State of Delaware
Final Report: Ozone Observations and Forecasts in 2015

A Report Prepared for the Delaware Department of Natural Resources and Environment

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

February 29, 2016

Table of Contents

List of Figures

Executive Summary

- The O_3 season of 2015 was the third consecutive historically low O_3 season for the State of Delaware. Only 2 days in 2015 reached the Code Orange threshold, and there were no Code Red or Code Purple days.
- Data suggest that a "step-down" in observed O_3 began in 2013, most likely due to continuing reductions in NO_x emissions.
- The summer seasons of 2013-2015 were marked by an atypical weather pattern that limited the development of heat waves in June-August and the inhibited the frequency of westerly transport aloft; this pattern likely contributed to the recent "step-down" in observed O_3 levels, but is not primarily responsible.
- Overall forecast skill (all days) was comparable to 2013-2014, with a median absolute error of 6.0 ppbv.
- Forecast skill for the Code Orange O_3 cases showed steady improvement compared to 2013-2014. Notably, both of the Code Orange O_3 days in 2015 were correctly forecasted. The false alarm rate was high (0.71) but an improvement on 2014 (0.75) and much better than 2013 (1.00).
- For the recent "step-down" period, high temperature and persistence have become unreliable as forecast predictors, increasing the difficulty of accurately forecasting Code Orange cases.
- As in 2013-2014, the numerical air quality models that provide forecast guidance had poor skill for the Code Orange O_3 cases and a propensity to over-predict peak O_3 in all cases. The North Carolina Department of Environment and Natural Resources (NCDENR) model had the most skillful performance overall of the four forecast models used in the model "ensemble."
- A new Code Orange threshold of 71 ppb will be in place for 2016, which is expected to roughly triple the number of O_3 exceedance days in Delaware relative to the 2013-2015 average. Under the new O_3 standard, 2015 would have had 5 Code Orange days and 1 Code Red day.

Ozone Observations in 2015

The ozone (O_3) season of 2015 was the third consecutive historically low O_3 season for Delaware, in terms of both mean and median O_3 values and the number of days that observed 8-hour average O_3 exceeded the National Ambient Air Quality Standard (NAAQS) of 75 ppb. In 2015, there were only 2 days with maximum observed 8-hour O_3 in excess of the Code Orange threshold of 76 ppbv, no days in exceedance of the Code Red threshold of 96 ppbv, and no days in exceedance of the Code Purple threshold of 116 ppbv (Figure 1). Thus, 2015 was one of two years with the lowest occurrence of O_3 NAAQS exceedance days since 2000, tied with 2013.

Although three years are not sufficient to definitively confirm a trend, the data in Figure 1 suggest that beginning in 2013, a "step-down" in observed O_3 began. The black lines in Figure 1 demonstrate that an initial "step-down" occurred during the period 2003-2012, when an average of 17.7 O_3 exceedance days were observed each year compared to an average of 38.3 exceedance days during the period 1990-2002. This initial decrease in observed O_3 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. during the period 1999-2002 associated with the so-called "NO_x SIP Rule." Another instance of a "step-down" in O_3 precursor emissions and concentrations occurred during the Great Recession in 2009, when only three O_3 exceedance days were observed. Figure 1 indicates that, likely due to continuing NO_x emissions reductions following the Great Recession, an additional "step-down" in O_3 has occurred, such that the average number of O_3 exceedance days dropped substantially to only 2.3 per year for the period 2013-2015. This trend is not unique to Delaware; the same recent "step-down" in observed O_3 is evident in the Philadelphia metropolitan area.

The "step-down" phenomenon is further illustrated by Figures 2-4, pie charts showing changes in the percentage of all observed Air Quality Index (AQI) color codes since 1990. The percentage of Code Yellow O_3 has not changed substantially, from 28% in 1990-2002 to 30% in 2003-2012 and more recently, 22% in 2013-2015. But during those same periods, Code Orange days have decreased from 19% in 1990-2002 to 11% in 2003-2012 to only 2% in 2013-2015. At the same time, Code Red and Code Purple days have essentially disappeared, and the percentage of Code Green days has increased dramatically to the point where 76% of days during the period 2013-2015 were Code Green.

As shown in Figure 5, the 2015 seasonal (May 1-September 30) mean and median maximum observed 8-hour O_3 were the fourth-lowest since 1990; only the values from 2013-2014 and the recession year of 2009 were lower. The daily time series of maximum observed 8-hour O_3 in Delaware during the summer of 2015 compared to the 2003-2012 average (Figure 6) demonstrates that daily maximum observed 8-hour O_3 in 2015 fell below the 2003-2012 average on most days. 2015 data are compared to the 2003-2012 average to reflect the change from the prior "step-down" period due to the initial impacts of the NO_x SIP Rule emissions reductions. This difference is further illustrated in Figure 7, the daily time series of the difference between maximum observed 8-hour O_3 in 2015 compared to the 2003-2012 average. The values of vast majority of the green lines in Figure 7 are negative, indicating that observed 8-hour O_3 in 2015 was less than the 2003-2012 average on most days.

The Impact of Meteorology in 2015

In 2015, the association between hot weather, characterized by maximum temperature (T_{max}) $\geq$ 90 °F, and high O_3 (Code Orange or higher days) continued to deteriorate. The historical strong association of hot weather with high O_3 weakened in Delaware in the wake of the NO_x SIP Rule, and the most recent "step-down" period of 2013-2015 has seen an even sharper breakdown in the hot weather/high O_3 relationship, as shown in Figure 8. Prior to 2003, over 80% of days with $T_{max} \geq$ 95 °F and about 70% of days with $T_{max} \geq$ 90 °F were high O_3 days. These percentages fell to around 55% of days with $T_{max} \geq$ 95 °F and around 30% of days with $T_{max} \geq$ 90 °F for the period 2003-2012. For 2013-2015, however, just over 10% of days with $T_{max} \geq$ 95 °F and only around 5% of days $T_{max} \geq$ 90 °F were high O_3 days.

In the current O_3 era of 2013-2015, hot weather is still necessary for high O_3, but it is clearly no longer sufficient. Now, most hot days are <u>not</u> high O_3 days. The challenge is identifying which hot days will be high O_3 days. In the wake of the "decoupling" of O_3 and temperature, other conditions, such as local scale wind circulation, convection, cloud cover, and local emissions, are becoming greater contributing factors for observed O_3 concentrations. The "classic" conceptual model for O_3-conducive weather patterns, driven by synoptic scale weather and transport effects, is no longer consistently applicable.

Figure 9 shows that temperatures during June, July, and August 2015 were about average compared to 1981-2010, which would suggest an average O_3 season. Certainly compared to the 2003-2012 average, 2015 was a historically low O_3 season, but compared to the 2013-2014 average, it was normal.

Figure 10 indicates that Delaware experienced slightly above average precipitation during June, July, and August, which may have partially contributed to the historically low O_3 levels. As was the case in 2013-2014, another contributing factor in 2015 was the overall synoptic weather pattern, which featured few extended periods of hot, dry, and stagnant weather in June, July, and August. During summers with average or above average O_3 concentrations, the semi-permanent Bermuda High extends westward over the Mid-Atlantic region. During June-August 2015, however, the Bermuda High was suppressed southward (Figure 11), which helped to prevent the development of extended heat waves across the Mid-Atlantic.

In contrast to the June-August period, September 2015 featured classic synoptic heat waves in the first and third weeks of the month, with the Bermuda High in its preferred position for high O_3 (Figure 12). In response, one of the two observed Code Orange O_3 days in Delaware occurred on September 17, in the middle of the second heat wave.

As in 2013-2014, another factor in limiting O_3 during 2015 appeared to be the predominant transport pattern aloft. During summer 2015, the transport pattern differed from the classic westerly transport pattern, in which air aloft at approximately 500-1500 m above ground level (AGL) flows from the Ohio River Valley, a source region for O_3 and O_3 precursors, such as NO_x. The summer 2015 transport pattern featured light onshore flow, which brought clean maritime air into Delaware and the northern Mid-Atlantic region, as shown in Figure 13. Winds in the transport layer aloft were suppressed and shifted southerly compared to normal, with the result that large scale westerly transport from the Ohio River Valley and the Midwest into Delaware was limited.

Thus, 2015, like 2013-2014, was marked by an atypical weather pattern that limited the development of heat waves in June-August and the inhibited the frequency of westerly transport aloft. This weather pattern likely contributed to the recent "step-down" in observed O_3 levels, but it appears to have played a minor role compared to ongoing reductions on O_3 precursor emissions.

Code Orange O_3 Days in 2015

The two observed Code Orange O_3 days in 2015 were single-day events that occurred on June 11 and September 17. Table 1 lists characteristics of the O_3 exceedance days. Of note is that both of the Code Orange days occurred in single day "spikes," although these days were part of regional high O_3 events along the I-95 Corridor. As shown in Figure 14, prior to 2013, roughly 50% or more of observed high O_3 days occurred in multi-day events (two or more days in a row), which made persistence, or previous day maximum 8-hour O_3, a reliable forecast tool. The exception is the recession year of 2009, when none of the three Code Orange days occurred in multi-day events. Similarly, Figure 14 shows that no high O_3 days occurred as part of multi-day events for the recent "step-down" period of 2013-2015 – they were all single day "spikes," which are more challenging to predict.

