of 2005 (U.S. EPA's Air Markets Program Data).

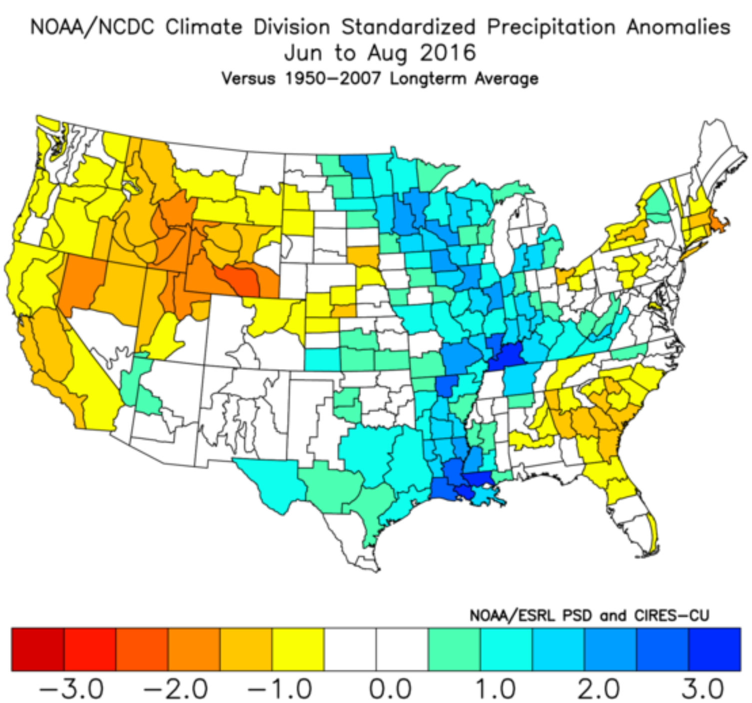

Figure 5. Precipitation anomalies (in inches) in the U.S. for June-August 2016 compared to the 1950-2007 average (courtesy of NOAA/ESRL).

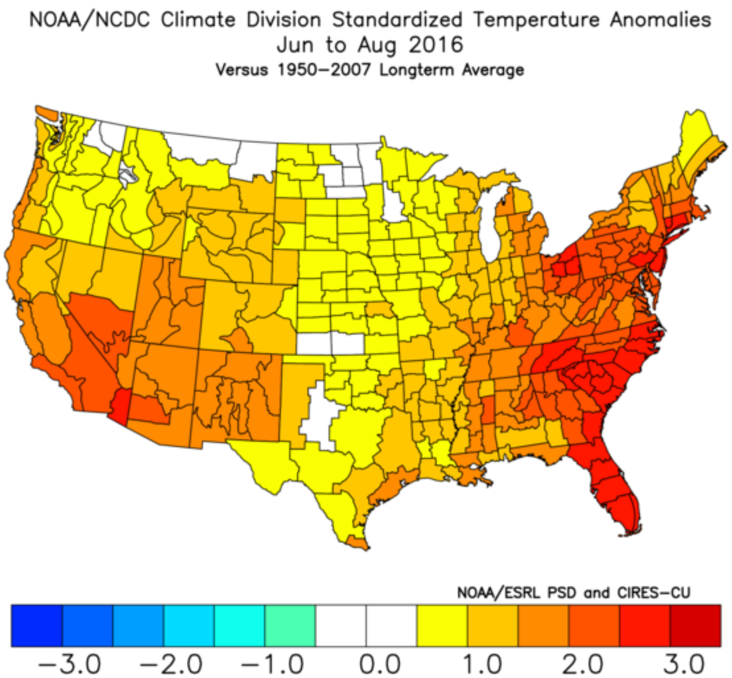

Figure 6. Temperature anomalies (in °F) in the U.S. for June-August 2016 compared to the 1950-2007 average (courtesy of NOAA/ESRL).

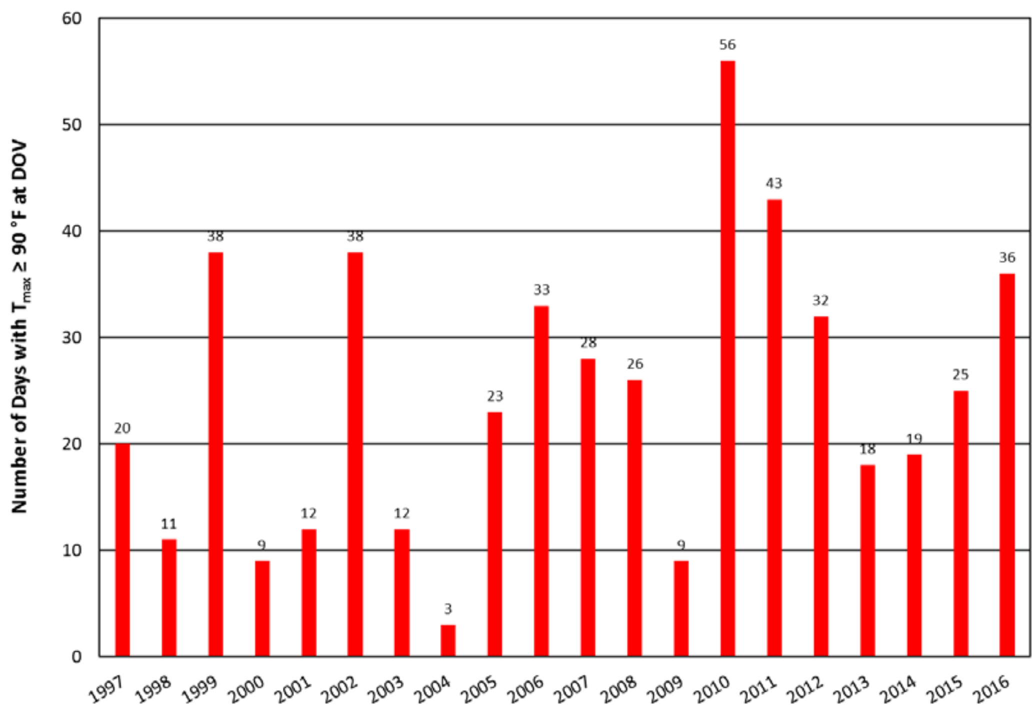

Figure 7. Number of days with maximum temperature (T_{max}) ≥ 90 °F at Dover Air Force Base (DOV) for May 1 to September 30, 1997-2016.

