State of Delaware
Final Report: Ozone Observations and Forecasts in 2017

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, 2018

Table of Contents

List of Tables

List of Figures

Executive Summary

- 2017 marked the second season under the new 8-hour average ozone (O_3) exceedance threshold of 71 ppbv.
- In 2017, there were 8 days with observed 8-hour average $O_3 \geq 71$ ppbv, no days 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 8 Code Orange days (71-85 ppbv) on the Air Quality Index (AQI) scale.
- The slight drop in the number of O_3 exceedance days (defined as observed 8-hour average $O_3 \geq 71$ ppbv) in 2017 compared to 2016 (11) is attributed to differences in meteorology; summer 2016 was unusually warm compared to summer 2017, which was about average in terms of temperature.
- The recent downward trend in the number of O_3 exceedance days in Delaware, which began in 2013, continued in 2017. This trend is not unique to Delaware; the same trend is evident across the entire Mid-Atlantic region.
- The 2017 seasonal (May 1 to September 30) mean and median of the maximum observed 8-hour average O_3 (49 ppbv) are the lowest observed in Delaware, going back to 1997.
- The period 2013-2017, termed the New Normal, is characterized by the decrease in observed O_3 exceedance days and average O_3 relative to the Post-NO_x SIP period of 2003-2012. These consistent decreases are attributed to continuing reductions in regional NO_x emissions, particularly from the upwind Ohio River Valley source region, as described in detail in the 2016 O_3 season final report for Delaware.
- The 2017 O_3 season reinforced key trends that have emerged during the New Normal period related to the influence of wildfire smoke, hot weather, persistence O_3, and climatology.
    - Smoke transported from upwind wildland fires continued to be a major factor for local O_3 production in 2017. At least 2 of the 8 O_3 exceedance days (May 17-18) and possibly a third day (July 22) were impacted by smoke. The May 18 event had the highest observed 8-hour average O_3 (84 ppbv at the Lums Pond and MLK monitors) and the largest number of exceeding monitors (all 4 in New Castle County) of the season.
    - Hot weather (maximum observed temperature ≥ 90 °F) is still necessary, but no longer sufficient for O_3 exceedances. Although there was a slight increase in the percent of hot days that corresponded with an O_3 exceedance in 2017 (30%) compared to other years in the New Normal period (12-22%), most hot days ($\geq 70\%$) are no longer O_3 exceedance days.
    - In 2017, persistence O_3 continued to be a weak forecast predictor. Persistence O_3 only explains 24% of the total variation in observed O_3 during the New Normal period, compared to 39% in the Pre-NO_x SIP period (1997-2002) and 33% in the Post-NO_x SIP period.
    - The breakdown in persistence O_3 made it challenging to identify the single day exceedances and geographically isolated exceedances in 2017, such as the July 21 exceedance at Lewes, which was influenced by a sea breeze.
    - Observations in 2017 reinforced the fact that the historical seasonal O_3 cycle has virtually disappeared in the New Normal period; O_3 climatology is now spread out across the entire O_3 summer season.
    - The climatology of the peak of the O_3 season has also shifted, with a drop in the percentage of O_3 exceedances observed in August and a concomitant increase in September.

- Forecast skill for O_3 exceedance days in 2017 was similar to 2016, with a false alarm ratio of 0.43 and hit rate of 0.50. The number of false alarms (3) was constant in 2016 and 2017, but the number of hits in 2017 (4) dropped compared to 2016 (6), due to the greater number of total exceedance days observed in 2016 (11) versus 2017 (8).
- Overall forecast skill (all days) for 2017 improved relative to previous years. Forecasts in 2017 were more accurate (median absolute error of 5.0 ppbv compared to 6.0 ppbv in 2013-2016) with less tendency towards over-forecasting (bias of only 0.7 ppbv in 2017 compared to 2.1 ppbv in 2016).
- The skill of numerical air quality models in predicting O_3 exceedance days in 2017 was mixed and generally inconsistent from 2016. The NOAA and North Carolina models had poor skill; in contrast, the two versions of the Baron Advanced Meteorological Services (BAMS) models had higher hit rates than the expert forecasts. Given the history of the BAMS models' high false alarm ratios in 2016 and their higher overall number of false alarms in 2017, expert forecasters did not recognize what turned out to be skillful guidance until the conclusion of the 2017 season.
- A continuation of the New Normal period is expected in 2018, with 6-12 O_3 exceedance days likely for Delaware. Since forecasters can no longer dependably rely on any of the historical O_3 forecasting tools, including hot weather, climatology, persistence O_3, statistical models, and numerical air quality models, a return to the consistently high hit rates (> 0.60) of the pre-2013 period is unlikely. The general trend of higher hit rates and lower false alarm ratios that has emerged over the past three seasons is expected to continue, however.

1. Ozone Observations in 2017

Summer 2017 marked the second season under the new 8-hour average ozone (O_3) exceedance threshold of **71 ppbv**, set by the 2015 daily O_3 National Ambient Air Quality Standard (NAAQS) of 70 ppbv. Figure 1 compares the 2017 O_3 season (May 1 to September 30) to previous years in the State of Delaware forecast area, which includes Kent, New Castle, and Sussex Counties. In 2017, there were **8 days** with observed 8-hour average $O_3 \geq 71$ ppbv, no days 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 8 Code Orange days (71-85 ppbv) on the Air Quality Index (AQI) scale.

The 8 observed O_3 exceedance days (defined as observed 8-hour average $O_3 \geq 71$ ppbv) in Delaware in 2017 tied the 8 exceedance days observed in 2014 and is slightly less than the 11 exceedance days observed in 2016. The drop in the number of exceedances in 2017 compared to 2016 is likely due to differences in meteorology; summer 2017 (June, July, and August) was about normal in terms of temperature (Figure 2), while summer 2016 was much warmer than average. This difference in temperature is reinforced by Figure 3, which shows the number of hot days, defined as maximum observed temperature (T_{max}) ≥ 90 °F at Dover Air Force Base (DOV), for May 1 to September 30, 1997-2017. There were 20 hot days in 2017, compared to 36 during the unusually warm summer of 2016. In terms of hot days, 2017 was analogous to 2014, which had 19 hot days; this is consistent with the fact that 2017 and 2014 had the same number of exceedance days (8). In addition, summer 2017 was slightly wetter than normal (Figure 4), which also likely helped to suppress O_3 production compared to summer 2016.

Figure 1 also shows that the recent downward trend in the number of O_3 exceedance days in Delaware, which began in 2013, continued in 2017. This period, termed the "New Normal," is attributable to continued reductions in regional NO_x emissions relative to the Post-NO_x SIP period (2003-2012), particularly from the upwind Ohio River Valley source region, as described in detail in the 2016 O_3 season final report for Delaware. The observed O_3 decreases that have occurred during the New Normal period are not unique to Delaware; the same trend is evident across the entire Mid-Atlantic region.

The seasonal (May 1 to September 30) mean and median of the maximum observed 8-hour average O_3 (Figure 5) have the same downward trend as O_3 exceedance days in Delaware. The 2017 seasonal mean and median (49 ppbv) are the lowest observed in Delaware, going back to 1997. Thus, not only have the highest O_3 values dropped during the New Normal period, but the average O_3 values as well.

2. Continuation of Key Trends in 2017

Details about the 8 O_3 exceedance days in Delaware during 2017 are listed in Table 1. Grey shading indicates the 4 days that were accurately forecasted, with alerts issued to the public. The information in Table 1, including the number of exceeding monitors in the forecast area, maximum observed 8-hour average O_3 mixing ratio, and T_{max} at DOV, reinforce key trends that have emerged during the New Normal period of 2013-2017.

2.1 Wildfire Smoke

Smoke transported from wildland fires has become a major factor for local O_3 production in Delaware. Smoke contains high concentrations of NO_x and reactive volatile organic compounds (VOCs), which are precursors for O_3 production. Since most of the Mid-Atlantic region, including Delaware, is in a "NO_x-limited" regime during the summertime, meaning that the formation of O_3 is enhanced by increasing the amount of ambient NO_x, the presence of additional NO_x from wildfire smoke can amplify local O_3 production substantially. Because regional anthropogenic NO_x emissions have been significantly reduced during the New Normal period, the most widespread and highest magnitude O_3 exceedances now occur on days when transported smoke is present.

In 2017, at least 2 of the 8 O_3 exceedance days (May 17-18) and possibly a third day (July 22) were impacted by smoke. The May event (Figure 6), exacerbated by dilute smoke transported from wildfires in Mexico, was particularly widespread across the northern Mid-Atlantic. In Delaware, the highest observed 8-hour average O_3 (84 ppbv at the Lums Pond and MLK monitors) and the largest number of exceeding monitors (all 4 in New Castle County) of the season occurred during this event, on May 18.

2.2 Hot Weather

Observations during 2017 continued to reinforce the trend that hot weather is still necessary, but no longer sufficient for O_3 exceedances during the New Normal period. Figure 7 shows the percent of hot days in Delaware with maximum observed 8-hour average $O_3 \geq 71$ ppbv (orange dots/line) and ≥ 86 ppbv (red dots/line) for 1997-2017. In 2017, there was a slight increase in the percent of hot days that corresponded with an O_3 exceedance (30%) compared to other years in the New Normal period (12-22%). Historically, however, the only other year with < 30% of hot days corresponding to an O_3 exceedance was the Great Recession year of 2009, which was analogous to the New Normal period in terms of low regional NO_x emissions due to the economic recession and unusually cool summer. In the current New Normal period, most hot days ($\geq$ 70%) are not O_3 exceedance days in Delaware, and forecasts of $T_{max} \geq 90$ °F are no longer a reliable predictor of O_3 exceedance events.