The first Code Orange day in Delaware, June 11, was attributable to wildfire smoke. 2015 was remarkable for the influence of wildfire smoke transported from fire events in the western U.S., Alaska, and Canada. Smoke can cause rapid O_3 formation because it contains high concentrations of NO_x and reactive hydrocarbons, which are both precursors for O_3 production. The maximum observed 8-hour O_3 mixing ratio was 94 ppbv in northern Delaware on June 11, just under the Code Red threshold. This was the highest observed O_3 value for the entire season. Figure 15 shows that on June 11, observed O_3 reached the Code Orange range along the I-95 Corridor from Washington, DC to Providence, Rhode Island, with an isolated Code Red O_3 observation in northeastern Maryland. On this day, smoke transported from wildfires in Alaska and western Canada mixed to the surface and was responsible for the widespread O_3 exceedances along the I-95 Corridor.

June 11 also was the first day of a 3-day heat wave; T_{max} was ≥ 90°F at Dover, Delaware on June 11-13. As discussed above, hot weather is historically associated with Code Orange O_3, but during this mid-June heat wave, the only Code Orange O_3 day in Delaware was June 11, presumably due to the influence of smoke. By June 12, the smoke had been dissipated by southerly flow aloft, such that observed 8-hour O_3 only reached 61 ppbv in northern Delaware, barely into the Code Yellow range, despite the persistent hot conditions (T_{max} = 92 °F at Dover). Observed O_3 on June 13 was also only 61 ppbv, even though it was the third day of the heat wave (T_{max} = 92 °F at Dover).

In contrast, the September 17 Code Orange day was a classic synoptic-based event. A strong upper-level "ridge" of high pressure extended over the Mid-Atlantic region, which promoted very warm weather, sunny skies, and stagnation. It was atypical in that it occurred during late September, when fewer hours of sunlight and a larger solar zenith angle usually limit O_3 formation sufficiently to prevent Code Orange events. The last time that Code Orange O_3 was observed in September in Delaware was in 2010. The event did not coincide with a true heat wave, as T_{max} only reached 85 °F at Dover on September 17, but it was sufficiently hot for O_3 production.

Skill of Ozone Forecasts in 2015

The skill of all O_3 forecasts in 2015 was comparable to recent years. A time series of forecasts and observations for 2015 is shown in Figure 16, and error statistics for forecasts during recent years (2013-2015) are given in Figure 17. Median absolute forecast error during the 2015 season was 6.0 ppbv, which is the same as 2014 and slightly lower than 2013 (6.5 ppbv). Figure 17 also shows that, similar to recent years, there was a bias toward over-prediction of O_3 in 2015 (3.0 ppbv).

As was the case in 2013-2014, the 2015 forecasts were below historical average in skill for the Code Orange cases (Figure 18), but they showed steady improvement compared to 2013 and 2014. Details on the calculation of skill scores are given in Appendix A. The false alarm rate in 2015 was 0.71, indicating that for five of the seven forecasted Code Orange days, Code Orange O_3 was not observed. Although the 2015 false alarm rate was high, it was an improvement on 2014 (0.75) and much better than 2013 (1.00). With the exception of the September 3 false alarm, which was a near miss (observed O_3 of 70 ppbv), the other four false alarm days were substantial misses of 8 ppbv or more. Single day “spikes” in O_3 are difficult to forecast in part because persistence is an important predictor. Single day Code Orange “spikes” have been even more difficult to forecast in recent years as the relationship between hot weather and O_3 has become more tenuous, as noted above.

The recent downward trend in the total number of false alarms leveled off in 2015, with a total of 5, compared to 3 in 2014 and 4 in 2013, versus 6 in 2012 and 10 in 2011. Both of the 2015 observed Code Orange days were correctly forecasted with health alerts issued to the public, for a hit rate of 0.50. This is a steady improvement on the hit rate of 0.00 in 2013 and 0.33 in 2014. The gradual improvement in false alarm rate and hit rate in 2015, the third year of the current “step-down” period, suggests that forecasters are adjusting to the new observed O_3 environment and that a return to the historical high skill for Code Orange forecasts is eminent for 2016-2017.

Performance of Ozone Numerical Forecast Models in 2015

Numerical air quality forecast models are increasingly able to provide reliable guidance for operational O_3 forecasts. These models become operational in the mid-2000s and have steadily improved in skill since that time. Beginning in the 2011-2012 O_3 seasons, a number of numerical air quality forecast models were routinely available to forecasters. Similar to 2013 and 2014, the O_3 model “ensemble” for 2015 included the NOAA-EPA model (NOAA), the North Carolina Department of Environment and Natural Resources model (NCDENR), and two versions of the Baron Advanced Meteorological Services model (BAMS), the CMAQ and RT.

Error statistics for the O_3 forecast models used as guidance in 2015 are shown in Figure 19, and skill scores for the Code Orange O_3 predictions are given in Figure 20. As has been the case in recent years, all of the forecast models suffered from an over-prediction bias, ranging from 2.2 ppbv (NOAA-EPA) to 3.1 ppbv (BAMS-CMAQ). The median absolute error of the models varied from 6.7 ppbv (NOAA) to 8.0 ppbv (BAMS-RT). All of the models had high false alarm rates, with NOAA and BAMS-RT having the highest at 1.00, meaning that they did not correctly predict either of the observed Code Orange days. Only the NCDENR model was able to accurately forecast both of the Code Orange days, while the BAMS-CMAQ correctly forecasted one (June 11).

Although the NOAA model had the greatest skill for all days, with the least bias and lowest median absolute error (analogous to the expert forecasts), it had very poor skill for the high O_3

days. For 2015, the NCDENR model was the most skillful overall, with comparable skill to the NOAA model for all days and a high hit rate for Code Orange forecasts. The model ensemble average forecasts also performed skillfully, with a lower false alarm rate compared to any individual model, but still slightly higher than the expert forecasts. These results are different from 2014, when the NOAA model had the most skill for high O_3 days, which underscores the need for continuous verification of the model guidance during the course of the summer season, so variations in model performance can be determined and the most skillful model can be weighted more heavily as operational guidance. This approach will continue for the 2016 summer season.

Outlook for 2016

U.S. EPA lowered the O_3 NAAQS from 75 to 70 ppb on October 26, 2015. Thus, for summer 2016, a new Code Orange threshold of 71 ppb will be in place. This change is expected to roughly triple the number of O_3 exceedance days in Delaware, relative to 2013-2015. For example, in 2015, there were 6 days with observed 8-hour $O_3 \geq 71$ ppb, compared to 2 days with observed 8-hour $O_3 \geq 76$ ppb. Similarly, in 2014, there were 8 days with observed 8-hour $O_3 \geq 71$ ppb, compared to 3 days with observed 8-hour $O_3 \geq 76$ ppb. Under the new O_3 standard, 2015 would have had 5 Code Orange days and 1 Code Red day.

Development of a new statistical model, trained on the period from 2004-2013, had been planned for Delaware. 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, however, likely because of the recent "step-down" in observed O_3 that began in 2013. As a result, a new statistical model for Delaware will need to be trained on the period 2013-2015 to reflect the recent changes in emissions of O_3 precursors. This model will be developed and tested during 2016, with the hope that it will help make Code Orange and higher O_3 days easier to identify.

Summary

The 2015 O_3 season was the third consecutive historically low O_3 season for Delaware. There were only two O_3 NAAQS exceedance days in 2015, tying 2013 for the lowest occurrence of O_3 exceedance days since 2000. Although three years are not sufficient to definitively confirm a trend, data suggest that a "step-down" in observed O_3 began in 2013. The previous "step-down" period, 2003-2012, is attributed to reductions in NO_x emissions associated with the NO_x SIP Rule. A brief interim "step-down" in O_3 also occurred during the Great Recession in 2009. Now, most likely due to continuing NO_x emissions reductions, an additional "step-down" in O_3 concentrations appears to be in progress.

The period 2013-2015 was marked by an atypical weather pattern that limited the development of heat waves in June-August and the inhibited the frequency of westerly transport aloft. This weather pattern likely contributed to the recent "step-down" in observed O_3 levels, but it appears to have played a minor role compared to ongoing reductions on O_3 precursor emissions.

The skill of all O_3 forecasts in 2015 was comparable to recent years (2013-2015). As was the case in 2013-2014, the 2015 forecasts were below historical average in skill for the Code Orange cases, but they showed steady improvement compared to 2013 and 2014. Notably, both of the observed O_3 exceedance days were correctly forecasted with health alerts issued to the public. The gradual improvement in false alarm rate and hit rate in 2015, the third year of the current "step-down" period, suggests that forecasters are adjusting to the new observed O_3 environment and that a return to the historical high skill for Code Orange forecasts is eminent for 2016-2017.

