Final Report

Ozone Air Quality in Delaware – 2012

William F. Ryan

Nathan T. Wiles

Department of Meteorology and Atmospheric Sciences

The Pennsylvania State University

February, 2013

Ozone Concentrations in 2012

Ozone (O_3) concentrations in Delaware have decreased significantly since the regional NO_x emissions reductions regulations, the so-called "NOx SIP Rule", went into effect ~ 2002 (**Figure 1**). This is reflected in mean peak O_3 concentrations (**Figure 2**) and is more pronounced in the upper percentile brackets (**Figure 3**). Unlike the longer term trend, mean ozone (O_3) concentrations during the summer of 2012 were, by some measures, the highest observed since 2002. In terms of "bad air" days, 2012 was slightly below the recent (10 year) average frequency of Code Red cases (1 observed compared to a 1.3 average, **Figure 4**) and slightly above the frequency of Code Orange cases (20 days in 2012 compared to an 18.0 average, **Figure 5**). As has been the case since 2007, there were no Code Purple (≥ 115 ppbv) days in 2012. In fact, there have only been two total Code Purple days during the last 10 years – compared a *yearly* average of 1.8 such days from 1990-2002. A plot of the overall reduction in bad air days since 2002 for a variety of threshold concentrations is shown in **Figure 6**. Pie charts showing changes in the frequency of all AQI color codes since 1990 are given in **Figure 7**.

Meteorology and High O_3 Concentrations

A time series of peak daily O_3 in Delaware during the summer of 2012 is compared to recent years in **Figure 8**. Observed O_3 in 2012 was consistently above average from late June to late July. This period was marked by very warm and dry weather and thus conducive to O_3 formation. The 34 day period from June 28 to July 31 featured 20 days ≥ 90°F and 26 days ≥ 88°F. Roughly one-third of the 90°F days reached the Code Orange threshold. Temperature and O_3 are strongly correlated overall but association of hot weather with Code Orange or higher O_3 has been steadily weakening since 2002. The probability of Code Orange or Code Red O_3 occurring for a given range of temperature is presented in **Figure 9**. While days ≥ 95°F are still likely to be bad air days, the frequency of Code Orange cases on slightly "cooler" days (90-94°F) has dropped from a 70% probability prior to 2003, to only 30%.

Given the association of temperature and O_3, the summer season of 2012 was clearly conducive to O_3 formation. For the JJA period, temperatures at ILG were ~ 2°F above average with a summed precipitation deficient of 2.2". Putting this in perspective, JJA of 2012 was warmer than average even by the standards of the past two decades (**Figure 10**) and in the 94th percentile for the full 117 year period of observations (**Figure 11**). To have nearly average O_3 levels during a very warm summer season is an indication of continuing progress in air quality efforts.

Code Orange O_3 days in Delaware during 2012 occurred primarily in cases of westerly transport above the surface. This is the standard transport pattern for poor air quality events in the mid-Atlantic and the frequency of these cases early in the summer of 2012 led to the large number of Code Orange cases observed from late June through July. In **Figure 12**, 850 mb (~ 1500 m AGL) geopotential height is averaged for the 15 worst O_3 cases in 2012 (peak 8-hour $O_3 \geq 80$ ppbv). Transport of air parcels at this level is roughly parallel to the isoheight lines. Therefore, Delaware experienced, on average, west-northwest winds aloft. This is corroborated by an average wind vector plot for the same days (Figure 13). It is worth noting that winds aloft are rather strong, on the order of 6-7 ms^{-1} (13-16 mph). This translates to a 24-hour transport distance of approximately 520-610 km (320-380 miles) or roughly the distance from Cleveland, Ohio to Wilmington, Delaware. Slightly lower in the atmosphere (925 mb), a similar transport pattern is found (Figure 14).