2.3 Persistence O_3

In 2017, persistence O_3 continued to be a weak forecast predictor. Historically, persistence O_3 was a useful forecast tool 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 promoted the build-up of O_3, making persistence O_3 an informative indicator of "tomorrow's" O_3 levels. Figure 8 shows the linear relationship between maximum observed 8-hour average O_3 and local 1-day lag (persistence) maximum observed 8-hour average O_3 for Delaware in the New Normal period, including 2017. The coefficient of determination (R^2 value) of the persistence regression equation indicates that persistence O_3 only explains 24% of the total variation in observed O_3 during the New Normal period, compared to 39% in the Pre-NO_x SIP period (1997-2002) and 33% in the Post-NO_x SIP period (Figure 9). This drop illustrates the current limits of persistence as an effective forecast tool.

Isolated single day exceedance "spikes" and exceedances at only 1 monitor location also continued to occur on a regular basis in 2017, underscoring the weakening of persistence O_3 as a forecast predictor. The exceedances on June 22 and September 25 were single-day events, and the July 21 (Lewes) and September 25 (Brandywine) exceedances were only observed by 1 monitor. Single day and isolated exceedances are difficult to forecast, as they are typically geographically isolated and depend on mesoscale weather features that are less accurately predicted by numerical weather models.

The July 21 exceedance at Lewes illustrates the difficulty forecasters face in accurately identifying isolated single day exceedances. This was a "missed" forecast, in that Code Orange O_3 (76 ppbv) was observed, but Code Yellow O_3 (70 ppbv) was forecasted. The July 21 exceedance was isolated; all of the other monitors in Delaware were comfortably inside the Code Yellow range, with observed 8-hour average O_3 of 65 ppbv at Killens and 62 ppbv at Seaford, dropping to the upper 50s ppbv at the New Castle County monitors (Figure 10). Adding to the forecasting challenge, false alarm forecasts were issued for the previous two days, July 19 and 20, meaning that Code Orange O_3 was forecasted but Code Yellow O_3 was observed, despite the fact that conditions were conducive for high O_3 and exceedances were observed at neighboring monitors in New Jersey and Maryland. A sea breeze on July 21 facilitated the isolated exceedance at Lewes, forming a convergence zone inland along the coast that allowed for enhanced O_3 production and build-up. The strength and duration of sea breezes are difficult to forecast, as they are mesoscale weather phenomena, with spatial and temporal dimensions that are less than the scales of most numerical weather models.

2.4 Climatology

Observations in 2017 reinforced the fact that the historical seasonal O_3 cycle has virtually disappeared in the New Normal period. Figure 11 illustrates the historical climatological O_3 cycle, as shown by the daily time series of observed 8-hour average O_3 in Delaware during the Post-NO_x SIP period (blue dots/line), overlaid with the best polynomial fit (black line). The historical seasonal O_3 cycle was characterized by an increase in May, the seasonal peak in June and July, a downward trend in August, and a drop off in September. This seasonal climatology provided a foundation to anchor the daily O_3 forecast, using the 25th percentile, median, 75th percentile, and 95th percentile values as benchmarks. The daily time series of maximum observed 8-hour average O_3 in Delaware during the summer of 2017 (red dots/line in Figure 11) departs from the historical O_3 seasonal cycle, consistent with other years in the New Normal period. The 2017 observations approximate a straight line, with periods of peak O_3 from mid-May to late September, demonstrating that the O_3 climatology is now spread out across the entire O_3 summer season. This change makes O_3 climatology a much weaker foundation for the daily O_3 forecast in the New Normal period compared to previous periods.

The climatology of the peak of O_3 season has also changed (Figure 12). During the Pre-NO_x SIP (blue bars) and Post-NO_x SIP (red bars) periods, O_3 season peaked during June, July, and August, with the highest percentage of O_3 exceedance days occurring in July. There was a smaller but consistent percentage of exceedance days in May, and a marked decrease in September. In the New Normal (green bars) period, however, this historic distribution of O_3 exceedances has shifted at the end of the season. Figure 12 illustrates that the percent of seasonal exceedances in May, June, and July is approximately the same, but the peak in August has dramatically decreased (down to ~7% in the New Normal from ~ 22% in the Pre- and Post-NO_x SIP), while there has been a concomitant surge in the percent of exceedances in September (up to ~20% in the New Normal from ~7% in the Pre- and Post-NO_x SIP). It is too early to know the exact cause of this shift, but it is likely related to come combination of meteorology (perhaps a shift towards more humid conditions in August and hotter conditions in September) and NO_x emissions (possibly due to shifts in energy demand or cap-and-trade program credit distributions).

3. Skill of Ozone Forecasts in 2017

Despite the challenges associated with the continuing breakdowns in historical forecast predictors, the skill of O_3 exceedance day forecasts in 2017 was comparable to 2016. Figure 13 shows the time series of 8-hour average O_3 forecasts and observations for 2017, highlighting the fact that the majority of observed O_3 exceedances (5) in 2017 were threshold exceedances (maximum 8-hour average O_3 of 71-74 ppbv); these days would not have been exceedances under the previous (2008) daily O_3 NAAQS of 75 ppbv. Also prominent in Figure 13 is the smoke-influenced May 18 exceedance day, which had the highest observed 8-hour average O_3 of the entire season (84 ppbv).

Figure 14 shows the false alarm ratio (previously called false alarm rate; red bars) and hit rate (blue bars) for O_3 exceedance day forecasts for 2011-2017. As noted above, the exceedance threshold was 76 ppbv prior to 2016. Details on the calculation of threshold skill scores are given in Appendix A. Forecast skill for O_3 exceedance days in 2017 was similar to 2016, but just slightly less due to the fewer overall exceedance days in 2017 (8) compared to 2016 (11). The false alarm rate in 2017 was 0.43 compared to 0.33 in 2016, while the hit rate in 2017 was 0.50, which falls just short of the 2016 value (0.55). While the number of false alarms (3) was constant in 2017 (Figure 15), the number of hits (4) in 2017 dropped compared to 2016 (6), again related to the greater number of total exceedance days observed in 2016 (more chances to score a forecast hit). Despite the small drop in exceedance day forecast skill in 2017, the fact that the hit rate was larger than the false alarm ratio for the third year in a row reinforces the fact that forecasters have adjusted to the New Normal period and the lower O_3 exceedance threshold.

Overall forecast skill (all days) for 2017 improved relative to previous years, as shown in Figure 16. Forecasts in 2017 were more accurate overall, with a median absolute error of 5.0 ppbv, down 1.0 ppbv from the consistent value of 6.0 ppbv in 2013-2016. Forecasts in 2017 also showed much less tendency toward over-forecasting, with a bias of only 0.7 ppbv compared to 2.1 ppbv in 2016 (unbiased forecasts have 0.0 ppbv bias).

4. Performance of Ozone Numerical Forecast Models in 2017

The skill of numerical air quality models in predicting O_3 exceedance days in 2017 was mixed and generally inconsistent from 2016. Figure 17 shows the false alarm ratios (red bars) and hit rates (blue bars) of the numerical model guidance, including the NOAA-EPA model (NAQFC), the North Carolina Division of Air Quality model (NCDAQ), and two versions of the Baron Advanced

Meteorological Services (BAMS) model, the CMAQ and RT, compared to the expert forecasts for Delaware in 2017.

The NCDAQ model had an extremely high false alarm ratio of 0.86 and a low hit rate of only 0.13, making it shockingly unskillful in 2017. This translated into 6 false alarm predictions and only 1 hit, meaning the NCDAQ model correctly identified only 1 of the 8 observed O_3 exceedance days.

In contrast, the BAMS models were surprisingly skillful in 2017, which was a sharp contrast to 2016, when they had false alarm ratios > 0.50 and hit rates < 0.50. The 2017 false alarm ratios were comparable to the expert forecasts (~0.45) but the hit rates were higher (0.75 for the RT and 0.63 for the CMAQ). Given the history of the BAMS models' high false alarm ratios in 2016, the expert forecasters were unable to recognize what turned out to be skillful guidance in 2017, especially early in the season. Often is it easier to recognize the skill of a numerical model at the end of the season than in an operational setting in the midst of the season. In addition, the BAMS models had a higher overall number of false alarms, especially in July, which made the model seem less reliable. For example, the BAMS models had difficulty handling the challenging period of July 18-22, a Tuesday through Saturday. Each day during this period, the expert forecast had the incorrect AQI color code, with false alarm forecasts on July 19-20 and missed forecasts on July 21-22. As shown in Table 2, the BAMS models consistently over-predicted O_3 during this period, calling for Code Orange exceedance days every day. This resulted in 3 false alarm predictions in a row on July 18-20. As a result, the expert forecasters discounted what turned out to be the correct BAMS exceedance day predictions for July 21-22. Given the seasonal skill of the BAMS models at identifying O_3 exceedance days in 2017, the expert forecasters will give the BAMS guidance more weight at the beginning of the 2018 season, and carefully evaluate its early season performance, given the models' inconsistent season to season skill in recent years.

Similar to 2016, the NAQFC model had poor skill in predicting the O_3 exceedance days in 2017, with a hit rate of only 0.38, comparable to 2016 (0.36; Figure 18). This corresponded to only 3 hits in 2017 (Figure 19). On the positive side, the NAQFC only had 1 false alarm in 2017, for a hit rate of 0.25, which was a substantial improvement over 2016 (0.56). The performance of the NAQFC model for all days (Figure 20) showed a slight improvement in accuracy in 2017, with a median absolute error of 5.0 ppbv compared to 6.0 ppbv in 2014-2016. The bias of the model was relatively consistent in magnitude but changed signs, switching from a small tendency to over-forecast in 2014-2016 (~ +2.0 ppbv) to a slight tendency toward under-forecasting in 2017 (-1.3 ppbv). The NAQFC's performance in 2017 underscores that the 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. Its downward trend in false alarms, if consistent in 2018, may prove useful guidance for avoiding false alarm forecasts, however.