During the recent "step-down" period of 2013-2015, hot weather and persistence have become unreliable forecast tools. The "classic" conceptual model for O_3-conducive weather patterns, driven by synoptic scale weather and transport effects, is no longer consistently applicable. As a result, forecast methods that rely on synoptic scale phenomena are less likely to be accurate for forecasts in the higher end of the pollutant distribution. This will make the contributions of numerical air quality models, which are steadily improving in accuracy, increasingly important in the coming years.

Figures and Tables

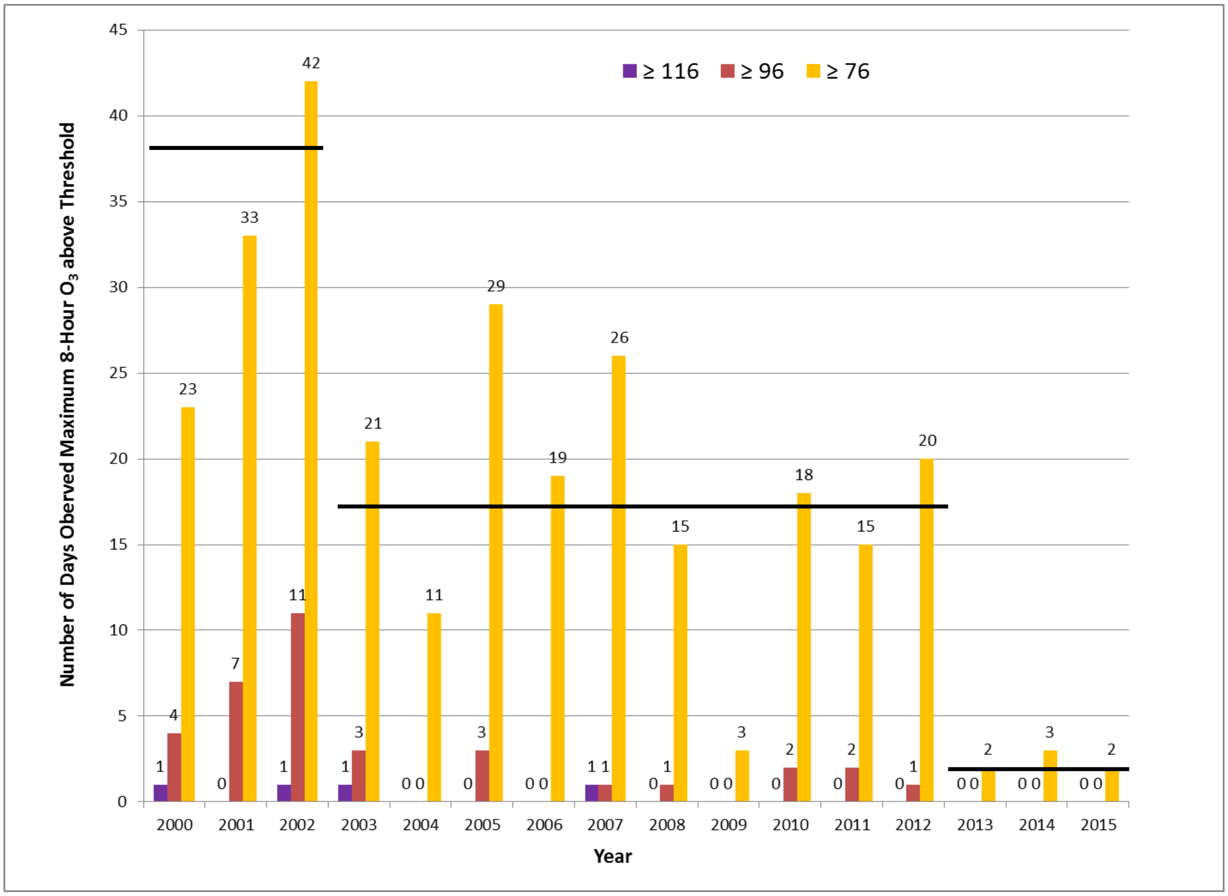

Figure 1. Frequency of days when maximum observed 8-hour O_3 exceeded thresholds of 115 ppbv (purple bars), 95 ppbv (red bars), and 75 ppbv (orange bars) in Delaware for 1990-2015. The black lines indicate the average number of days with observed 8-hour $O_3 \geq 76$ ppbv for the period 1990-2002 (38.3 days), 2003-2012 (17.7 days), and 2013-2015 (2.3 days).

Table 1. Details regarding Code Orange (8-hour $O_3 \geq 76$ ppbv) observed O_3 days in Delaware in 2015. Temperatures were measured at Dover, Delaware (DOV).

Date	Maximum 8-Hr O_3 (ppbv)	Day of Week	Number of Exceeding Monitors	Maximum Temperature at DOV (°F)	Conditions
6/11	94	Thursday	2	91	Transported wildfire smoke
9/17	82	Thursday	3	85	Classic hot and stagnant

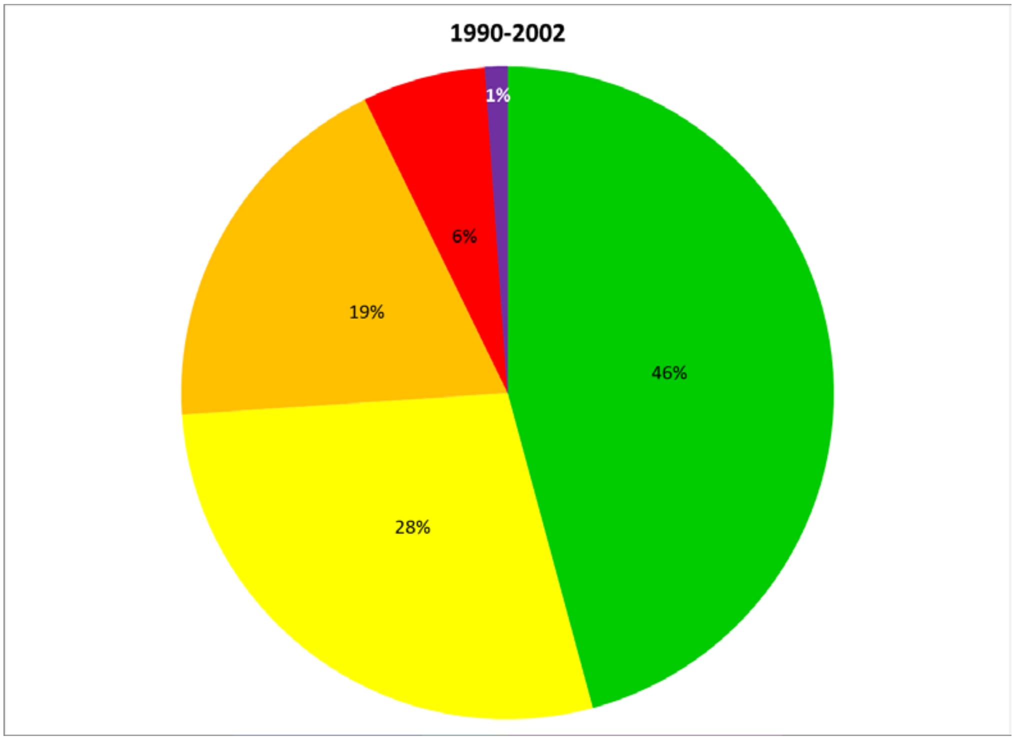

Figure 2. Frequency of AQI color codes for maximum observed 8-hour O_3 in Delaware for 1990-2002.

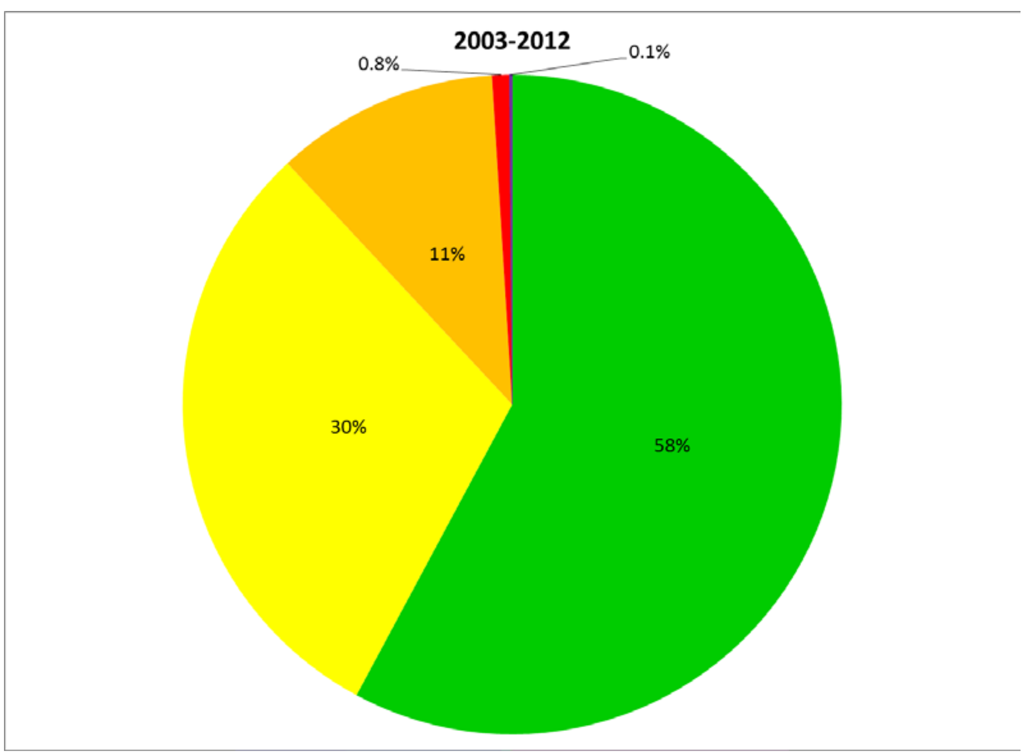

Figure 3. Frequency of AQI color codes for maximum observed 8-hour O_3 in Delaware for 2003-2012.