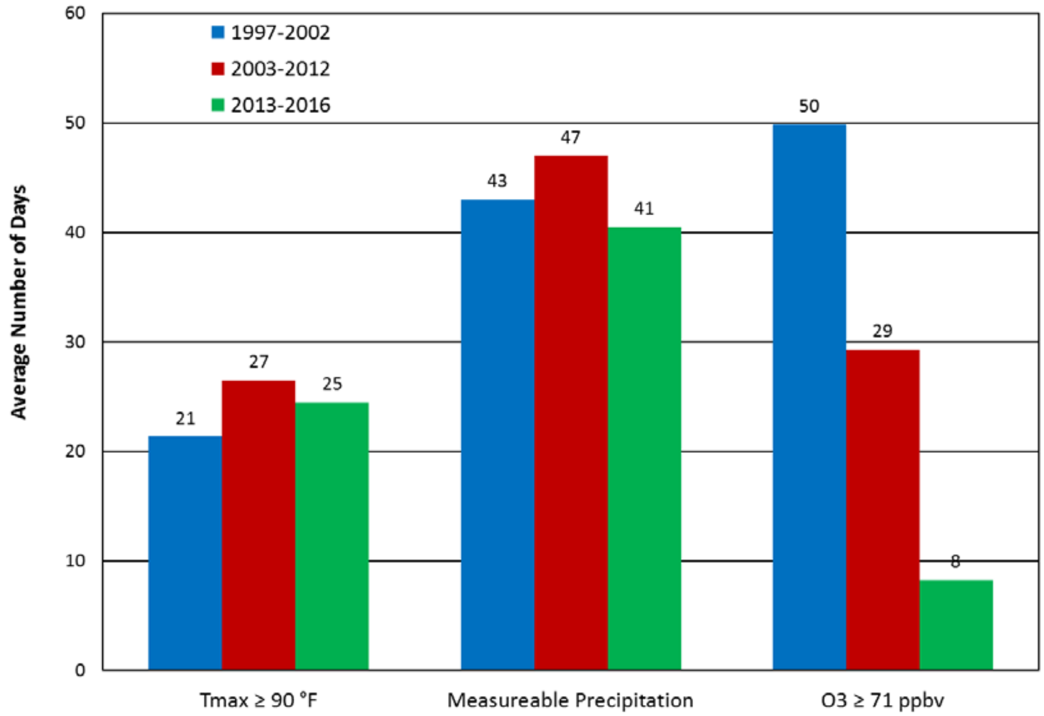

Figure 8. Average number of days with maximum temperature (T_{max}) ≥ 90 °F at DOV, measureable precipitation at DOV, and maximum 8-hour observed O_3 ≥ 71 ppbv in Delaware, May 1 to September 30 for specified periods.

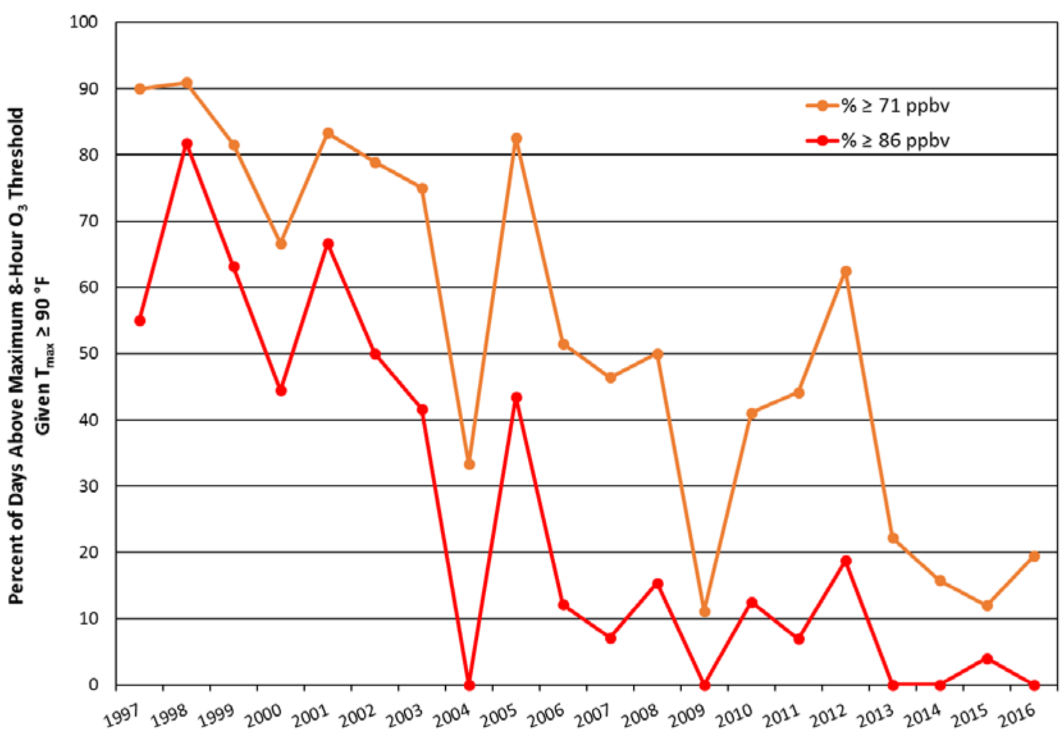

Figure 9. Percent of days with maximum temperature (T_{max}) ≥ 90 °F at DOV and maximum 8-hour observed O_3 ≥ 71 ppbv (orange dots) or ≥ 86 ppbv (red dots) in Delaware for 1997-2016.

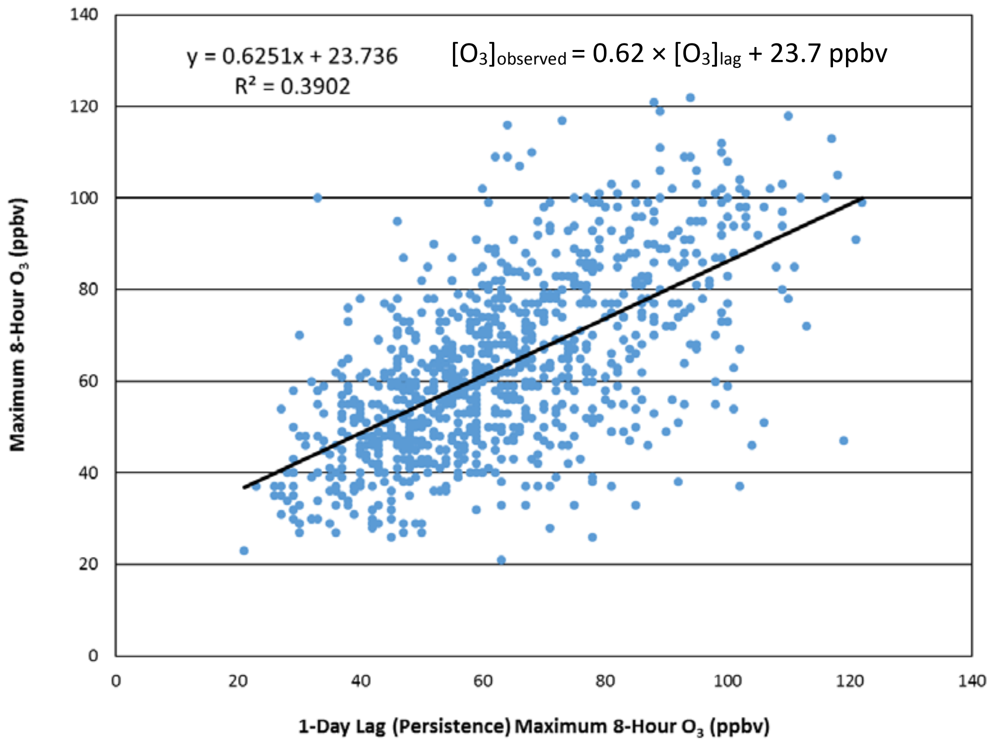

Figure 10. Linear relationship between seasonal (May 1 to September 30) maximum 8-hour observed O_3 and local 1-day lag (persistence) maximum 8-hour observed O_3 in Delaware for 1997-2002.