5. Outlook for 2018

A continuation of the New Normal period is expected in 2018, with 6-12 O_3 exceedance days likely for Delaware. Identifying O_3 exceedance days will remain difficult, given the consistent trends that make accurate forecasts challenging, highlighted in Section 2. Forecasters can no longer dependably rely on any of the historical O_3 forecasting tools, including hot weather, climatology, persistence O_3, statistical models, and numerical air quality models. As a result, a return to the consistently high hit rates (> 0.60) of the pre-2013 period is unlikely. The general trend of higher hit rates and lower false alarm ratios that has emerged over the past three seasons is expected to continue, however.

To address some of the uncertainty in transport of smoke from wildland fires, new high resolution aerosol products from the GOES-16 satellite will be available for the 2018 O_3 season. These high accuracy products will be updated every 15 minutes from sunrise to sunset, providing the temporal resolution necessary to track smoky air during mornings and early afternoons, leading up to the mid-afternoon O_3 forecast submission deadline. Smoke aerosols and associated NO_x are not included as boundary conditions in the NAQFC model for air that originates outside the continental United States (CONUS), and much of the thick, transported smoke that impacts air quality in the Mid-Atlantic comes from fires burning in Mexico, Alaska, and Canada. As a result, the NAQFC tends to grossly under-predict on days when smoke impacts local O_3 production. This additional shortcoming of the numerical air quality models in the New Normal period further limits their skill and necessitates the use of alternative method to estimate the growing influence of wildfire smoke. The 2018 O_3 season will be the first opportunity for forecasters to test the preliminary maturity GOES-16 aerosol products and evaluate their usefulness in air quality forecasting.

In addition, NOAA has developed a new post-processed version of its O_3 model guidance using an analog Kalman filter method that is expected to be available during the 2018 season. This post-processed product has the promise of reducing some of the known seasonal drift of the NAQFC guidance, which may in turn help improve the model's skill in identifying O_3 exceedances.

Tables and Figures

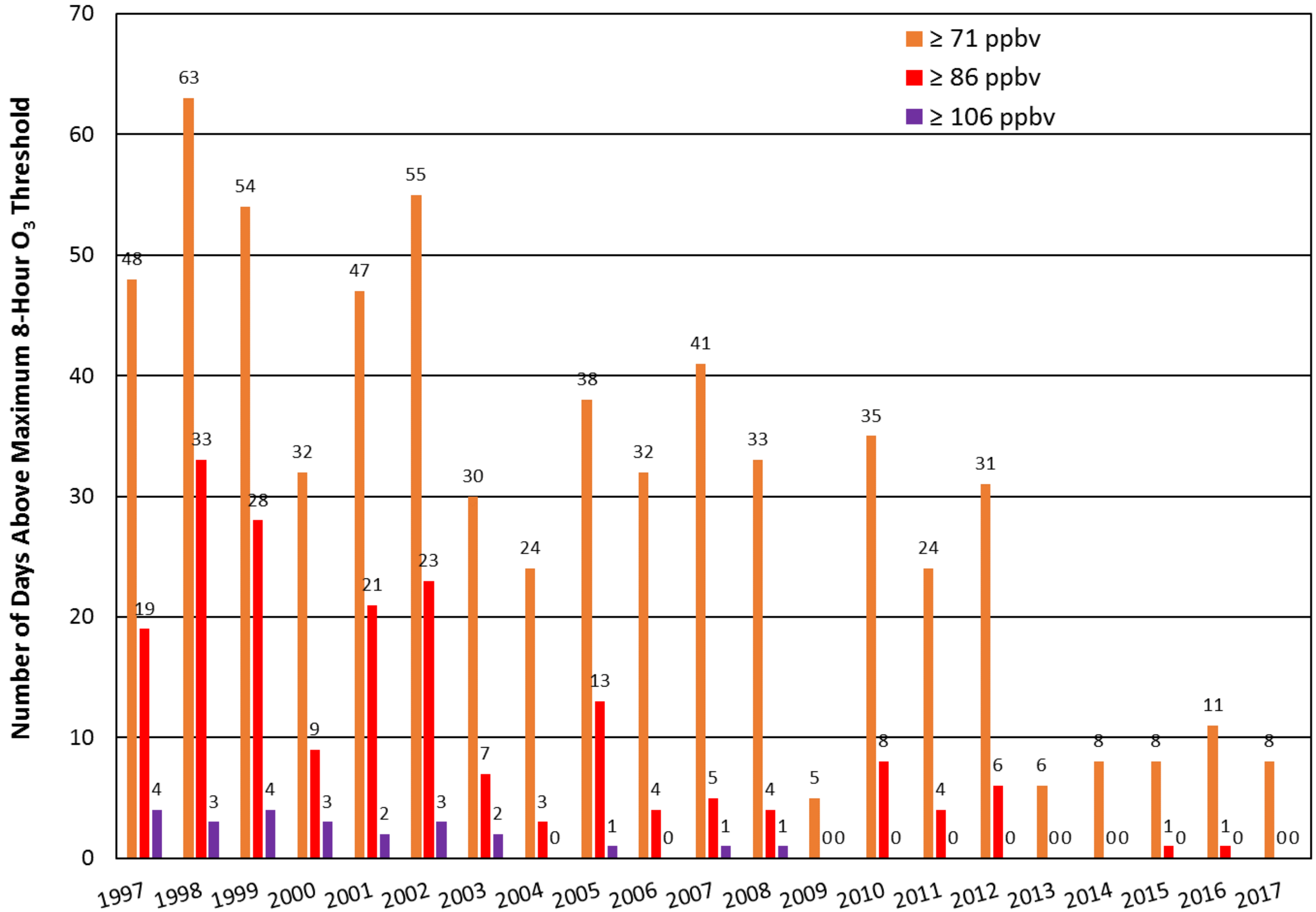

Figure 1. Number of days with maximum observed 8-hour average O_3 at or above thresholds of 71 ppbv (orange bars), 86 ppbv (red bars), and 106 ppbv (purple bars) in Delaware for 1997-2017

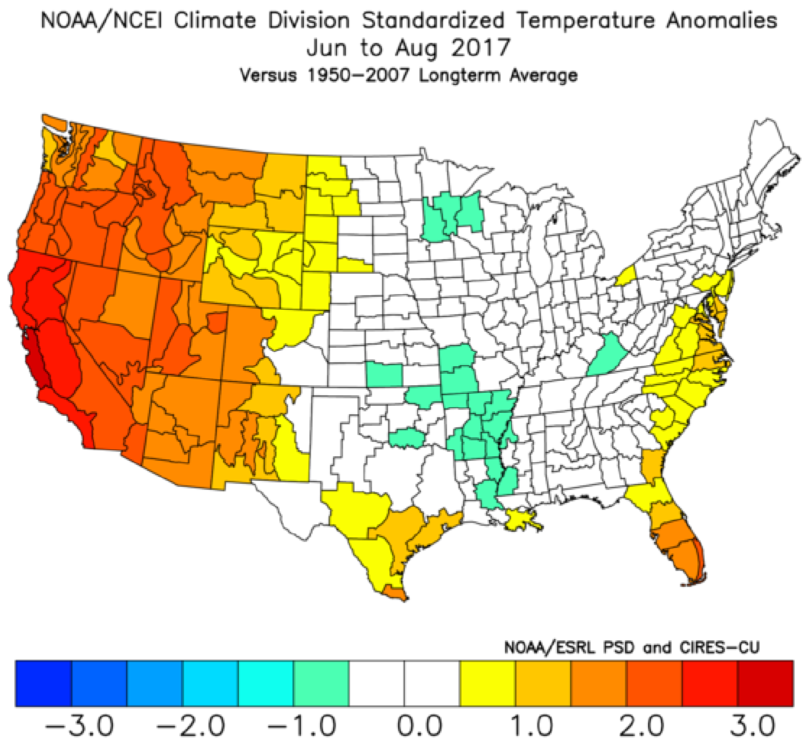

Figure 2. Temperature anomalies (in °F) in the U.S. for June-August 2017 compared to the 1950-2007 average (courtesy of NOAA/ESRL).

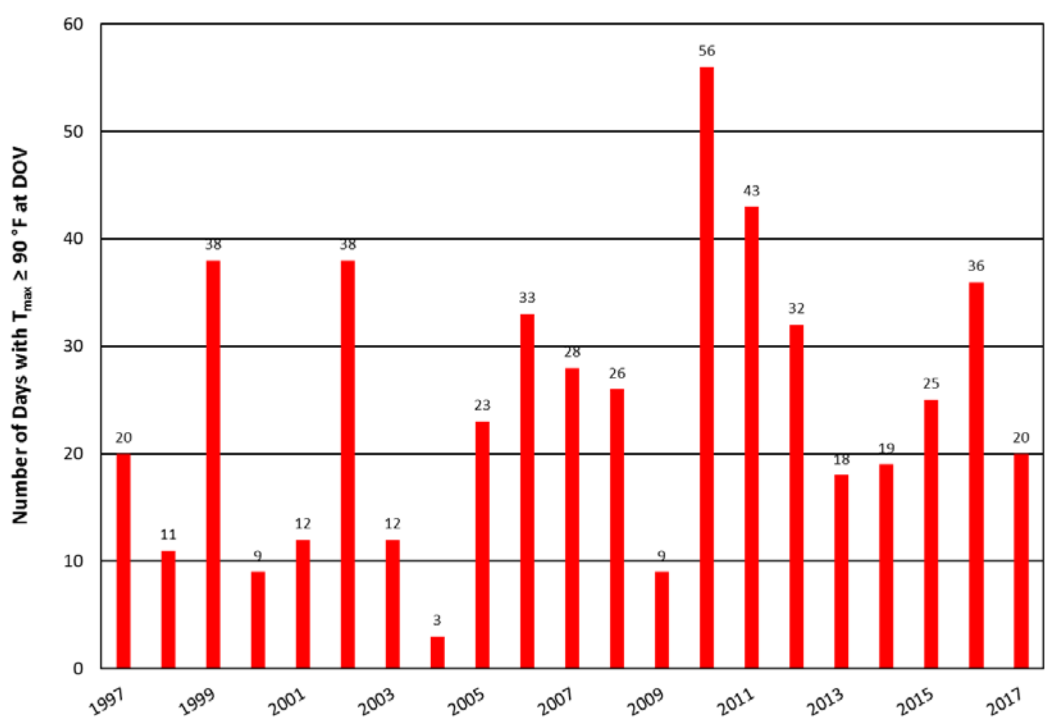

Figure 3. Number of days with maximum temperature (T_{max}) ≥ 90 °F at Dover Air Force Base (DOV) for May 1 to September 30, 1997-2017.