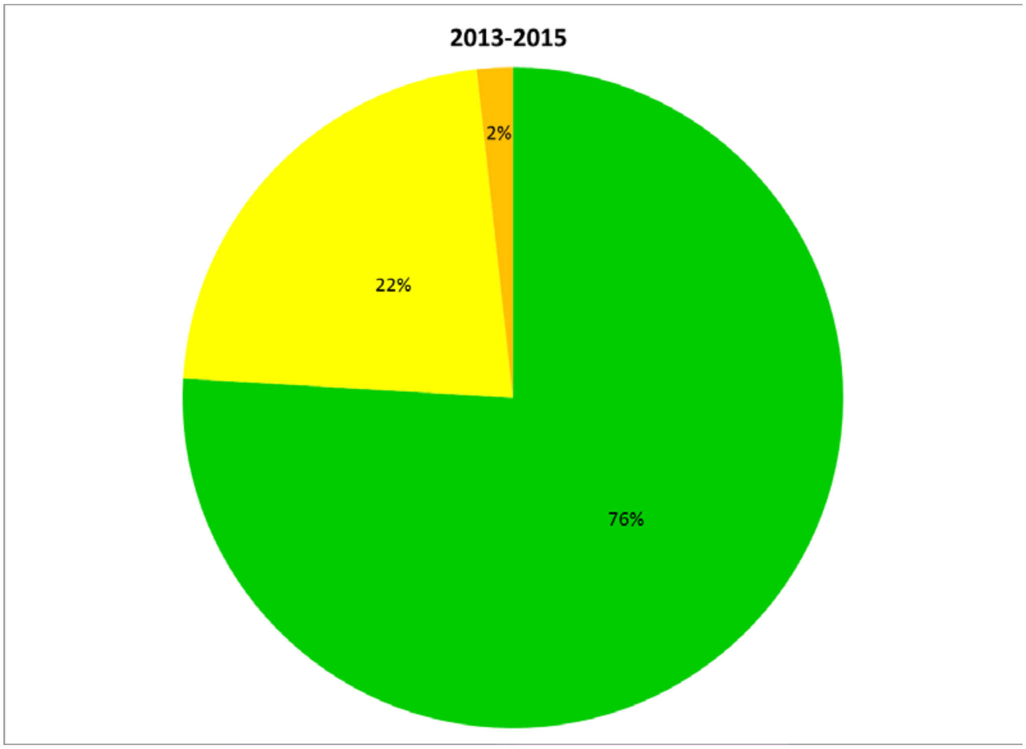

Figure 4. Frequency of AQI color codes for maximum observed 8-hour O_3 in Delaware for 2013-2015.

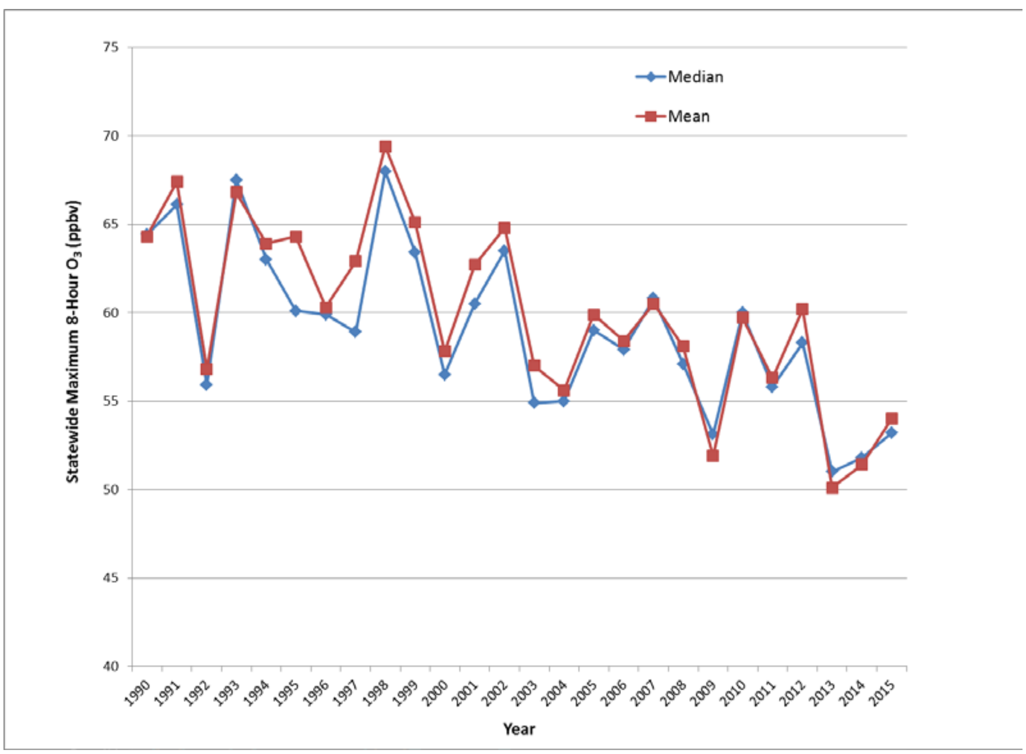

Figure 5. Seasonal (May-September) mean and median maximum observed 8-hour O_3 in Delaware for 1990-2015.

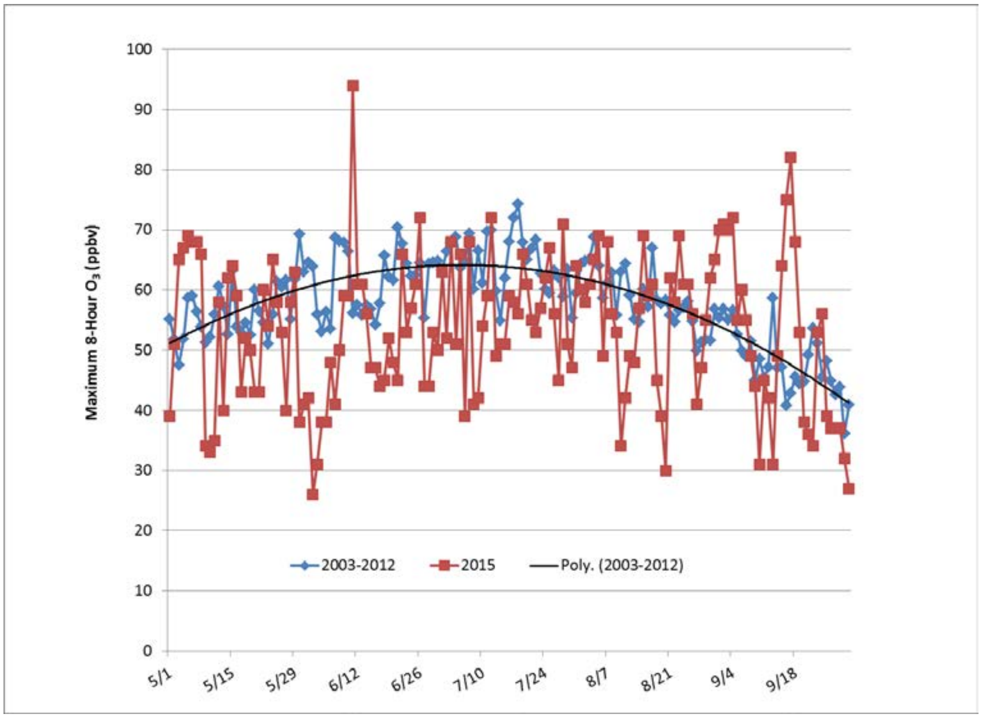

Figure 6. Daily time series of maximum observed 8-hour O_3 in Delaware for 2015 (red line) compared to the 2003-2012 average (blue line). The black line is the best polynomial fit to the 2003-2012 average.