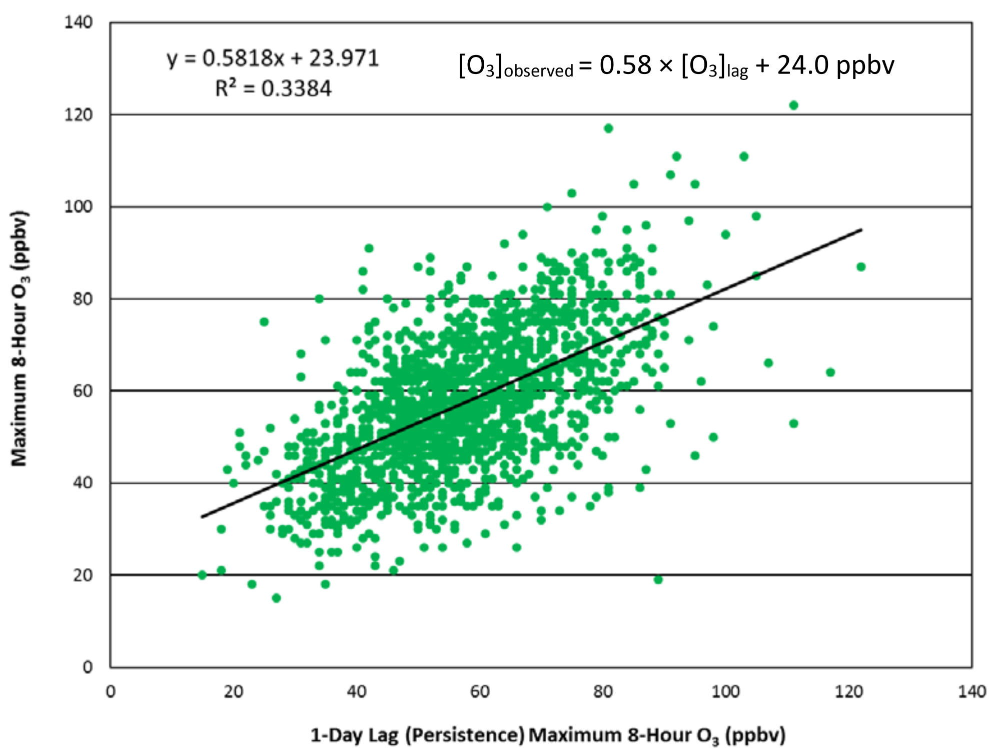

Figure 11. Linear relationship between seasonal (May 1 to September 30) maximum 8-hour observed O_3 and local 1-day lag (persistence) maximum 8-hour observed O_3 in Delaware for 2003-2012.

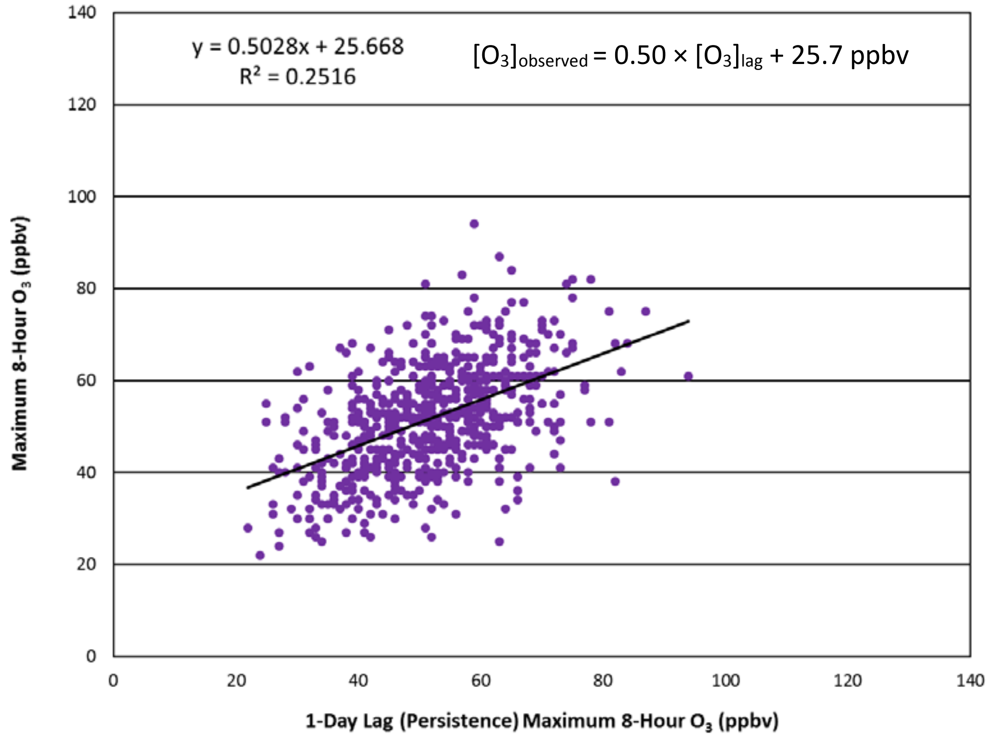

Figure 12. Linear relationship between seasonal (May 1 to September 30) maximum 8-hour observed O_3 and local 1-day lag (persistence) maximum 8-hour observed O_3 in Delaware for 2013-2016.

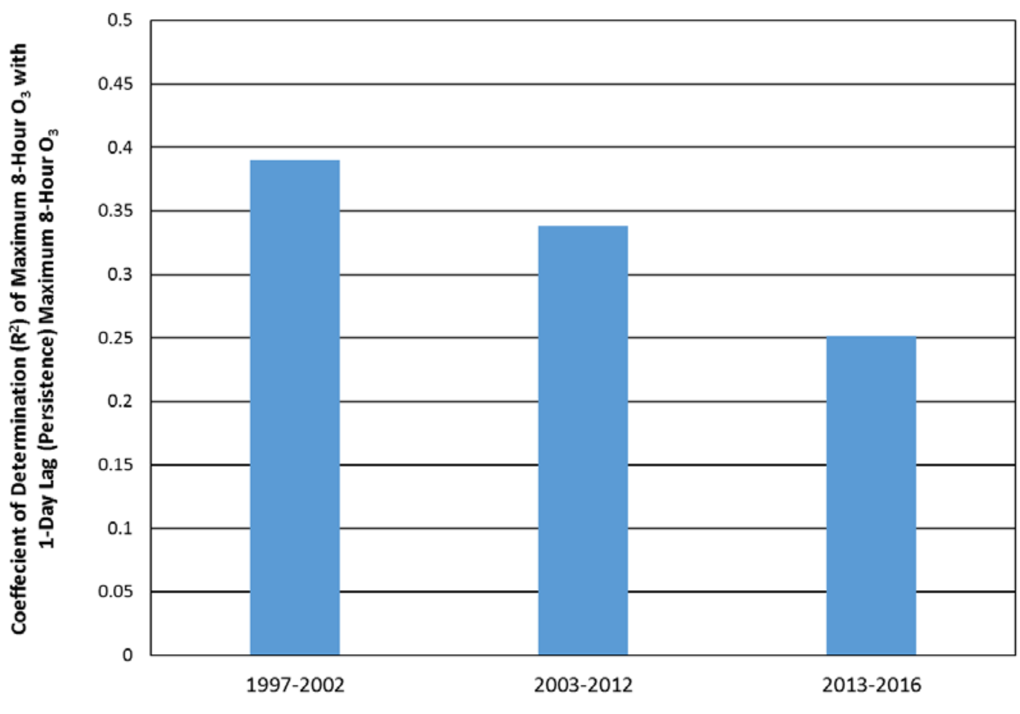

Figure 13. Coefficient of determination (R^2) of linear relationship between seasonal (May 1 to September 30) maximum 8-hour observed O_3 and local 1-day lag (persistence) maximum 8-hour observed O_3 in Delaware for given periods.