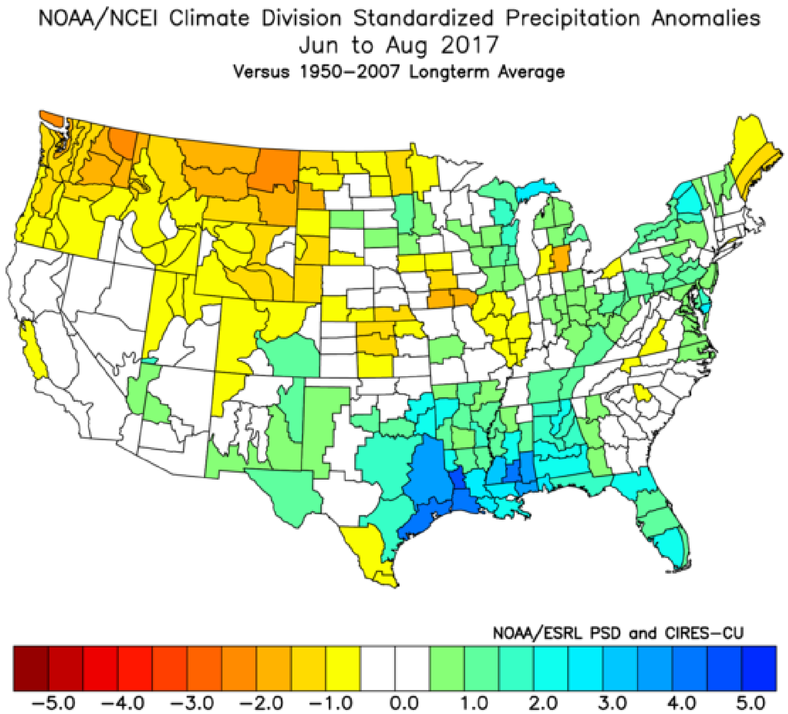

Figure 4. Precipitation anomalies (in inches) in the U.S. for June-August 2017 compared to the 1950-2007 average (courtesy of NOAA/ESRL).

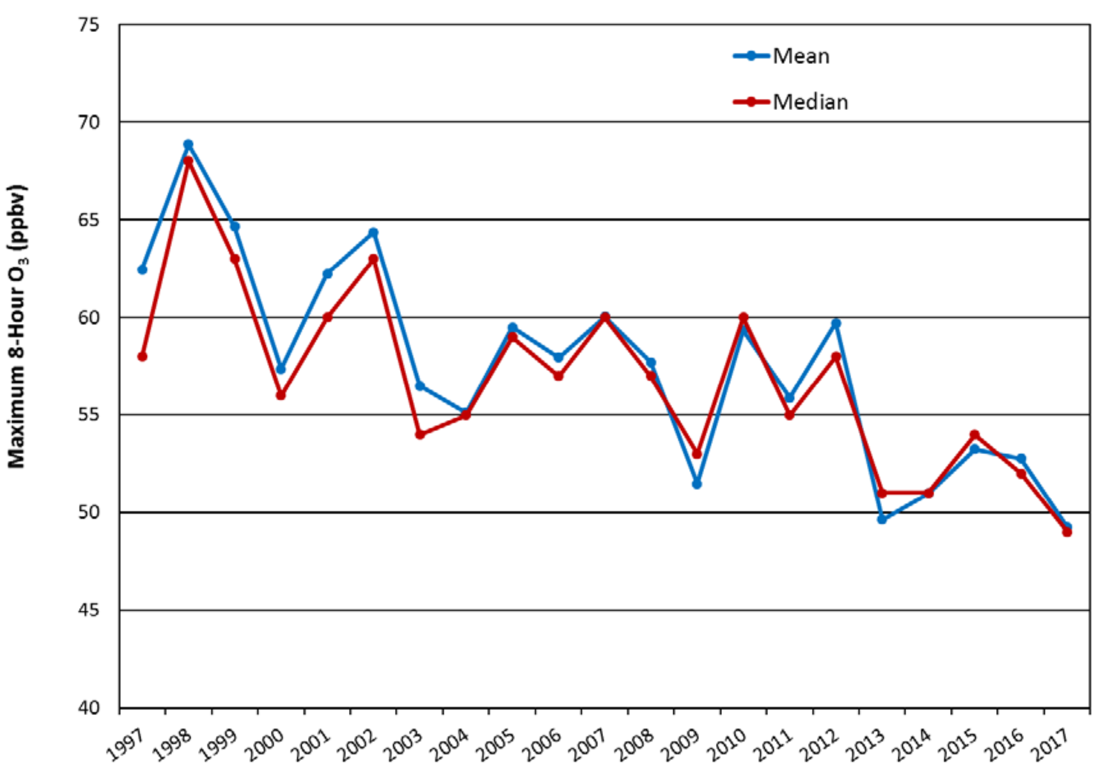

Figure 5. Seasonal (May 1 to September 30) mean and median maximum observed 8-hour average O_3 in Delaware for 1997-2017.

Table 1. Details regarding observed O_3 exceedance days in Delaware in 2017. Grey shading indicates days that O_3 exceedances were correctly forecasted, with alerts issued to the public. T_{max} indicates maximum temperature at Dover Air Force Base (DOV).

Date	Day of Week	Max Temp DOV (°F)	Max 8-Hour Average O_3 (ppbv)	Number of Monitors	Conditions
5/17	Wed	92	74	3	Hot; SW flow; smoky air mass upwind
5/18	Thu	91	84	4	Hot; persistent SW flow; smoky air mass upwind and locally
6/12	Mon	93	72	2	Bermuda High/classic high ozone pattern (hot and sunny)
6/13	Tue	95	74	2	Bermuda High/classic high ozone pattern (hot and sunny)
6/22	Thu	86	71	3	Bermuda High; stagnant; surface convergence
7/21	Fri	94	76	1	Sea breeze at Lewes; light NW surface winds
7/22	Sat	93	73	2	Expected thunderstorms arrived 1 hour too late; possible smoke
9/25	Mon	86	76	1	Strong upper-level ridge; subsidence ahead of Hurricane Maria

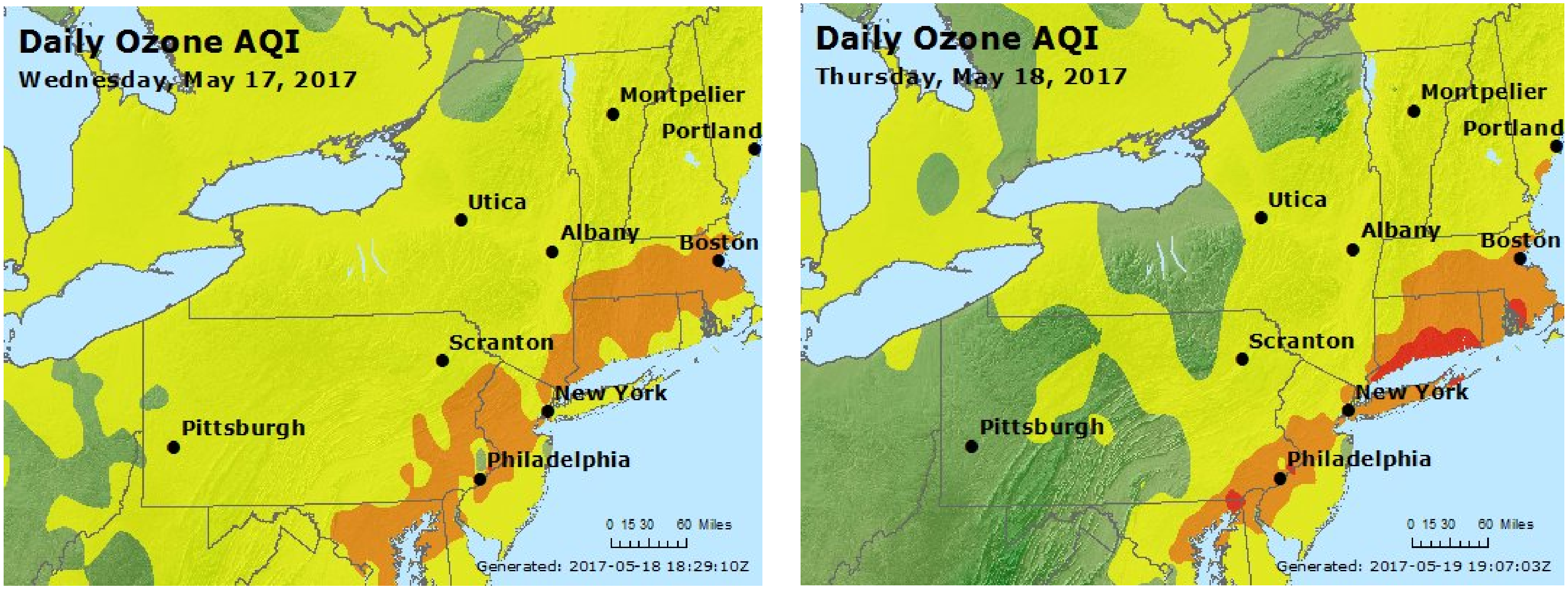

Figure 6. Observed O_3 air quality index (AQI) color codes for the northern Mid-Atlantic and southern New England regions on May 17 (left) and May 18 (right), 2017. Dilute smoke transported from wildfires in Mexico contributed to regional Code Orange O_3 conditions on both days with pockets of Code Red O_3 on May 18.