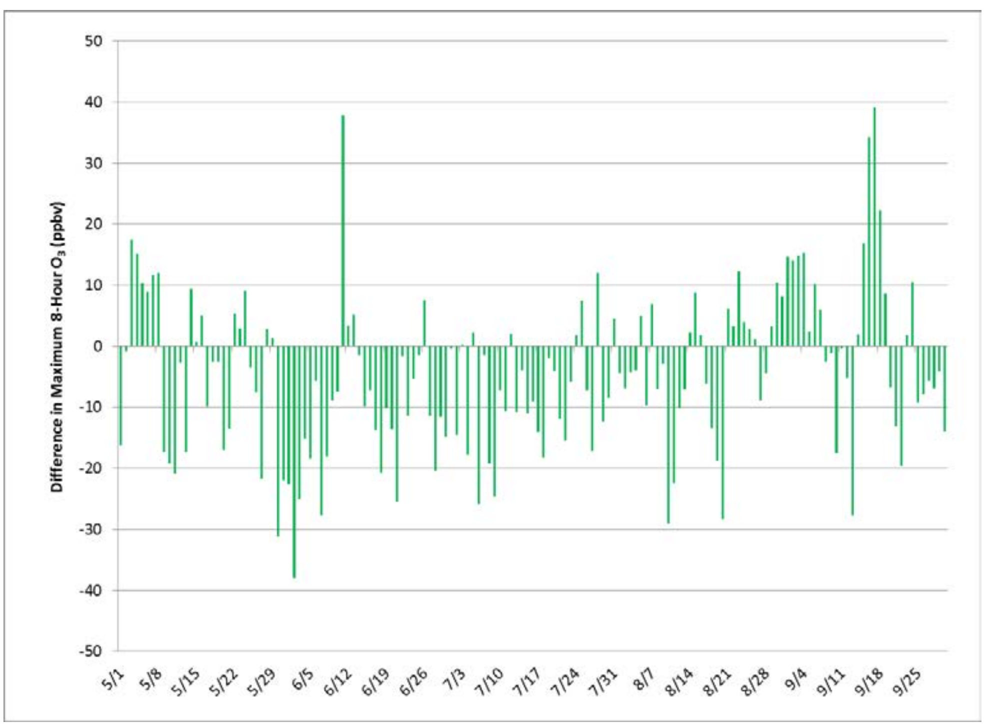

Figure 7. Daily time series of the difference between maximum observed 8-hour O_3 for May 1 to September 30, 2015 compared to the 2003-2012 average.

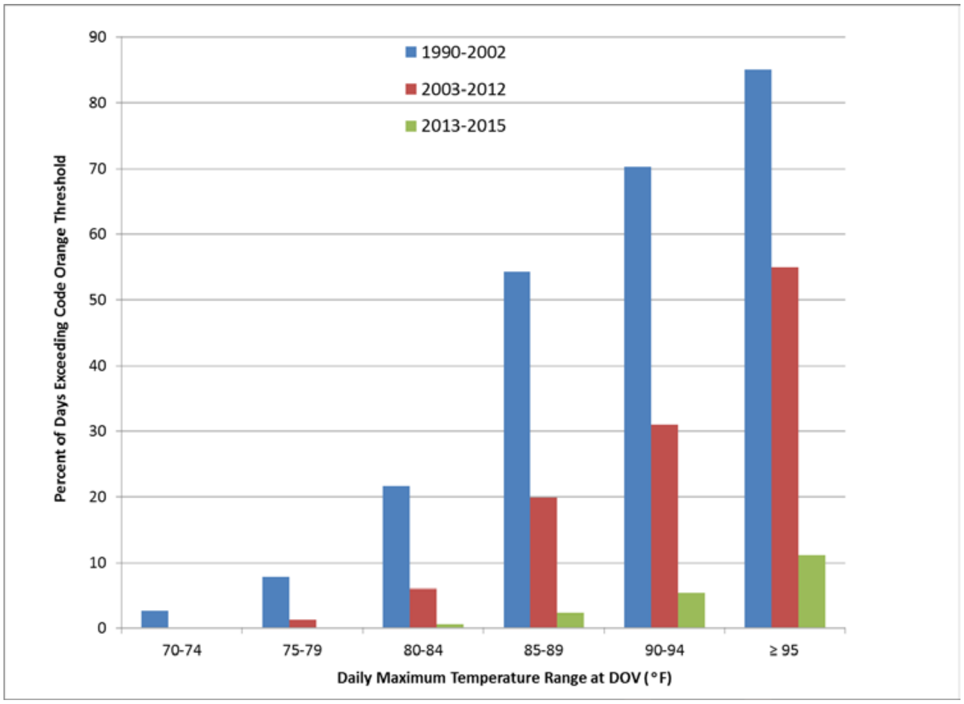

Figure 8. Percent of days exceeding the observed 8-hour O_3 Code Orange threshold in Delaware for bins of maximum air temperature measured at Dover, Delaware (DOV).

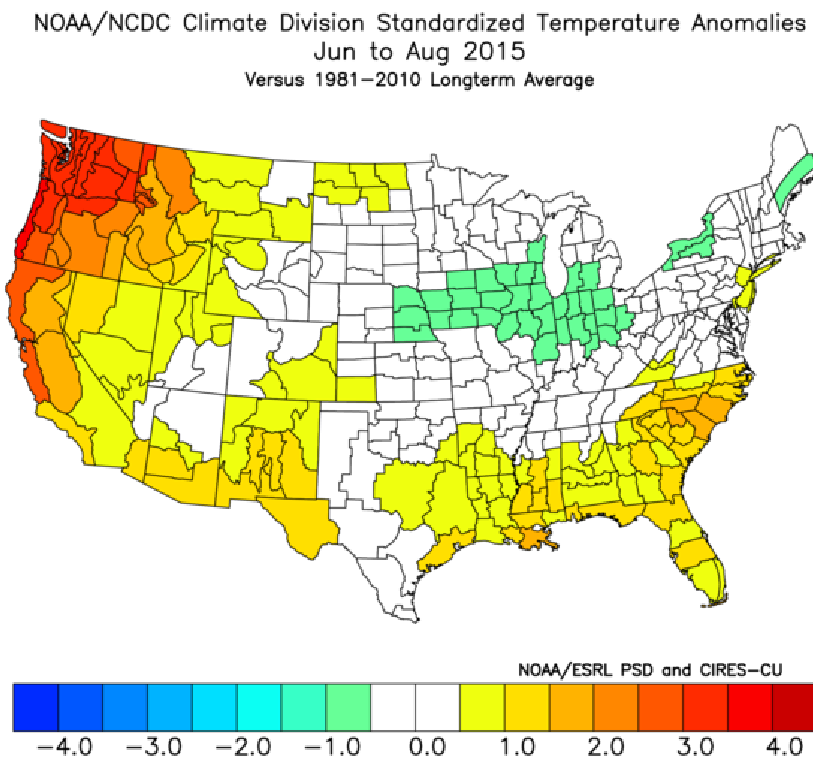

Figure 9. Temperature anomalies (in °F) in the U.S. for June-August 2015 compared to the 1981-2010 average (courtesy of NOAA/ESRL).

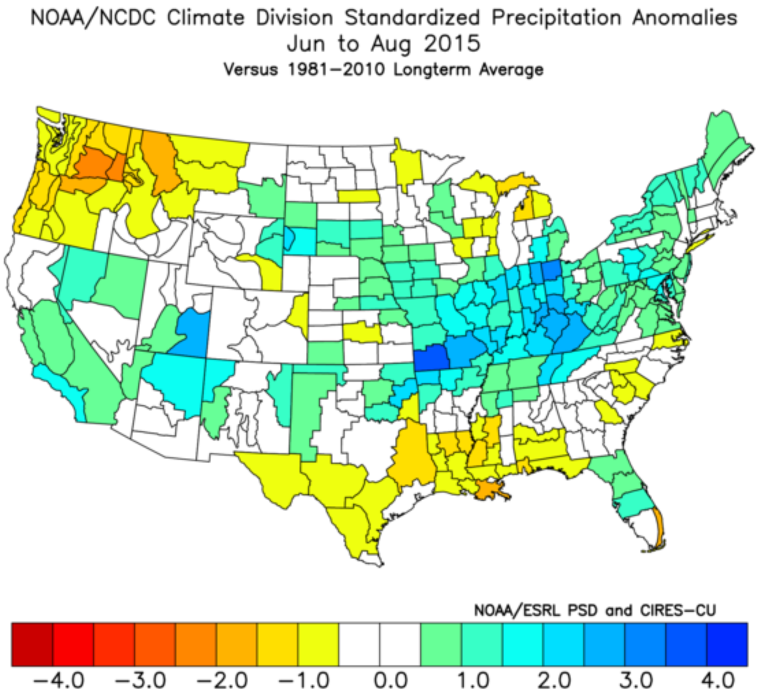

Figure 10. Precipitation anomalies (in inches) in the U.S. for June-August 2015 compared to the 1981-2010 average (courtesy of NOAA/ESRL).