As an example of the impact of westerly transport, a time series of winds aloft from the State of Maryland wind profiler at Horn Point on the Eastern Shore of Maryland for the high O_3 case of July 17 is shown in Figure 15. Sustained northwest winds are observed overnight on July 16 and into the afternoon hours of July 17. As this transport direction places Delaware downwind of both the Ohio River Valley and the I-95 Corridor, it is likely that this air mass contains high concentrations of both O_3 and its precursors. A time series of O_3 concentrations at elevated monitor locations in Virginia (Shenandoah National Park, 1 km AGL) and Pennsylvania (Methodist Hill, 0.7 km AGL) for this event shows regional scale concentration ≈ 60 ppbv (Figure 16). When the surface-based inversion breaks in Delaware during the late morning of July 17, the transported high O_3 air aloft is mixed quickly downward. Because the high O_3 concentrations aloft are regional in scale, concentrations across Delaware move rapidly upward in lock step as mixing continues. By mid-day, statewide O_3 concentrations show little variability from location to location with a magnitude nearly equal to the inferred "regional load" - in the 60-70 ppbv range (Figure 17). Later in the day, of course, local effects cause large local scale variations in peak O_3 concentrations. We see a slightly stronger example of the transport effect during the high O_3 event of June 10 with upwind and local late morning concentrations higher by ≈ 5 ppbv (Figure 18 and Figure 19).

Forecast Skill

Overall forecast skill is given in Figure 20 and Table 1 for both the public forecasts as well as the numerical O_3 model forecast guidance. Model performance will be addressed in more detail below. Overall forecast skill in 2012 was slightly better than in 2011. The main improvements were a reduction in forecast bias (+2.6 ppbv to +1.0 ppbv) and a reduction in mean absolute error (improvement of 0.8 ppbv). The reduction in mean absolute error suggests more consistency in the forecasts. In terms of AQI, the correct color code was forecast on 73% of all days. The majority of the missed color code forecasts (58%) were errors at the good/moderate threshold (Green/Yellow).

Of most interest is forecast skill in the high end of the O_3 distribution. In Figure 21, forecast skill in high O_3 cases is shown and compared to the reference "persistence" forecast. Persistence is simply today's observed O_3 forecast for tomorrow. Because O_3 is a relatively long lived pollutant, persistence is typically a skillful forecast, explaining ≈ 35-40% of the variance in O_3 in the mid-Atlantic region and therefore a better skill reference than a random measure. The skill measures shown in Figure 21, and in several succeeding figures, are described in more detail in Appendix A. As seen in Figure 21, the public forecast was significantly more skillful than persistence in all measures but there is also considerable room for improvement. For example, the probability of detection, or hit rate, was only 0.65 in 2012. That is, 65% of observed Code Orange days were correctly forecast. As a general rule, a hit rate of 0.75 or better is an achievable goal. These “missed” forecasts (7 total) were seen in a variety of weather situations but included two so-called “onset” days - the first day of a multi-day poor air quality event. In these cases, O_3 concentrations rose from the good range (< 60 ppbv) to Code Orange in less than 24 hours. An example of this phenomenon is shown in Figure 22 for June 27-28. Another unusual missed forecast was May 12 with Code Orange observed although maximum surface temperature remained in the mid-70’s F.

The false alarm rate in 2012 was 0.31, that is, approximately 30% of days with Code Orange forecasts did not observe Code Orange. This is slightly higher than the typical goal of ≤ 0.25 but it must be noted that, on 4 of the 6 false alarms days, Code Orange concentrations were observed in *both* the Philadelphia and Baltimore-Washington forecast areas. AirNow maps of the four false alarm days are given in Figure 23. Note that in one case (July 6), O_3 concentrations on the Maryland border (Millington, MD) were high enough to cause the mapping algorithm to extend the Code Orange O_3 contour into central Delaware - even though no Delaware monitors observed Code Orange levels. As a result, we feel that reducing missed Code Orange forecasts is more of a priority than reducing false alarms. Greater attention to transported O_3 in “onset” cases may help to reduce the false alarm rate in 2013.