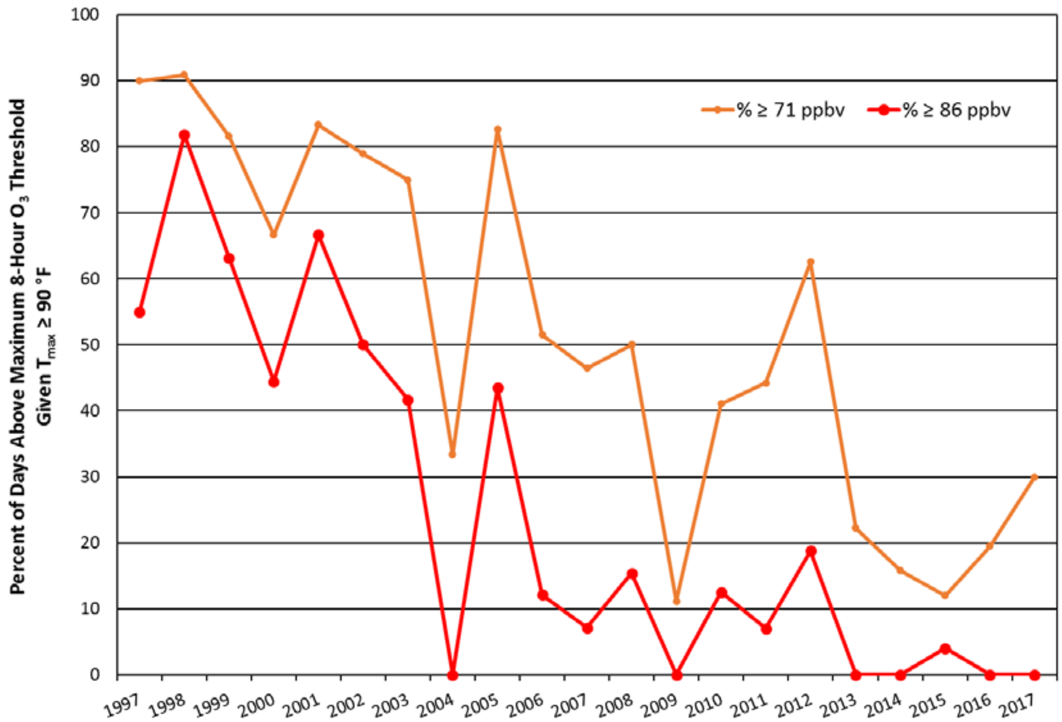

Figure 7. Percent of days with maximum temperature (T_{max}) ≥ 90 °F at DOV and maximum observed 8-hour average O_3 ≥ 71 ppbv (orange dots) or ≥ 86 ppbv (red dots) in Delaware for 1997-2017.

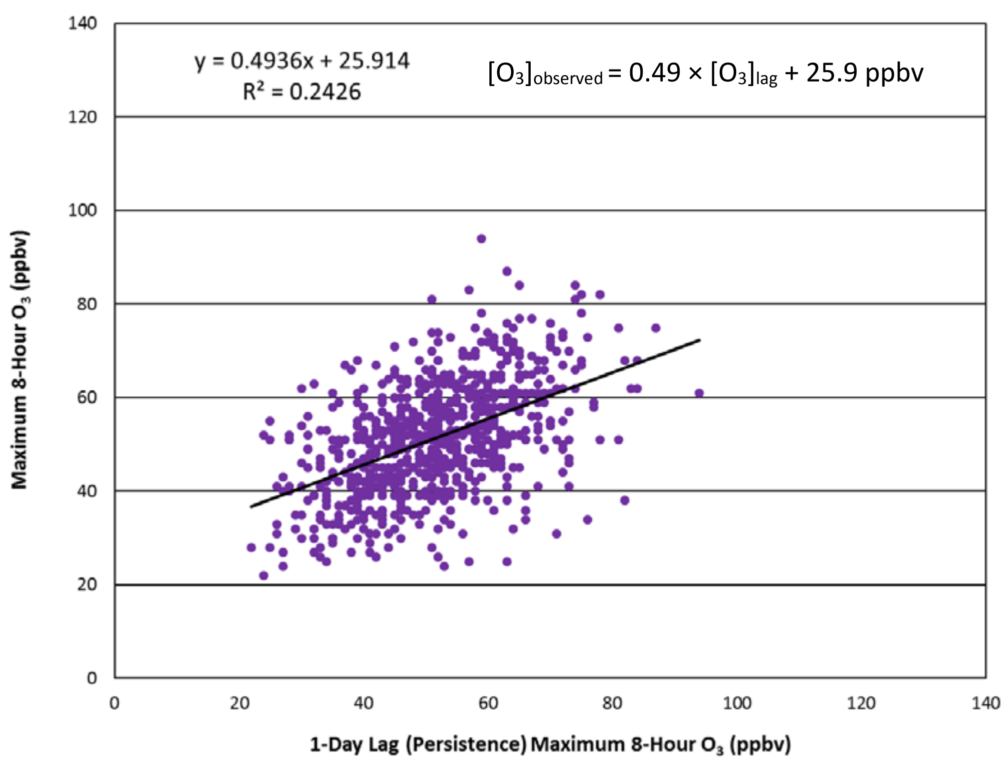

Figure 8. Linear relationship between seasonal (May 1 to September 30) maximum observed 8-hour average O_3 and local 1-day lag (persistence) maximum observed 8-hour average O_3 in Delaware for 2013-2017.

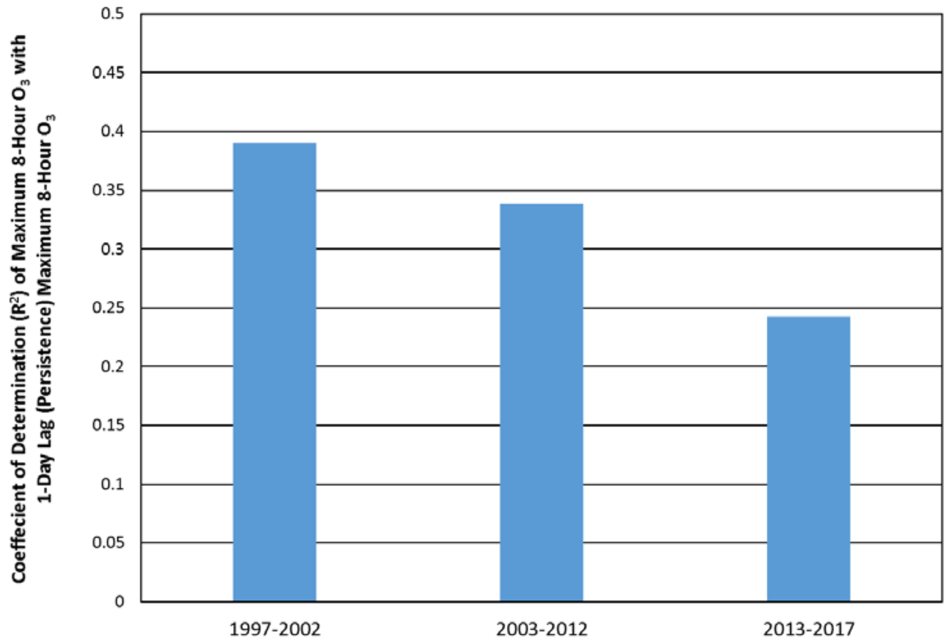

Figure 9. Coefficient of determination (R^2) of linear relationship between seasonal (May 1 to September 30) maximum observed 8-hour average O_3 and local 1-day lag (persistence) maximum observed 8-hour average O_3 in Delaware for given periods.

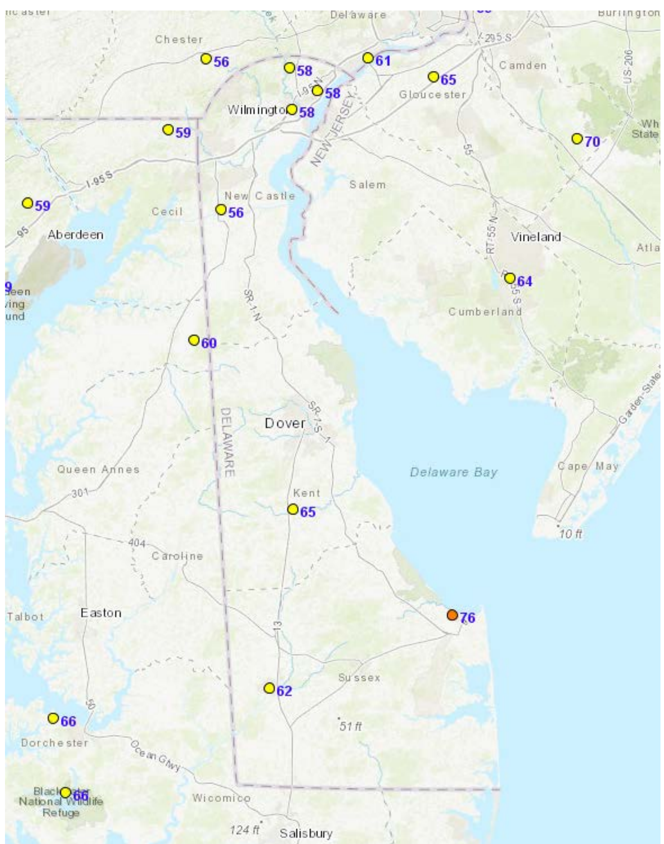

Figure 10. Maximum observed 8-hour average O_3 on July 21, 2017, showing the isolated O_3 exceedance at the Lewes, Delaware monitor (courtesy of AirNow-Tech Navigator).