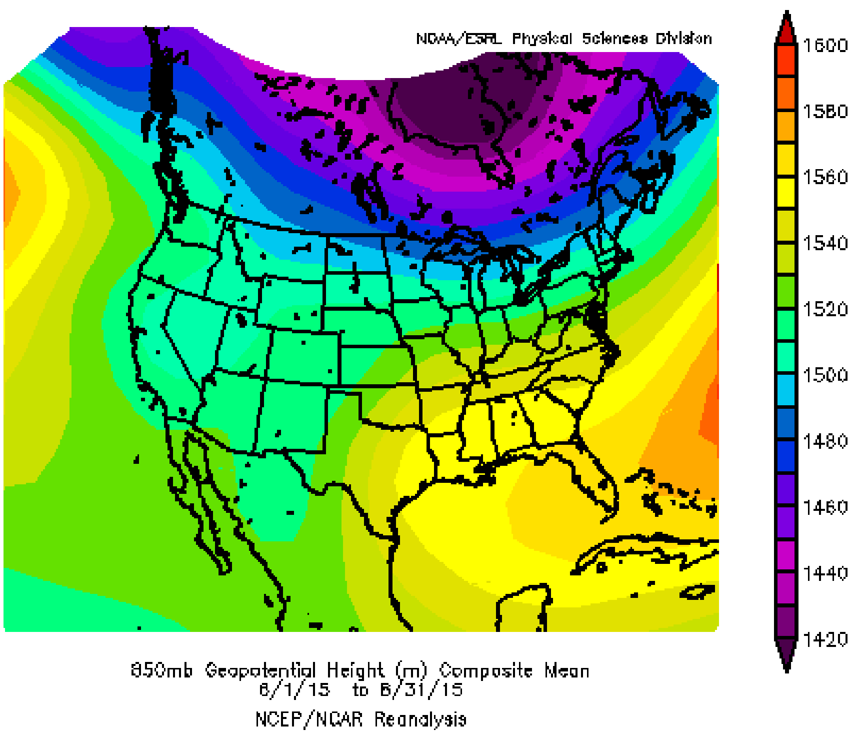

Figure 11. Geopotential height composite mean at 850 mb (~1500 m AGL) in the U.S. for June-August 2015 (courtesy NOAA/ESRL) showing the semi-permanent Bermuda High (yellow colors in Atlantic Ocean) suppressed southward and eastward compared to a summer with high O_3 observations.

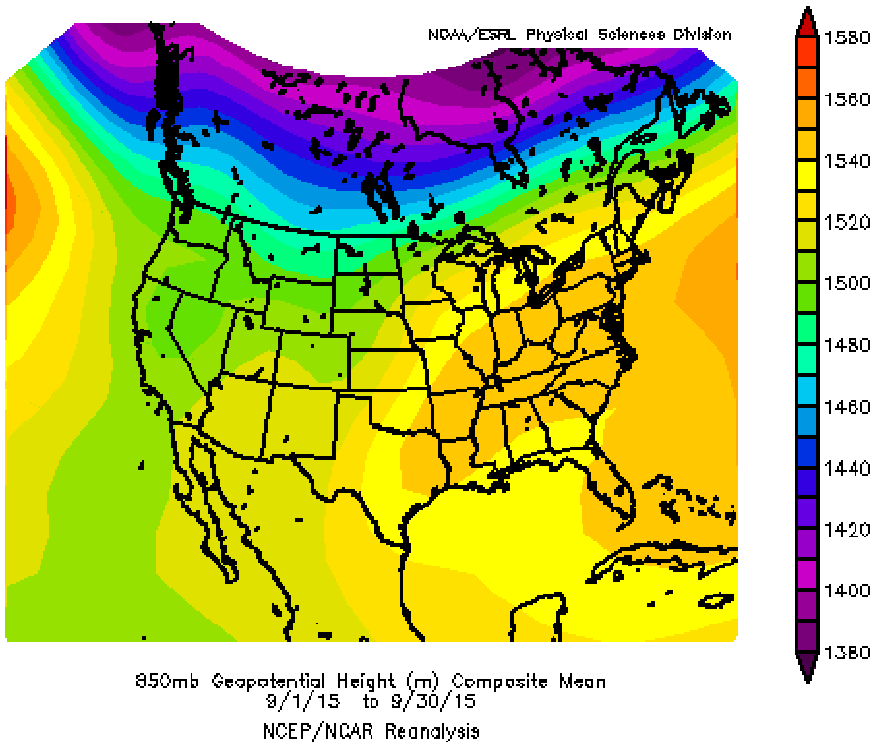

Figure 12. Geopotential height composite mean at 850 mb (~1500 m AGL) in the U.S. for September 2015 (courtesy NOAA/ESRL) showing extension of Bermuda High over the eastern U.S.

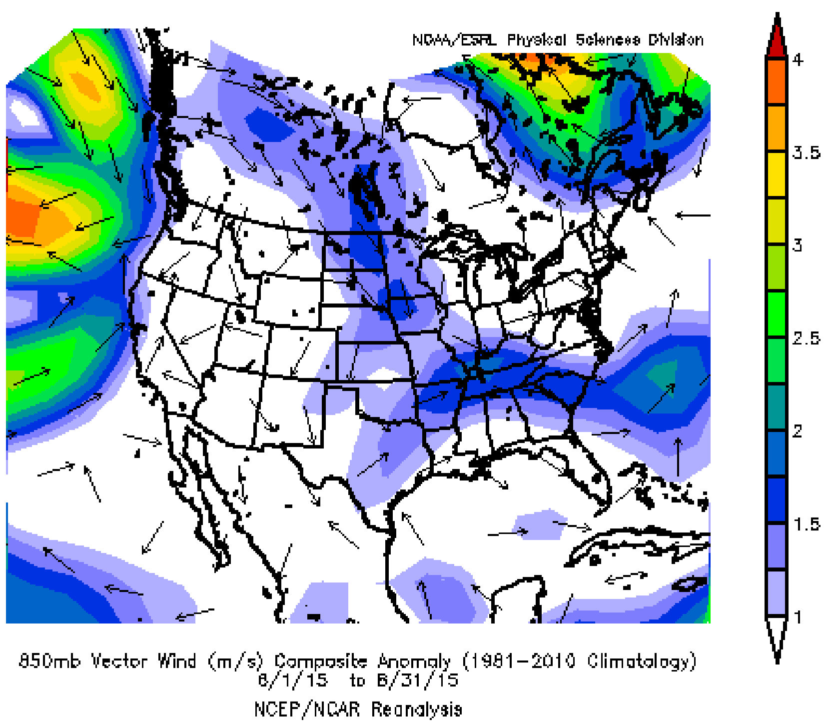

Figure 13. Vector wind composite anomaly at 850 mb (~1500 m AGL) in the U.S. for June-August 2015 (courtesy NOAA/ESRL).

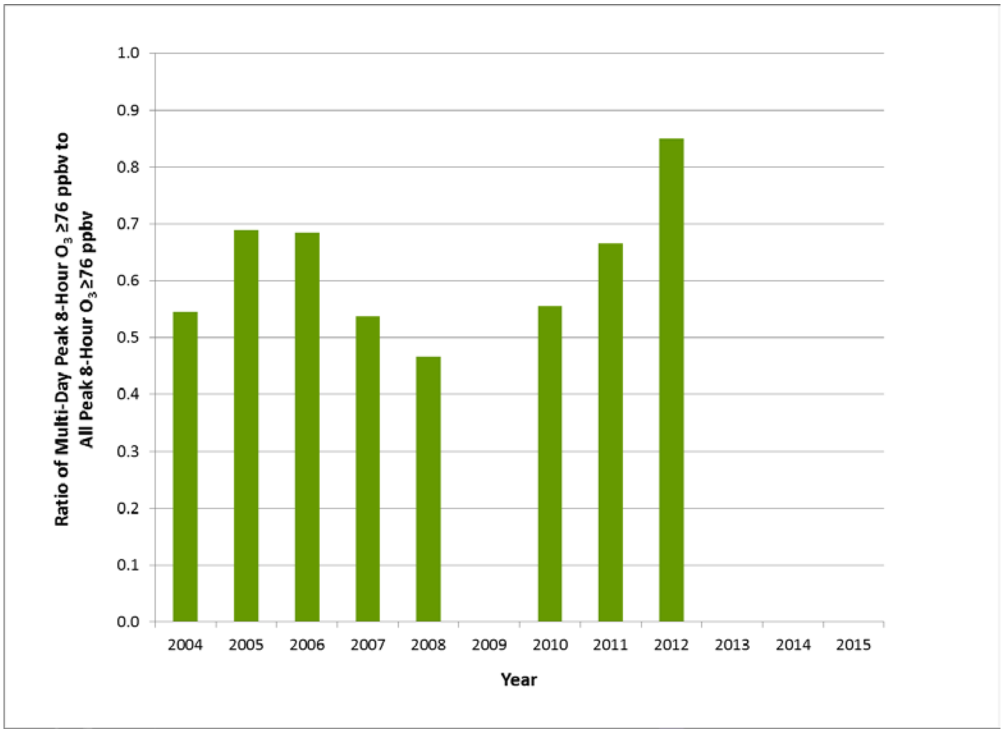

Figure 14. Ratio of multi-day (2 or more in a row) observed Code Orange or higher days to all Code Orange or higher days in Delaware for 2004-2015.