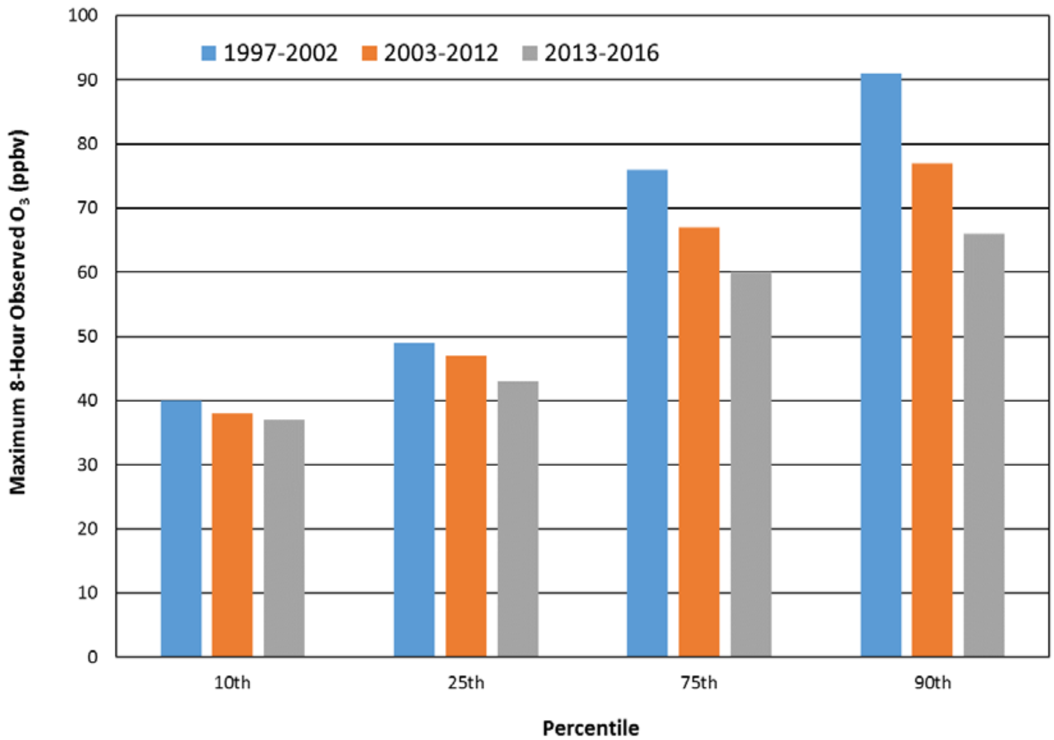

Figure 14. Maximum 8-Hour O_3 value corresponding to the 10th, 25th, 75th, and 90th percentiles in Delaware for given periods.

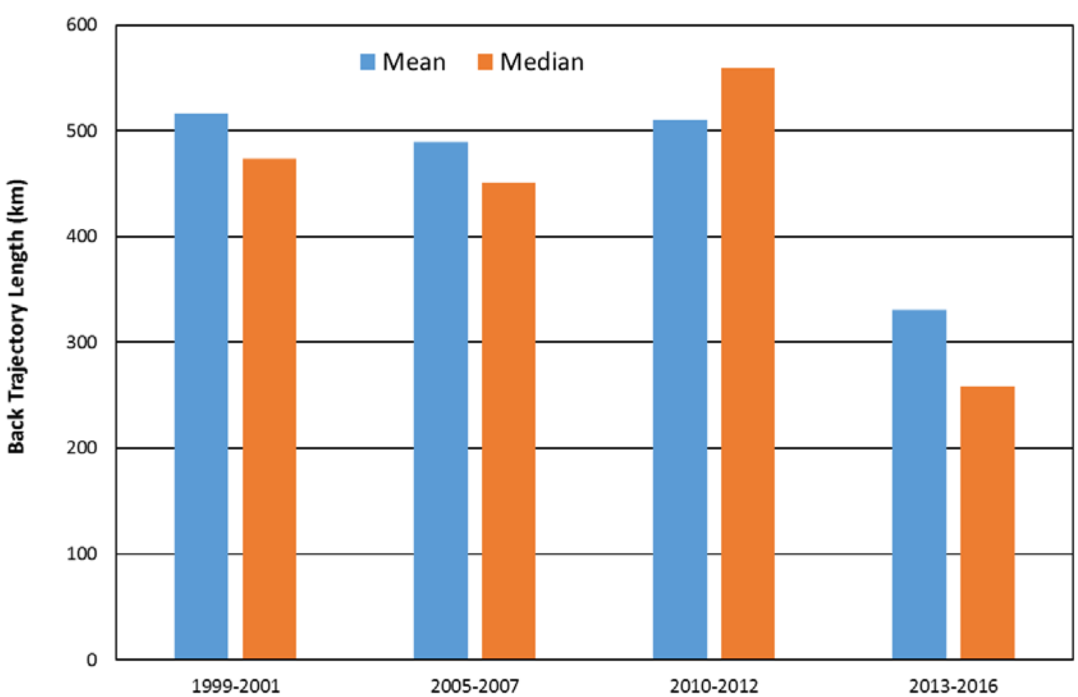

Figure 15. Mean (blue bars) and median (orange bars) of 24-hour backward airmass trajectory analyses, initiated at 1000 m AGL and ending at KPHL, for the highest 96th percentile maximum 8-hour observed O_3 days in the Philadelphia metropolitan area for given periods.