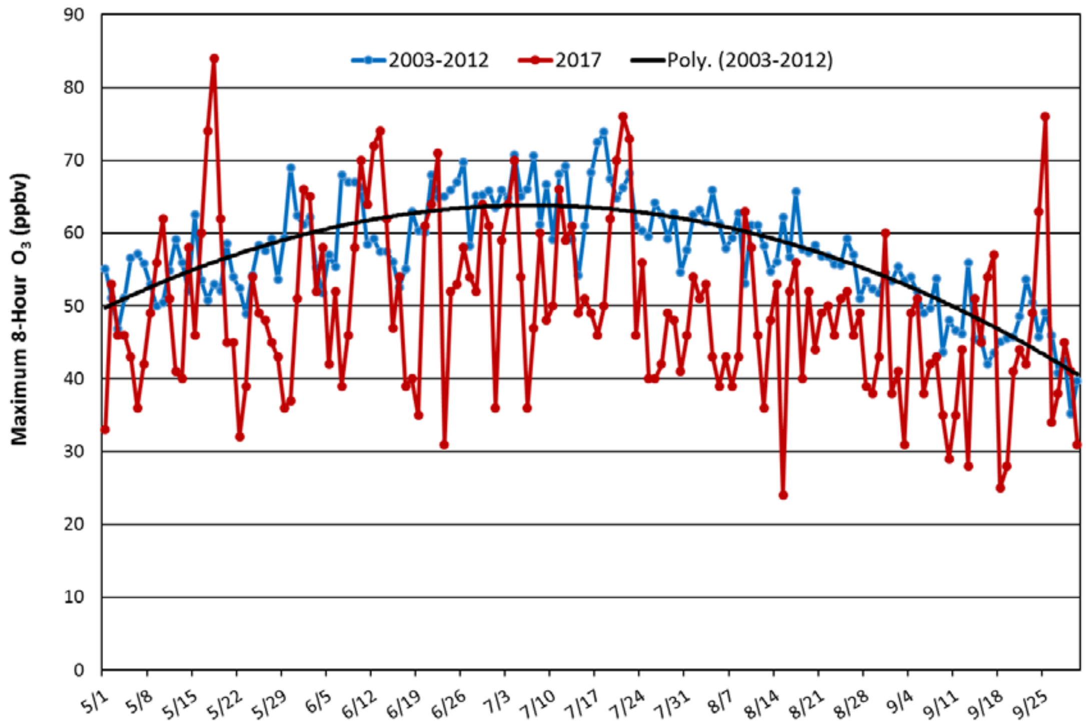

Figure 11. Daily time series of maximum observed 8-hour average O_3 in Delaware for 2017 (red line) compared to the 2003-2012 average (blue line). The black line is the best polynomial fit to the 2003-2012 average.

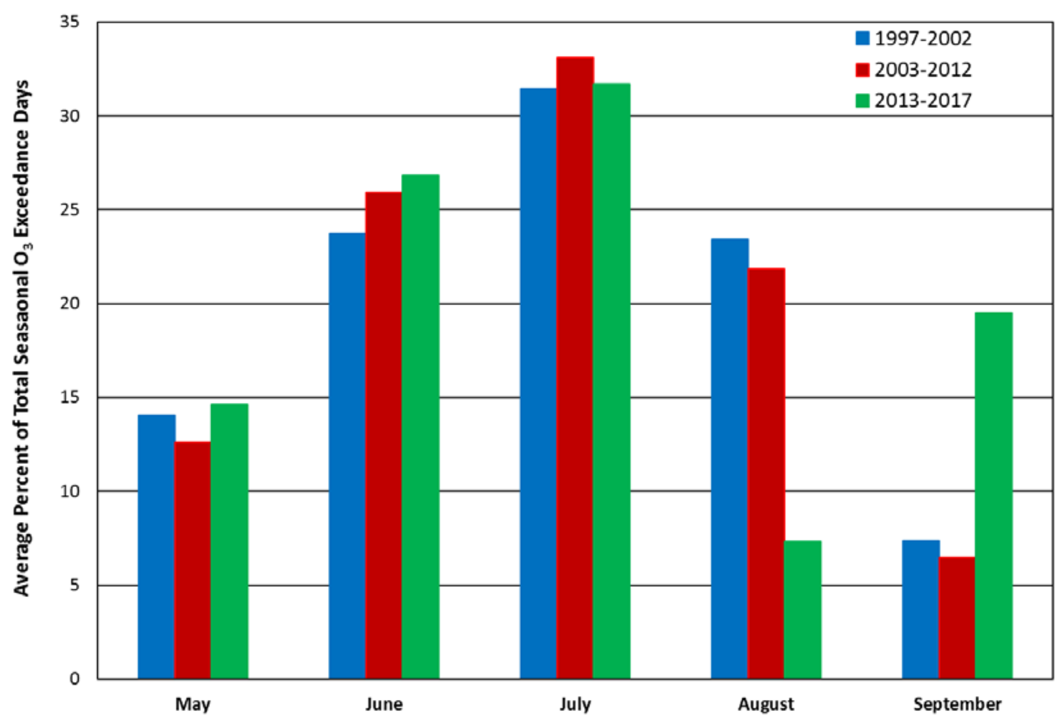

Figure 12. Average percent per month of total seasonal number of days that maximum observed 8-hour average O_3 exceeded the threshold of ≥ 71 ppb in Delaware for the given periods.

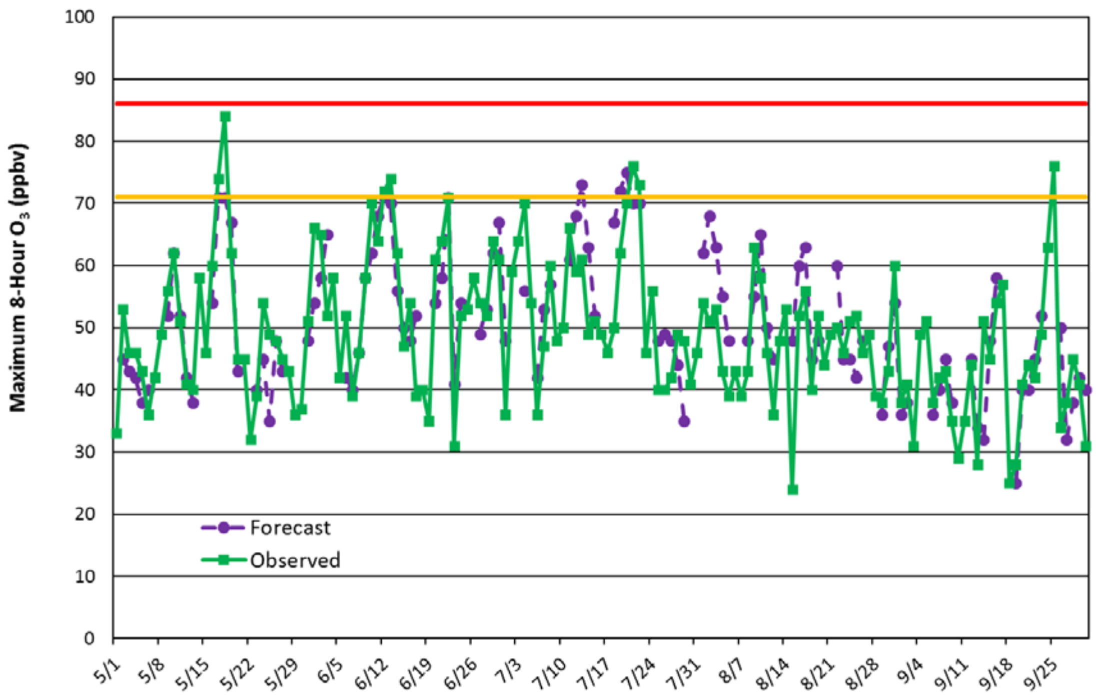

Figure 13. Maximum 8-hour average O_3 forecasts and observations for Delaware during May 1 to September 30, 2017. The orange and red lines indicate the Code Orange and Code Red thresholds, respectively.