Forecast Model Performance

A summary of forecast model performance is given in Table 1 for all cases. Skill scores for high O_3 cases are given in Figure 24. The SUNY and BAMS models performed best overall with the NOAA model subject to frequent false alarms. Each of the models, however, varied in skill depending on the specific skill score measure in question. For example, the SUNY model issued few false alarms but had a weaker hit rate. Based on research results from 2011 showing that a combination (ensemble) of model forecasts provided better results than any single model, we prepared a set of ensembles for operational use in 2012. The results of the model ensembles for Code Orange cases are given in Figure 24 and Figure 25. As seen in Figure 25, the best performing ensemble was the “all of the above” ensemble where all three forecast models were weighted evenly. Similar results were obtained in the Philadelphia metropolitan area.

The performance of the NOAA model in Delaware was investigated in more detail. It appears that the NOAA model suffers from a “seasonal drift” in forecast bias. That is, the NOAA model tends to over-predict peak O_3 more as the summer moves on. In Figure 26, the NOAA model bias for forecasts in the Code Yellow or Orange ranges is given by month. While the model has a negative bias (under-prediction) in May, the bias shifts steadily to an over-prediction mode thereafter and, by August, the NOAA model is strongly over-predicting peak O_3. Another way of looking at this effect is by calculating the cumulative bias in the forecast models (Figure 27). Cumulative bias is simply the running sum of forecast model bias. The slope of the cumulative bias line indicates the direction of the forecast bias.

The NOAA model shows a steady high bias beginning in mid-July and continuing to the end of the forecast season. The other models do not display this systematic bias and it appears that the "damping" of the NOAA model bias by the ensemble weighting of other models accounts for a portion of the skill of the ensemble models.

Extreme O_3 Case of June 29

Although, as shown above, the frequency of days with extreme O_3 concentrations has been steadily falling since the early 2000's, weather conditions can still drive individual cases of very high O_3. Such a case occurred in June 29, 2012. Code Red conditions were observed in the mid-Atlantic region as a whole and in Delaware as well. A brief summary of this event is provided below.

Code Orange to Code Red O_3 levels were observed as early as June 27 in the Midwest. Due to prevailing westerly winds aloft, this high O_3 was steadily transported to the east on June 28 (Figure 29). Although concentrations were in the good range in Delaware on June 27, they quickly rose to the Code Orange range by June 28 (Figure 30). Westerly transport continued apace through the evening of June 28 and into the morning of June 29 (Figure 31). Observations in remote areas along the spine of the Appalachians west of Delaware on June 28-29 show that the regional load of O_3 was on the order of 60-80 ppbv, the highest observed this summer (Figure 32). Winds from the northwest extended down to the surface on the morning of June 29 in Delaware (Figure 33) and although a bay breeze developed east of the I-95 Corridor later in the afternoon, the ocean breeze did not extend very far into Delaware (Figure 34). As a result, following two days of sunny, hot weather, with sustained transport from the west, O_3 concentrations reached the Code Red range in Delaware (Figure 35). As evening fell on June 29, a "derecho" – a grouping of severe thunderstorms accompanied by extremely strong straight line winds – approached from the west (Figure 36). As this line of storms crossed the mid-Atlantic, it brought the poor air quality event to an end although at the cost of significant damage across the region (Figure 37).

Outlook for 2013

There are two issues that may impact O_3 forecast skill in 2013. First, NOAA has proposed terminating their operational air quality model in March, 2013. This decision is, in part, due to budget constraints but also reflects ambivalence at NOAA and NWS regarding the creation of forecast guidance for a product (air quality forecasts) that are not directly issued by NWS forecast offices. The comment period for this plan has closed, we submitted a comment opposing the termination, but we have not been advised of any final decision. The appointment last week of a new director of the NWS, Louis Uccellini, gives us some faint hope. The loss of the NOAA model will impact O_3 forecasting in a negative manner. The model itself is very successful in the mid-Atlantic region and, when included in a larger model ensemble, provides very skillful forecasts in high O_3 cases. In addition, only the NOAA and BAMS forecast models are reliably available every day.