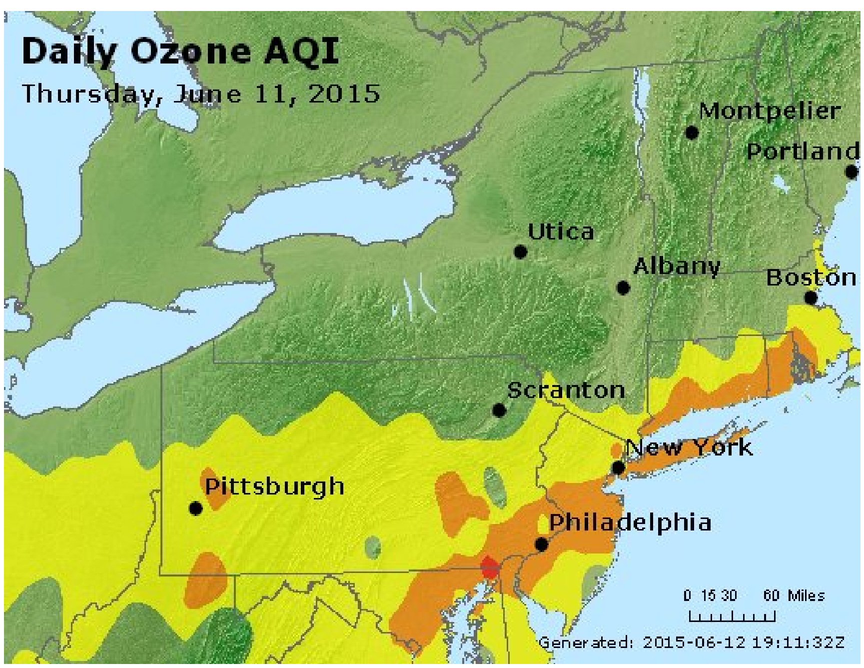

Figure 15. Observed O_3 Air Quality Index (AQI) color codes for the northern Mid-Atlantic and southern New England regions on June 11, 2015, when transported wildfire smoke contributed to Code Orange O_3 along the I-95 Corridor, with an isolated Code Red O_3 observation in northeastern Maryland.

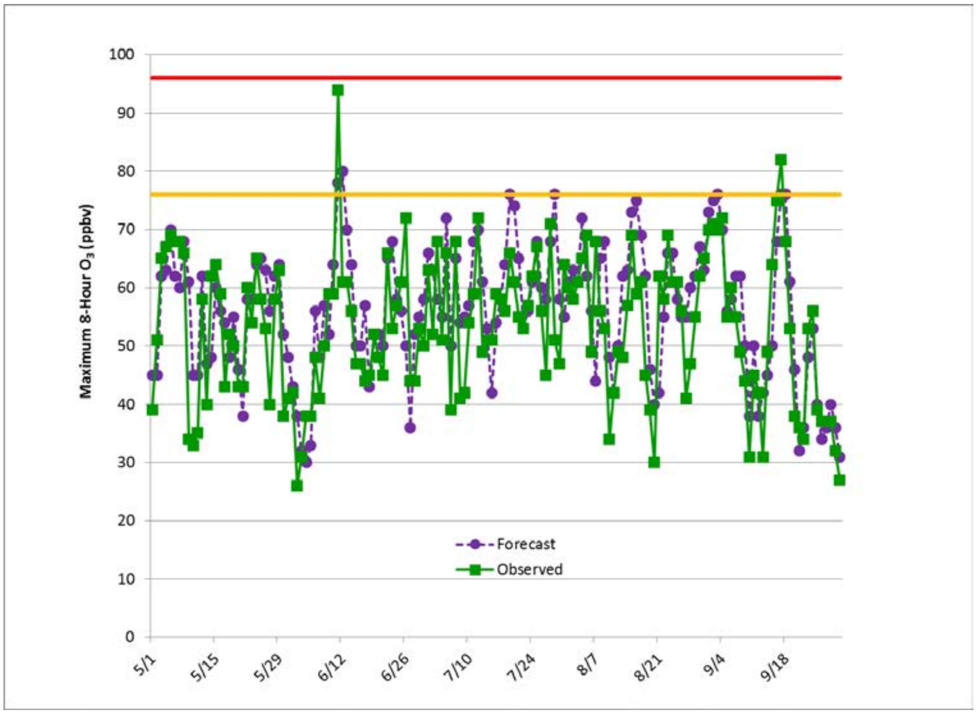

Figure 16. Maximum 8-hour O_3 forecasts and observations for Delaware during May 1 to September 30, 2015. The orange and red lines indicate the Code Orange and Code Red thresholds, respectively.

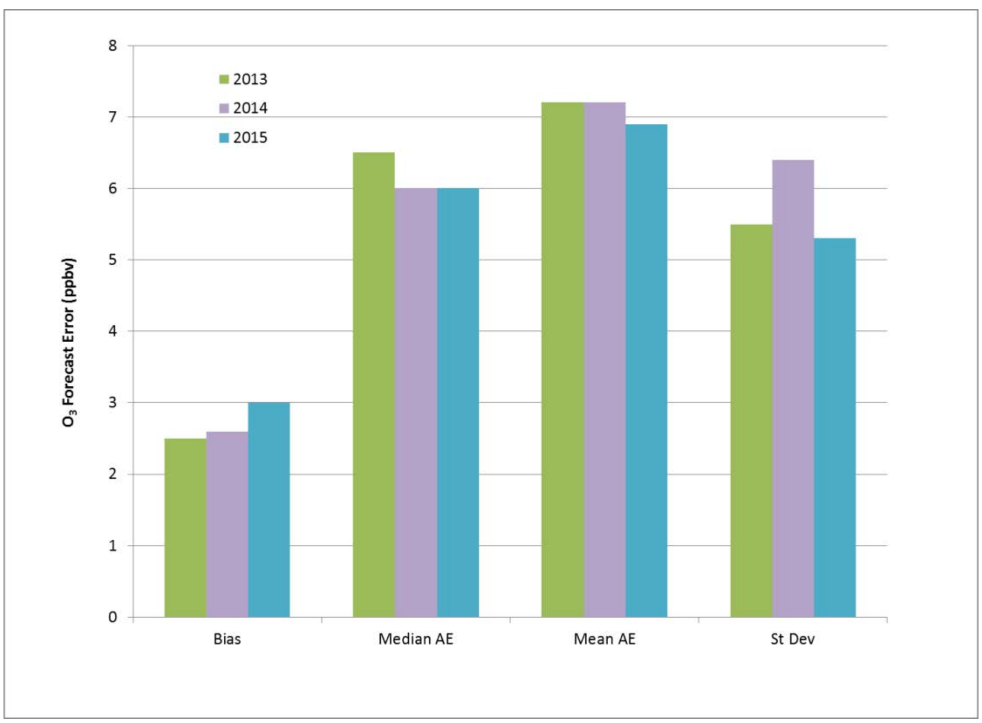

Figure 17. Error statistics for all maximum 8-hour O_3 forecasts in Delaware for 2013-2015. "Median AE" refers to median absolute forecast error, "Mean AE" refers to mean absolute error, and "StDev" refers to the standard deviation of the mean absolute error.

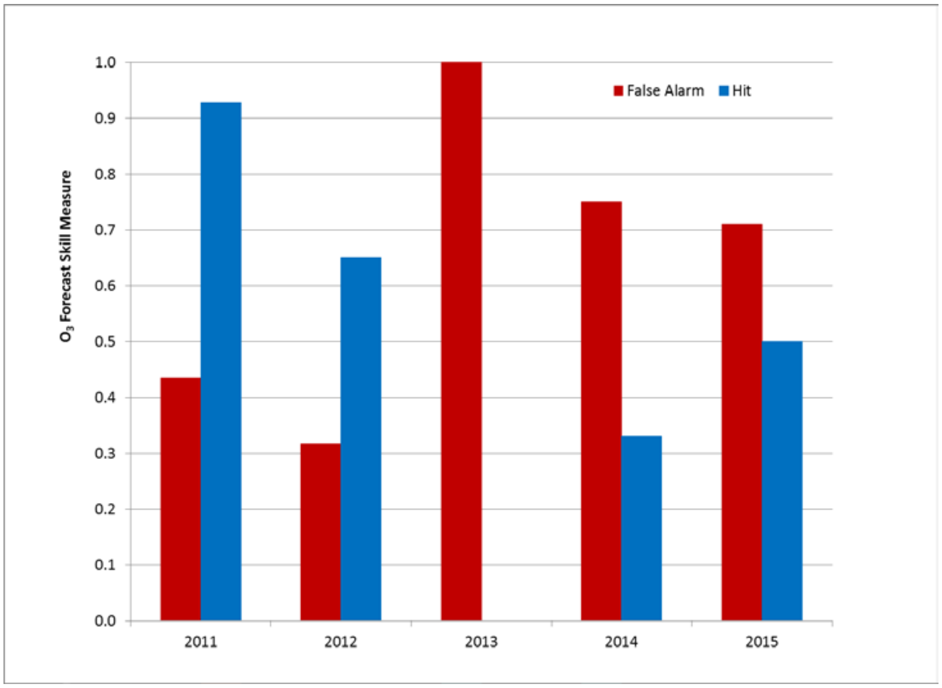

Figure 18. False alarm rate and hit rate for Code Orange O_3 forecasts in Delaware for 2011-2015.