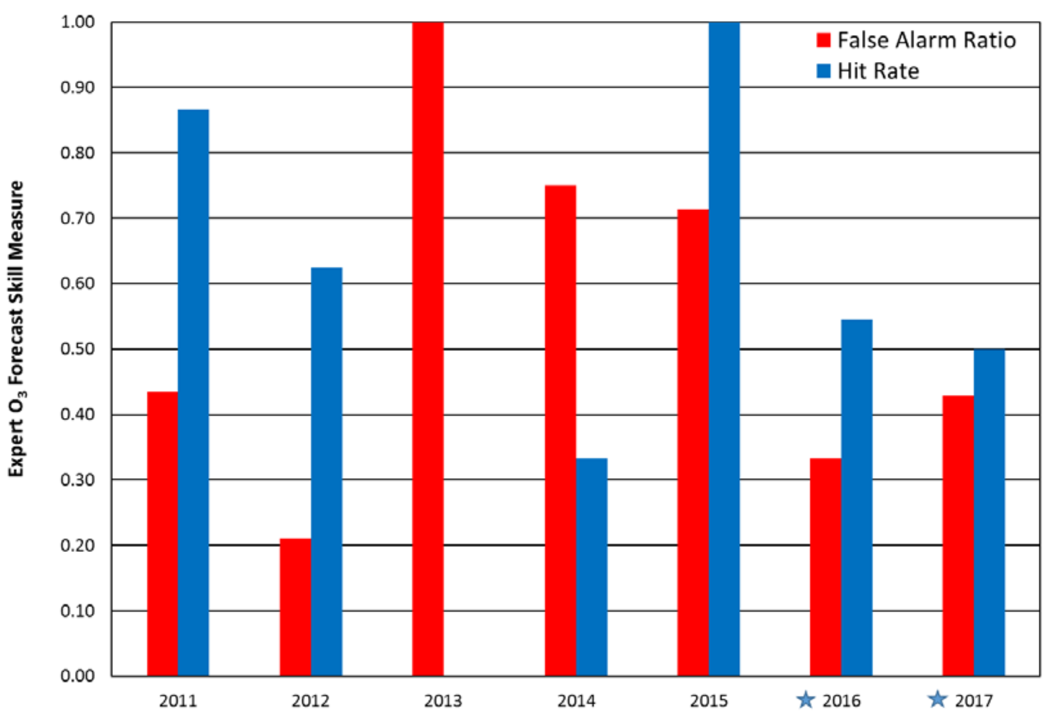

Figure 14. False alarm ratio and hit rate for threshold forecasts of maximum 8-hour average O_3 in Delaware for 2011-2017. Note that for 2011-2015, the threshold was 76 ppbv, but in 2016, it lowered to 71 ppbv (designated by the blue stars).

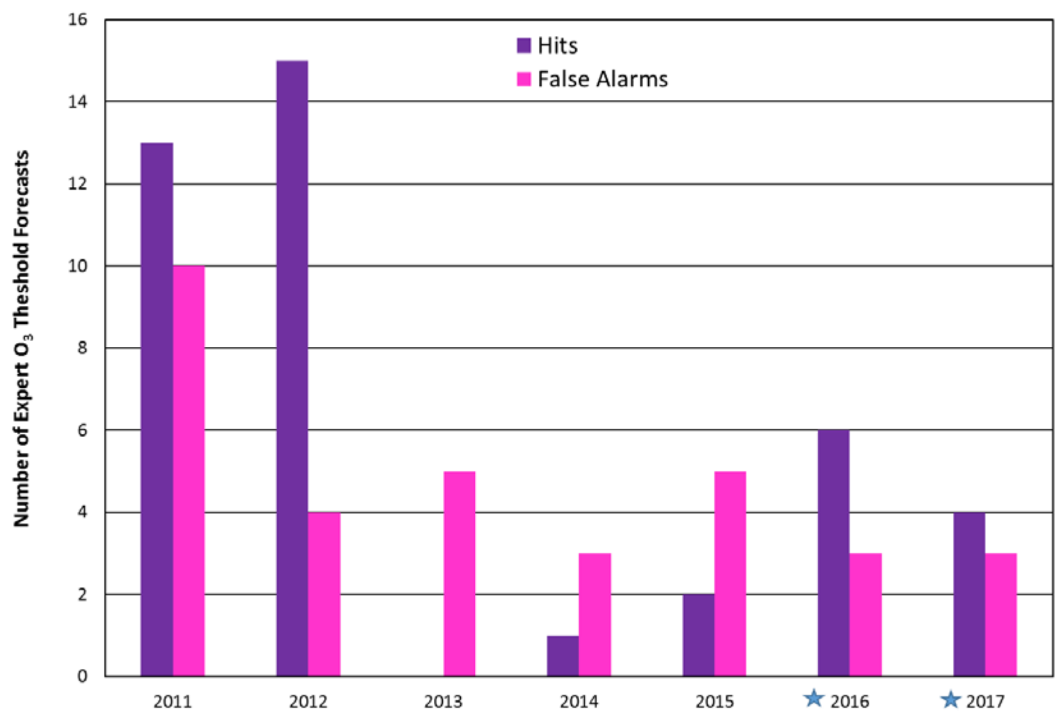

Figure 15. Number of false alarms and hits for threshold forecasts of maximum 8-hour average O_3 in Delaware for 2011-2017. Note that for 2011-2015, the threshold was 76 ppbv, but in 2016, it lowered to 71 ppbv (designated by the blue stars).

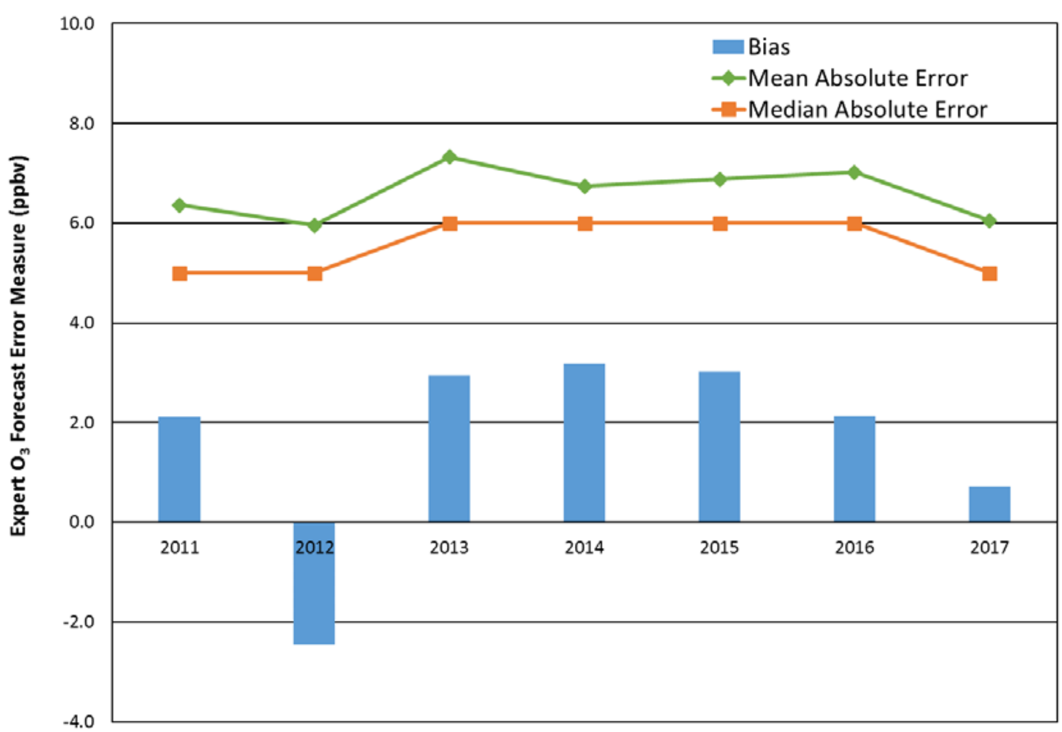

Figure 16. Error statistics for all maximum 8-hour average O_3 forecasts, May 1 to September 30, in Delaware for 2013-2017.

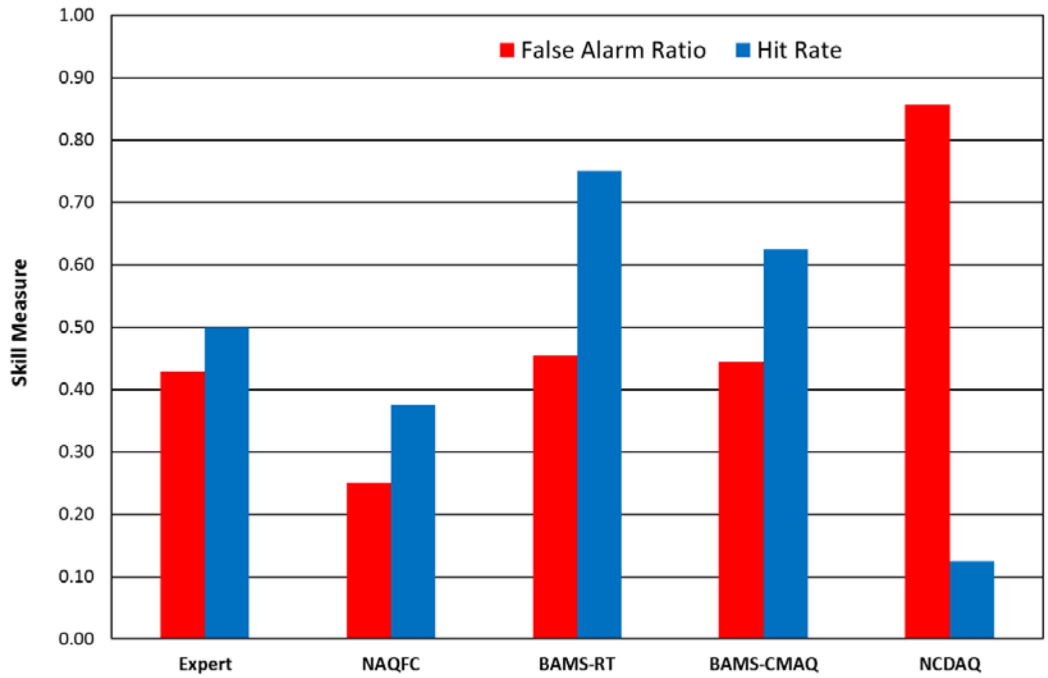

Figure 17. False alarm ratio and hit rate for expert and numerical air quality model threshold predictions of maximum 8-hour average O_3 in Delaware for 2017. Two variations of the Baron Meteorological Services (BAMS) models are shown (CMAQ and RT), as well as the North Carolina Division of Air Quality (NCDAQ) model.