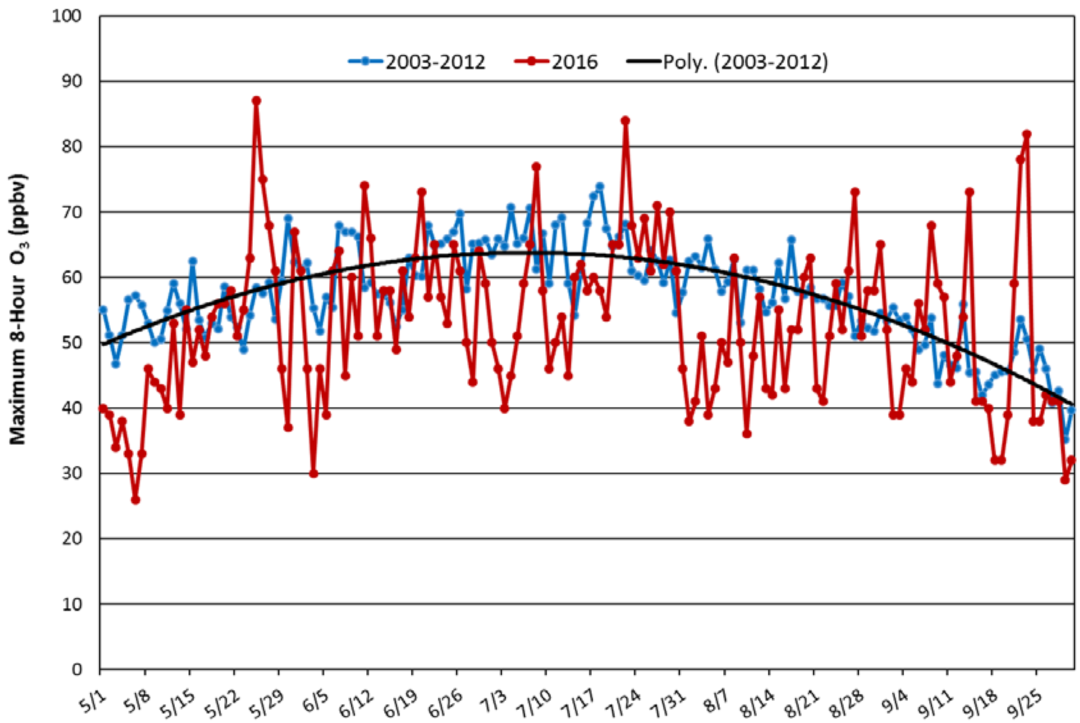

Figure 16. Daily time series of maximum 8-hour observed O_3 in Delaware for 2016 (red line) compared to the 2003-2012 average (blue line). The black line is the best polynomial fit to the 2003-2012 average.

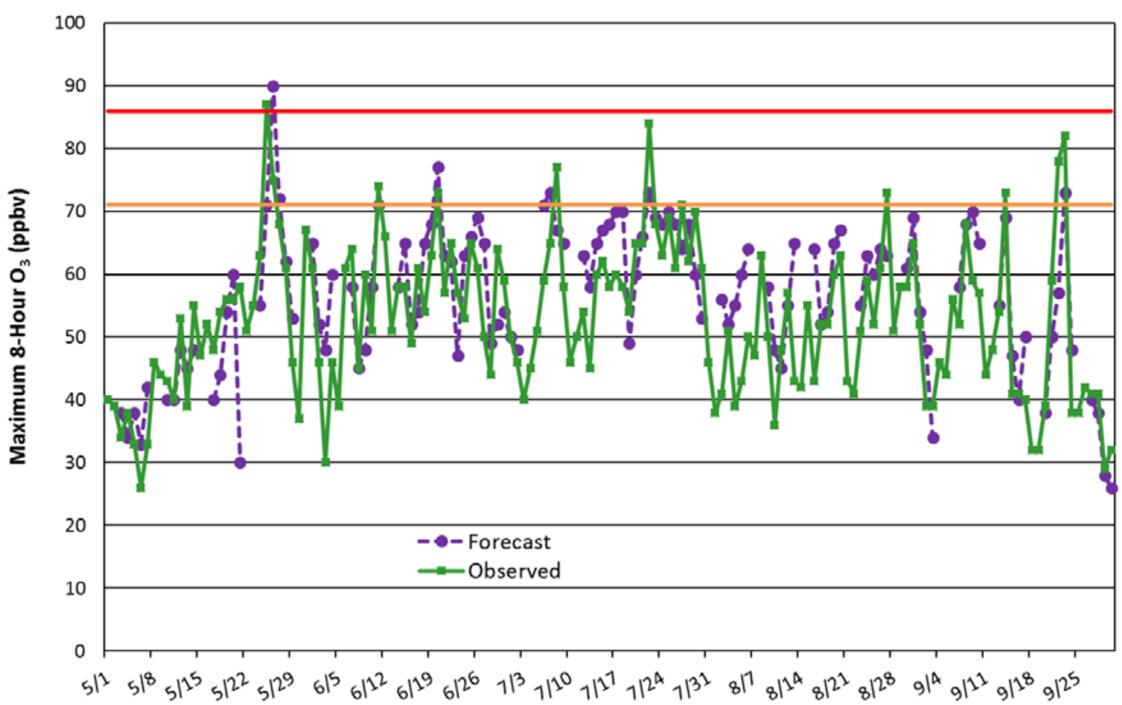

Figure 17. Maximum 8-hour O_3 forecasts and observations for Delaware during May 1 to September 30, 2016. The orange and red lines indicate the new, lower Code Orange and Code Red thresholds, respectively.

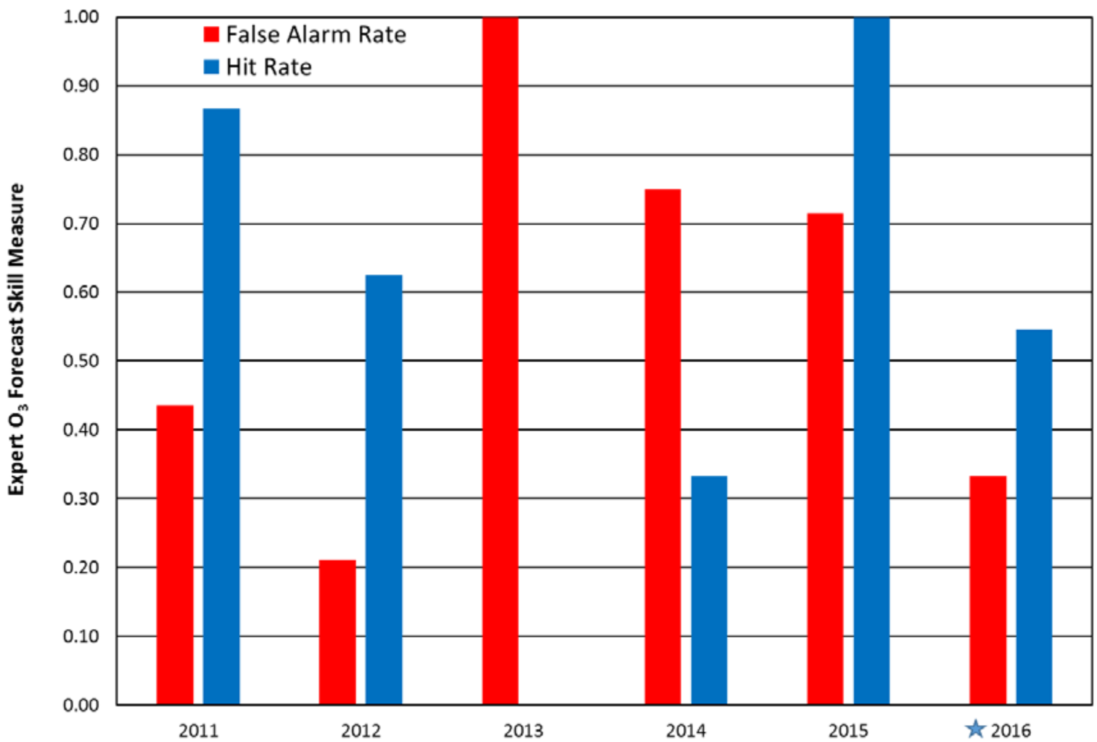

Figure 18. False alarm rate and hit rate for threshold forecasts of maximum 8-hour O_3 in Delaware for 2011-2016. Note that for 2011-2015, the threshold was 76 ppbv, but in 2016, it lowered to 71 ppbv.