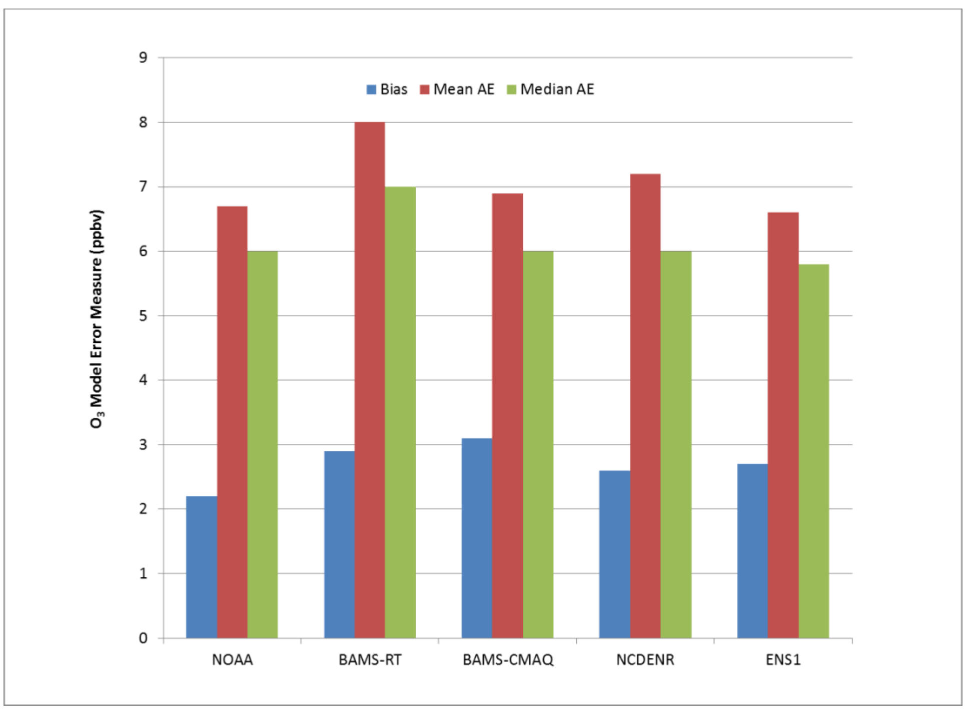

Figure 19. Error statistics (as in Figure 17) for air quality numerical forecast model guidance for all maximum 8-hour O_3 predictions in Delaware in 2015. Two variations of the Baron Meteorological Services (BAMS) models are shown (CMAQ and RT). The ENS1 is a mean value from the NOAA, BAMS-RT, BAMS-CMAQ, and NCDENR model forecasts.

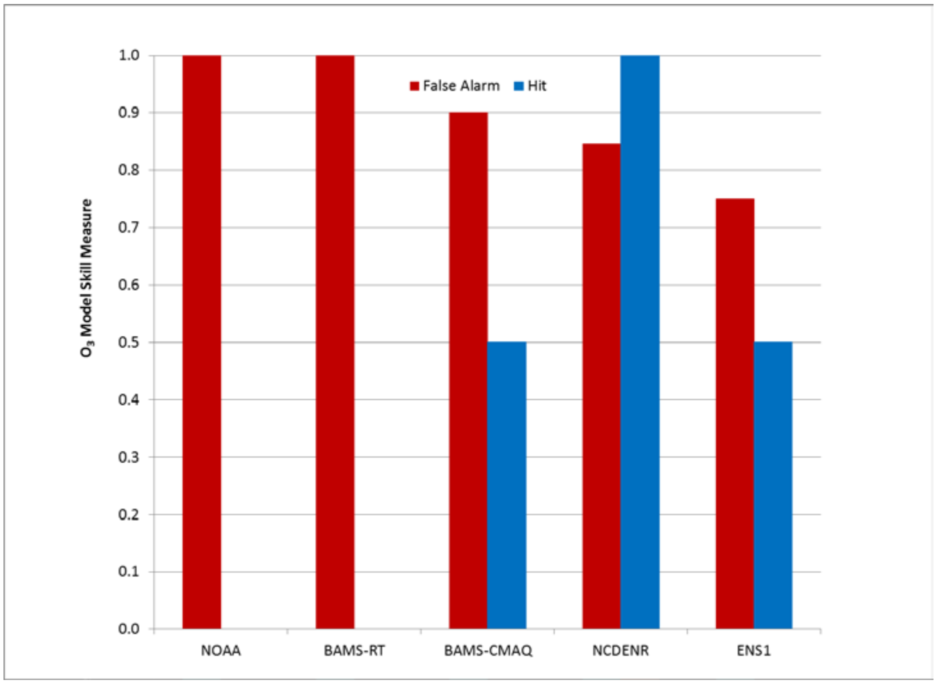

Figure 20. False alarm rate and hit rate for Code Orange O_3 predictions by air quality numerical forecast models in Delaware in 2015.

Appendix A. Skill Measures for Threshold Forecasts

The determination of the skill of a threshold forecast (e.g., Code Orange air quality) begins with the creation of a contingency table of the form:

Contingency Table for Threshold Forecasts			
		Observed	
		Yes	**No**
Forecast	**Yes**	a	b
	No	c	d

For example, if Code Orange O_3 concentrations are both observed and forecast ("hit"), then one unit is added to "a." If Code Orange O_3 is forecast but not observed ("false alarm"), then one unit is added to "b."

Basic Skill Measures

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

$$\textbf{Bias (B)} = \frac{a+b}{a+c}$$

Bias determines whether the same *fraction* of events are both forecast 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.

$$\textbf{False Alarm Rate (F)} = \frac{b}{a+b}$$

This is a measure of the rate at which false alarms (high O_3 forecast but not observed) occur.

$$\textbf{Hit Rate (H)} = \frac{a}{a+c}$$

The hit rate is often called the "probability of detection"

Miss Rate = 1 – H

Correct null forecasts:

$$\textbf{Correct Null (CNull)} = \frac{d}{c+d}$$

Accuracy:

$$\textbf{Accuracy (A)} = \frac{a+d}{a+b+c+d}$$

Other Skill Measures

Generalized skill scores (SS_{ref}) measure the improvement of forecasts over some given reference measure. Typically the reference is persistence (current conditions used as forecast for tomorrow) or climatology (historical average conditions).

$$\textbf{Skill Score (SS}_{\textbf{ref}}\textbf{)} = \left(\frac{A - A_{ref}}{A_{perf} - A_{ref}}\right) * 100\% = \text{ nn}\%$$

The skill score is typically reported as a percent improvement of accuracy (A) with respect to a reference forecast. The reference forecast accuracy (A_{ref}) is typically climatology or persistence. The perfect forecast (A_{perf}) is usually 1 (e.g., for hits) or 0 (e.g., for false alarm).

Additional measures of skill can be determined. The Heidke skill score (HSS) compares the proportion of correct forecasts to a no skill random forecast. That is, each event is forecast randomly but is constrained in that the marginal totals (a + c) and (a + b) are equal to those in the original verification table.

$$\textbf{HSS} = \frac{2(ad - bc)}{(a+c)(c+d) + (a+b)(b+d)}$$

For this measure, the range is [-1,1] with a random forecast equal to zero.

Another alternative is the **critical success index (CSI) or the Gilbert Skill Score (GSS)** also called the **"threat" score**.

$$\textbf{CSI} = \frac{a}{a+b+c} = \frac{H}{1+B-H}$$

For this measure, the range is [0,1]. Since the correct null forecast is excluded, this type of measure is effective for situations like tornado forecasting where the occurrence is difficult to determine due to observing bias, i.e., tornados may occur but not be observed. This can also be the case for air quality forecasting when the monitor network is less dense. Note, however, that the random forecast will have a non-zero skill.

The **Peirce skill score (PSS), also known as the "true skill statistic"** is a measure of skill obtained by the difference between the hit rate and the false alarm rate:

$$\textbf{PSS} = \frac{ad - bc}{(a+c)(b+d)} = \text{H} - \text{F}$$

The range of this measure is [-1,1]. If the PSS is greater than zero, then the number of hits exceeds the false alarms and the forecast has some skill. Note, however, that if d is large, as it is in this case, the false alarm value (b) is relatively overwhelmed. The advantage of the PSS is that determining the standard error is relatively easy.

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.