Table 2. Comparison of maximum observed and forecasted 8-hour average O_3 (in ppbv) to guidance from two versions of the Baron Advanced Meteorological Services model (BAMS), the CMAQ and RT, for July 18-22, 2017. Colored shading indicates the corresponding AQI color code.

Date	Observed	Expert Forecast	BAMS-RT	BAMS-CMAQ
7/18	50	67	75	72
7/19	62	72	82	85
7/20	70	75	81	84
7/21	76	70	80	79
7/22	73	70	83	78

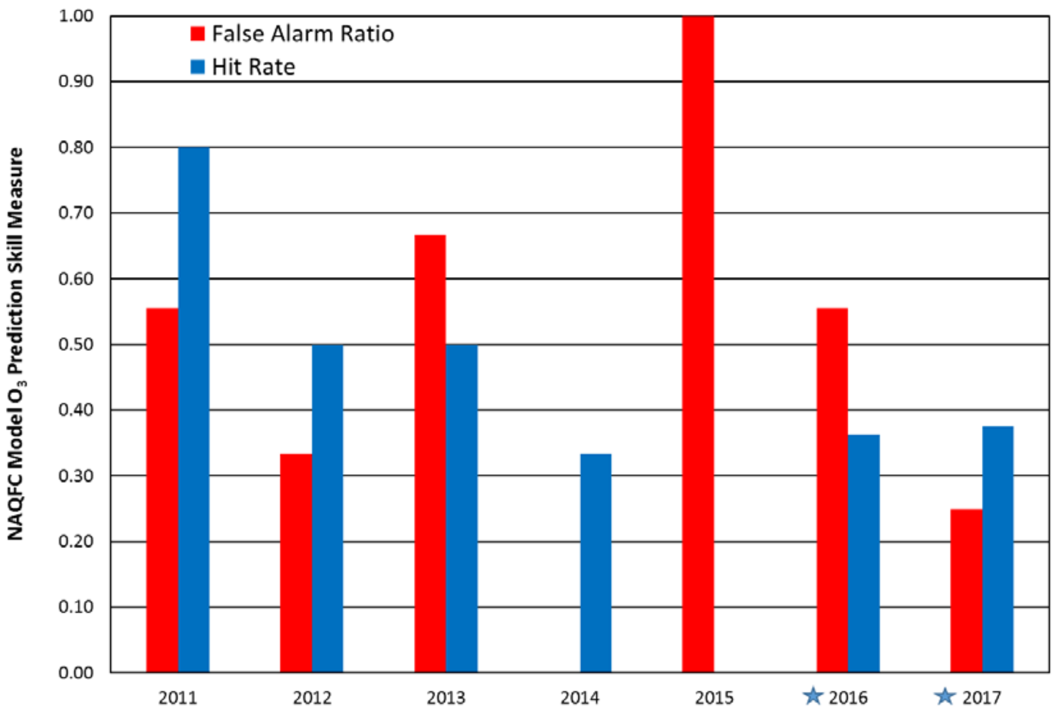

Figure 18. False alarm ratio and hit rate for threshold predictions of maximum 8-hour average O_3 by the NAQFC model in Delaware for 2011-2017. Note that for 2011-2015, the threshold was 76 ppbv, but in 2016, it lowered to 71 ppbv (designated by the blue stars).

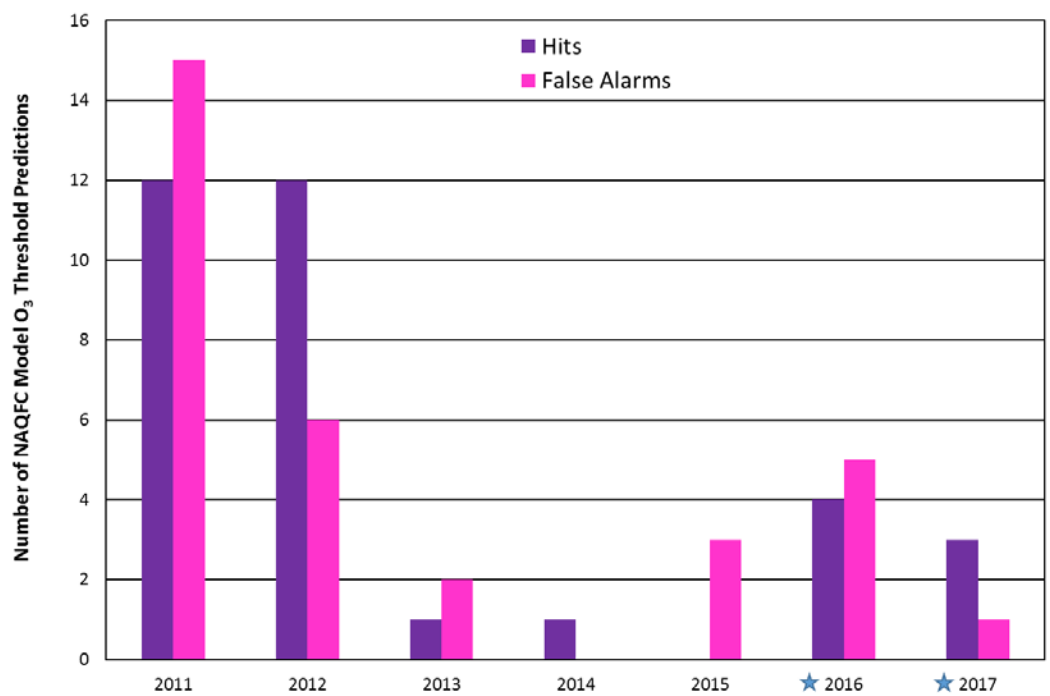

Figure 19. Number of false alarms and hits for threshold predictions of maximum 8-hour average O_3 by the NAQFC model in Delaware for 2011-2017. Note that for 2011-2015, the threshold was 76 ppbv, but in 2016, it lowered to 71 ppbv (designated by the blue stars).

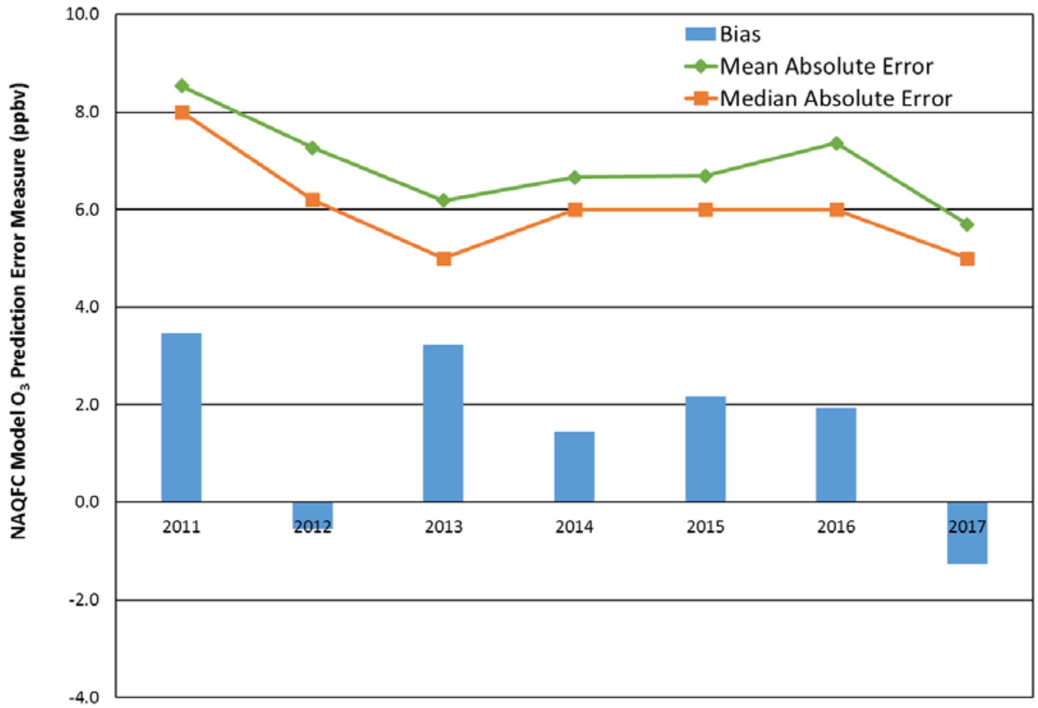

Figure 20. Error statistics for all maximum 8-hour average O_3 predictions by the NAQFC model, May 1 to September 30, in Delaware for 2013-2017.

Appendix A. Skill Measures for Threshold Forecasts

The determination of the skill of threshold forecasts (e.g., O_3 exceedances) begins with the creation of a 2x2 contingency table of the form of Figure 21. 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."

		Observed	
		Yes	No
Forecasted	Yes	a	b
	No	c	d

Figure 21. 2 x 2 contingency table for threshold forecasts.

Basic Skill Measures

Two common scalar skill measures are calculated using the values from the 2x2 contingency table: false alarm ratio, formerly termed false alarm rate, and hit rate.

False alarm ratio (FAR) is the fraction of threshold forecasts which are made but do not verify.

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

Hit rate (H), also called probability of detection (POD), is the ratio of correct threshold forecasts to the total number of O_3 exceedances that were observed.

$$H = \frac{a}{a+c} \qquad (2)$$

Reference

Wilks, D. S., *Statistical Methods in the Atmospheric Sciences*, Third Edition, Academic Press, 676 pp., 2011.