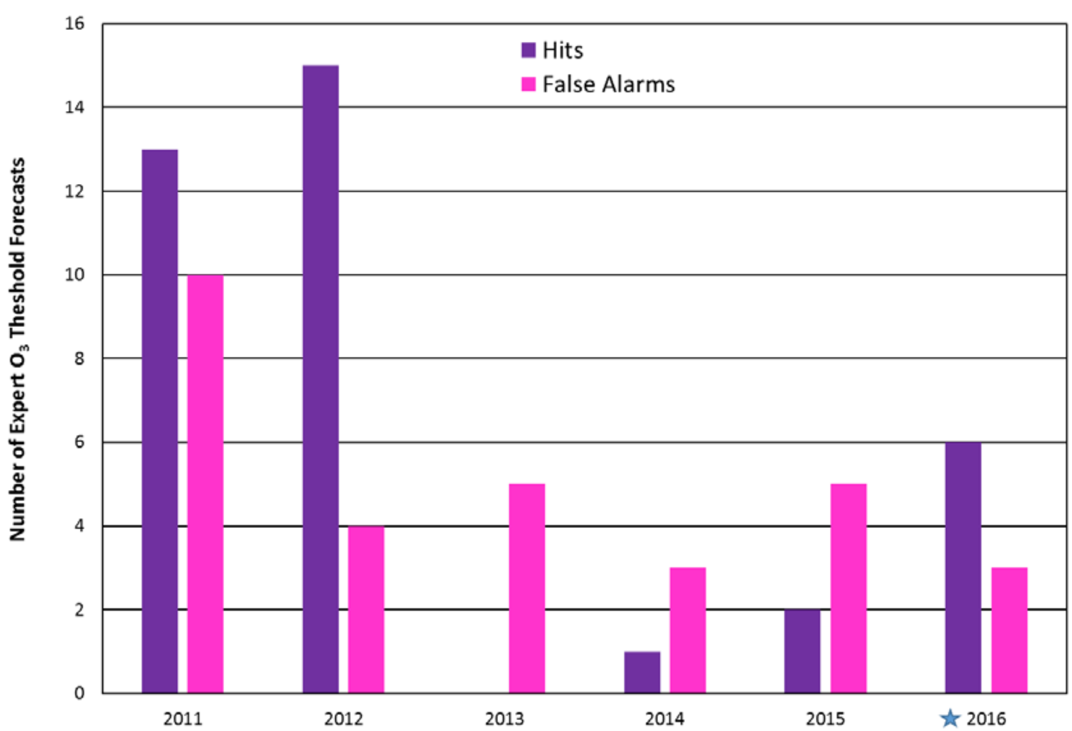

Figure 19. Number of false alarms and hits for threshold forecasts of maximum 8-hour O_3 in Delaware for 2011-2016. Note that for 2011-2015, the threshold was 76 ppbv, but in 2016, it lowered to 71 ppbv.

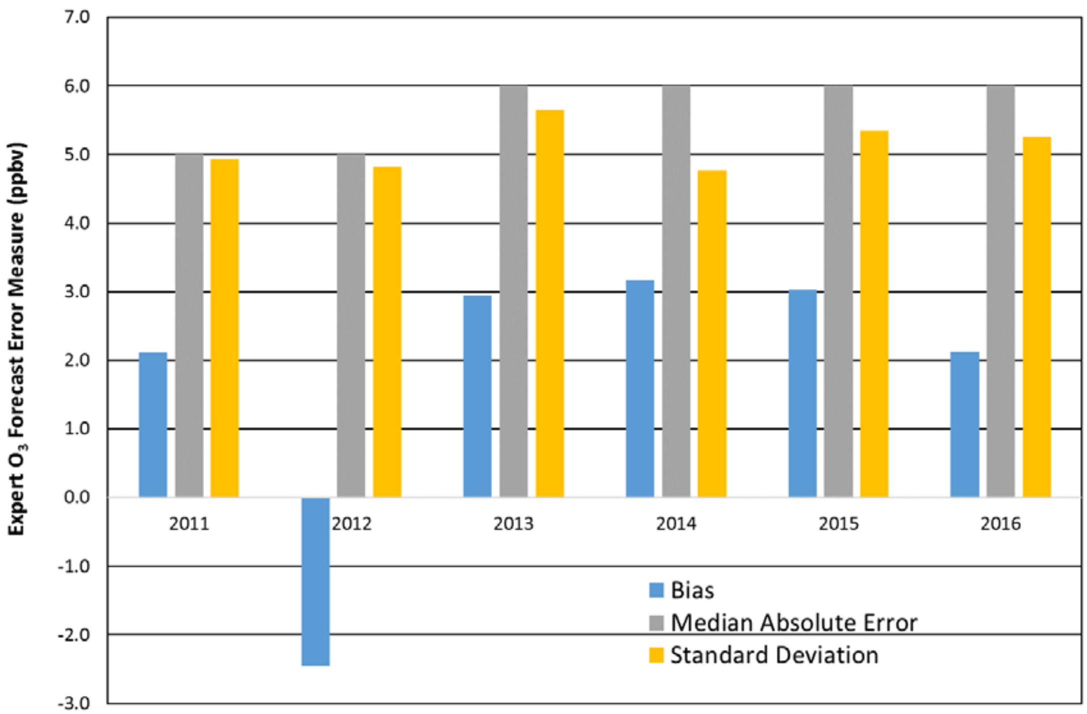

Figure 20. Error statistics for all maximum 8-hour O_3 forecasts, May 1 to September 30, in Delaware for 2013-2016.

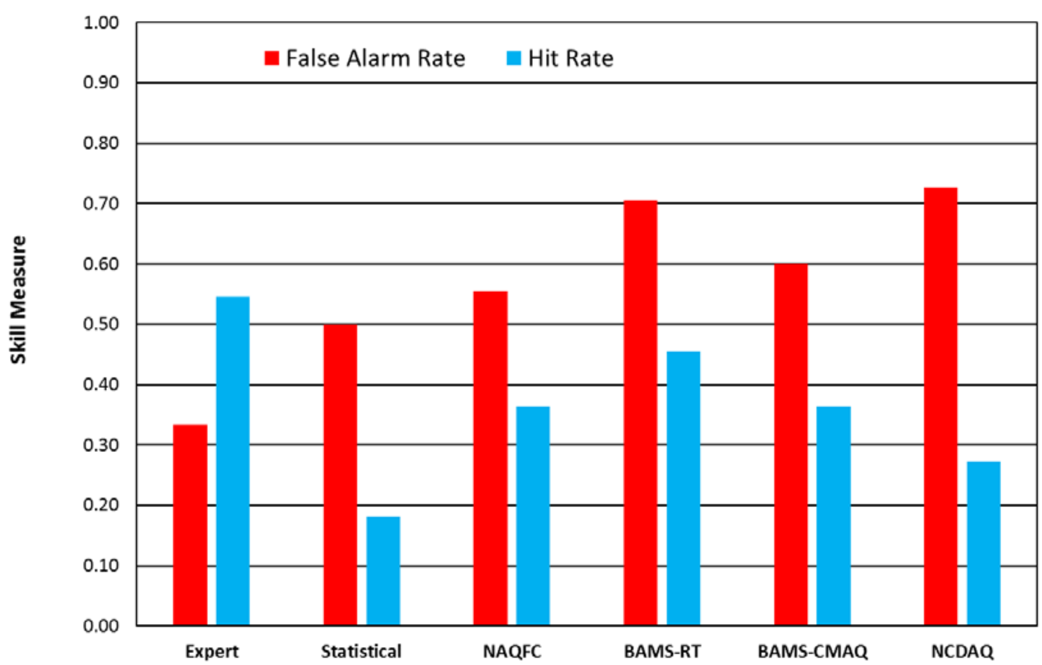

Figure 21. False alarm rate and hit rate for expert, statistical model, and numerical air quality model threshold predictions of maximum 8-hour O_3 in Delaware for 2016. Two variations of the Baron Meteorological Services (BAMS) models are shown (CMAQ and RT), as well as the North Carolina Division of Air Quality (NCDAQ) model.

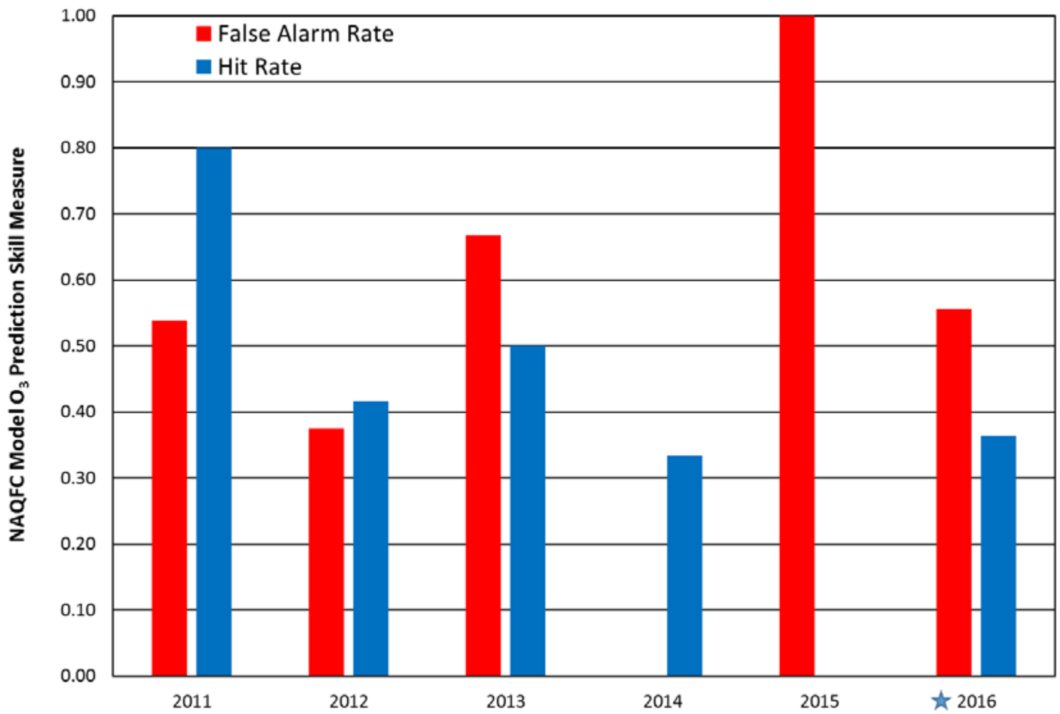

Figure 22. False alarm rate and hit rate for threshold predictions of maximum 8-hour O_3 by the NAQFC model in Delaware for 2011-2016. Note that for 2011-2015, the threshold was 76 ppbv, but in 2016, it lowered to 71 ppbv.

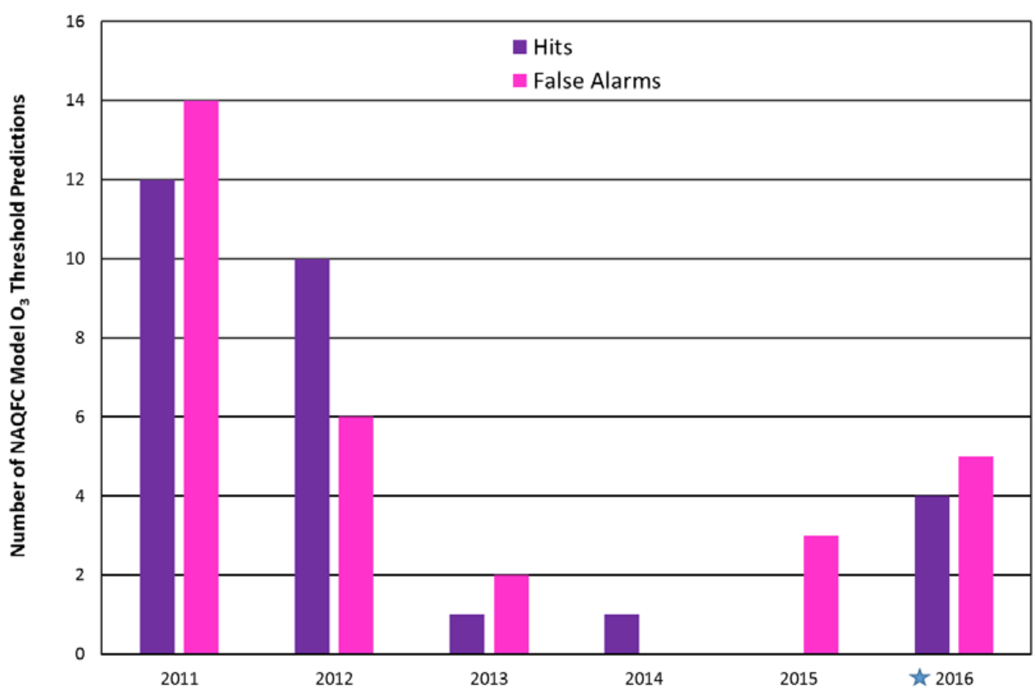

Figure 23. Number of false alarms and hits for threshold predictions of maximum 8-hour O_3 by the NAQFC model in Delaware for 2011-2016. Note that for 2011-2015, the threshold was 76 ppbv, but in 2016, it lowered to 71 ppbv.

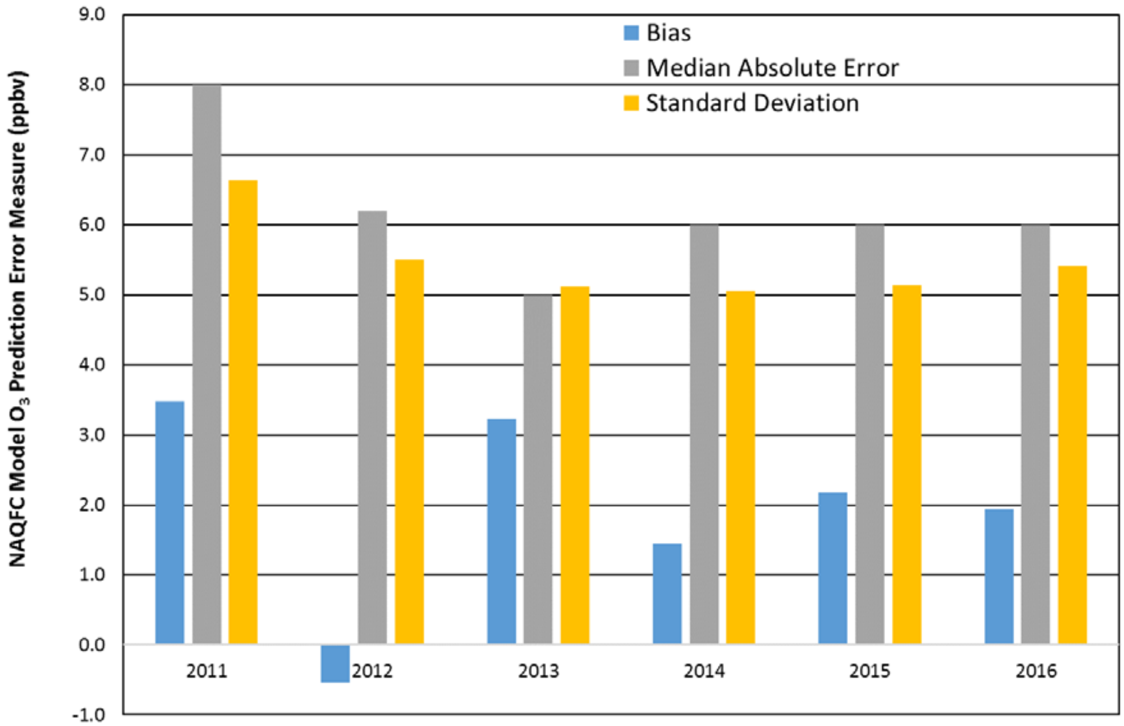

Figure 24. Error statistics for all maximum 8-hour O_3 predictions by the NAQFC model, May 1 to September 30, in Delaware for 2013-2016.

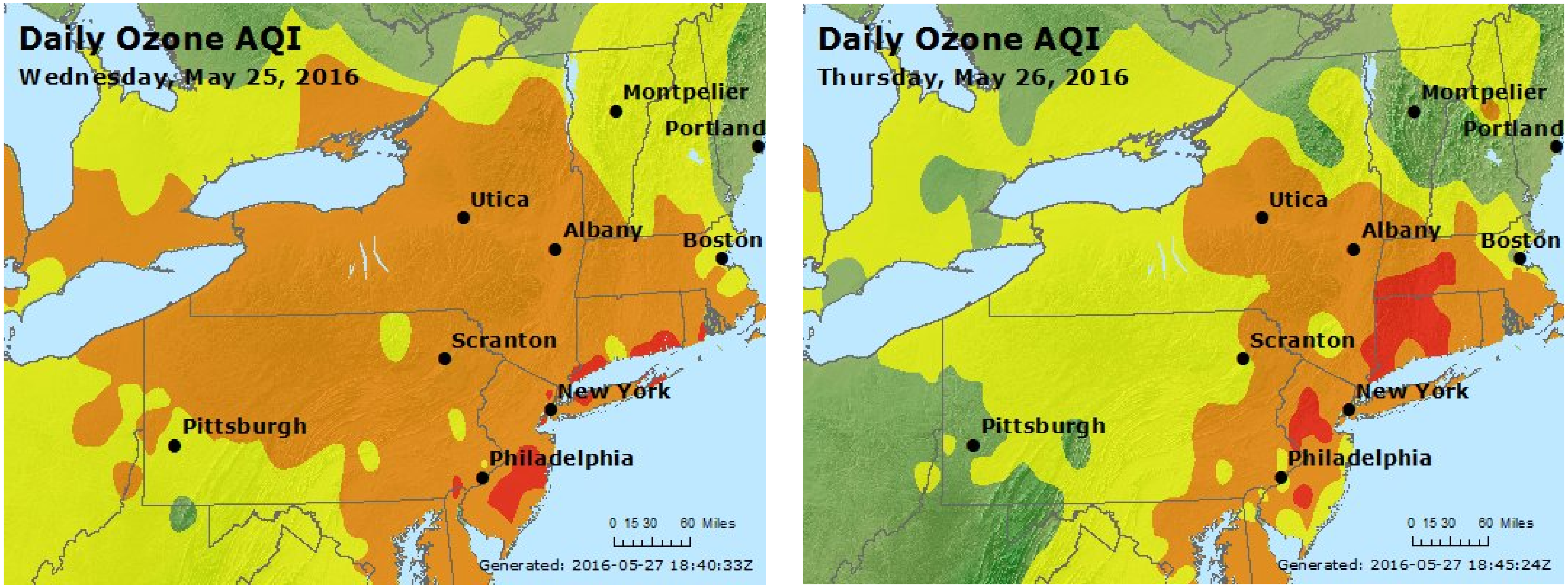

Figure 25. Observed O_3 air quality index (AQI) color codes for the northern Mid-Atlantic and southern New England regions on May 25 (left) and May 26 (right), 2016, when dilute wildfire smoke transported from the Fort McMurray wildfire in Alberta, Canada contributed to regional Code Orange O_3 conditions with pockets of Code Red O_3.

Appendix A. Skill Measures for Threshold Forecasts

The determination of the skill of a threshold forecast (e.g., O_3 exceedance) begins with the creation of a contingency table of the form of Table 3. For example, if an O_3 exceedance is both observed and forecasted ("hit"), then one unit is added to "a." If an O_3 exceedance is forecasted but not observed ("false alarm"), then one unit is added to "b."

Table 3. 2 x 2 contingency table for threshold forecasts.

		Observed	
		Yes	**No**
Forecasted	**Yes**	a	b
	No	c	d

Basic Skill Measures

A basic set of skill measures are determined from the contingency table and then used as the basis for further analysis.

Bias (B) determines whether the same fraction of events are both forecasted and observed. If B = 1, then the forecast is unbiased. If B < 1 there is a tendency to under-predict and if B > 1 there is a tendency to over-predict.

$$B = \frac{a+b}{a+c} \qquad (1)$$

False alarm rate (FAR) is a measure of the rate at which threshold forecasts are made but do not verify.

$$FAR = \frac{b}{a+b} \qquad (2)$$

The hit rate (HR) is a measure of the rate at which threshold forecast are made and verify. Hit rate is often called the "probability of detection."

$$HR = \frac{a}{a+c} \qquad (3)$$

References

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

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