As a result, the ensemble forecasts that showed great promise in 2012 may not be available in 2013. In addition to the possible loss of the NOAA model, the SUNY model is no longer operational as of the date of this report, although there are plans to restart later in 2013. Our plan for 2013 is to investigate the value of including additional models in our forecast preparation. The North Carolina Department of Natural Resources (NCDNR) plans to run their O_3 model through 2013. We analyzed this model in 2012 and it showed good skill but suffered from unreliability – forecasts not available by the deadline. Environment Canada (EC) has upgraded their meteorological and chemistry model and we will assess the skill of this model in 2013. Previous EC O_3 models have suffered from an over-prediction bias in the mid-Atlantic. The experimental High Resolution Rapid Refresh Model with Chemistry (HRRR-Chem) is currently running at NOAA-ESRL. This is not technically an operational model so that reliability is an issue. For all three of the new models, 2013 will only be an assessment year as we have no comprehensive information on their forecast skill or bias. It may be that we can cobble together another ensemble version for use in 2014.

The other lingering issue relates to the dissemination of our forecasts through NWS forecast offices. As you know, the Mount Holly NWS office, in conjunction with Sterling NWS, stopped posting our Code Orange forecasts on their warnings page. While they will post Code Red forecasts, Code Red O_3 cases have become extremely rare and peak O_3 in those cases are so close to the threshold that we have no forecast skill in these cases and typically only forecast Code Orange concentrations except in unusual cases. As a practical matter, then, no air quality warnings appear on the NWS warnings home page for the I-95 Corridor from Washington to Philadelphia. An example of the impact of this was seen on June 20, 2012. Air quality alerts were issued from North Carolina to Massachusetts but, in the Washington DC-Philadelphia Corridor only the southern Delmarva Peninsula (served by the Wakefield NWS office) showed an alert on their home page (Figure 38 and Figure 39). Widespread Code Orange O_3 was observed on June 20 (Figure 40). We were assured by our contacts at the main NOAA office that the alerts would return although on a modified basis but have not yet heard if they plan to do so.

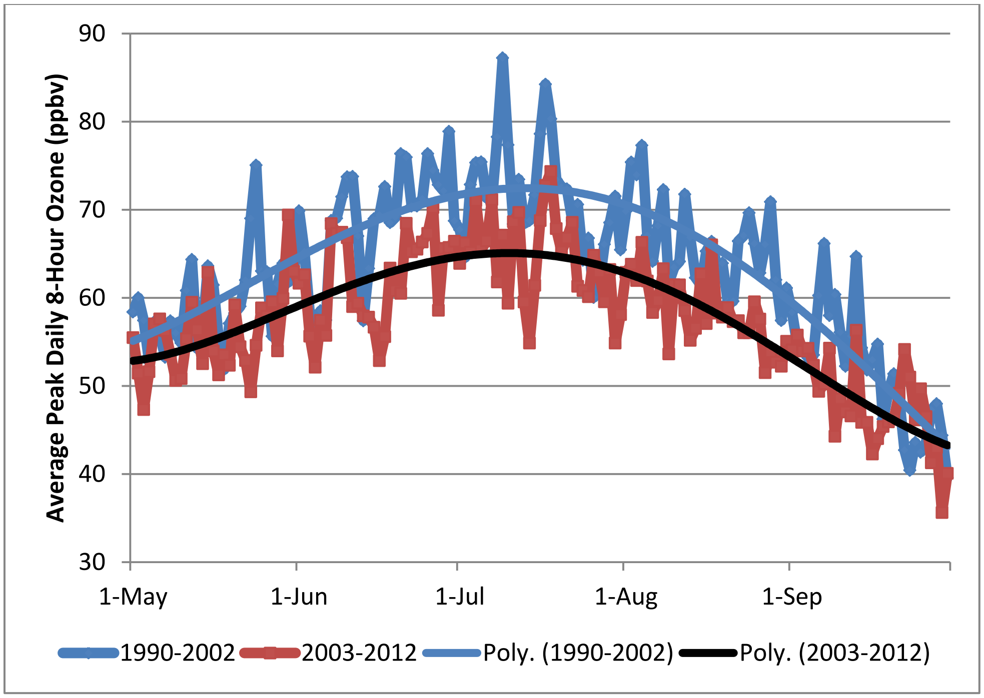

Figure 1. Average peak daily 8-hour O_3 in Delaware for 1990-2002 (blue) and 2003-2012 (red) with best cubic fit lines (blue and black respectively).

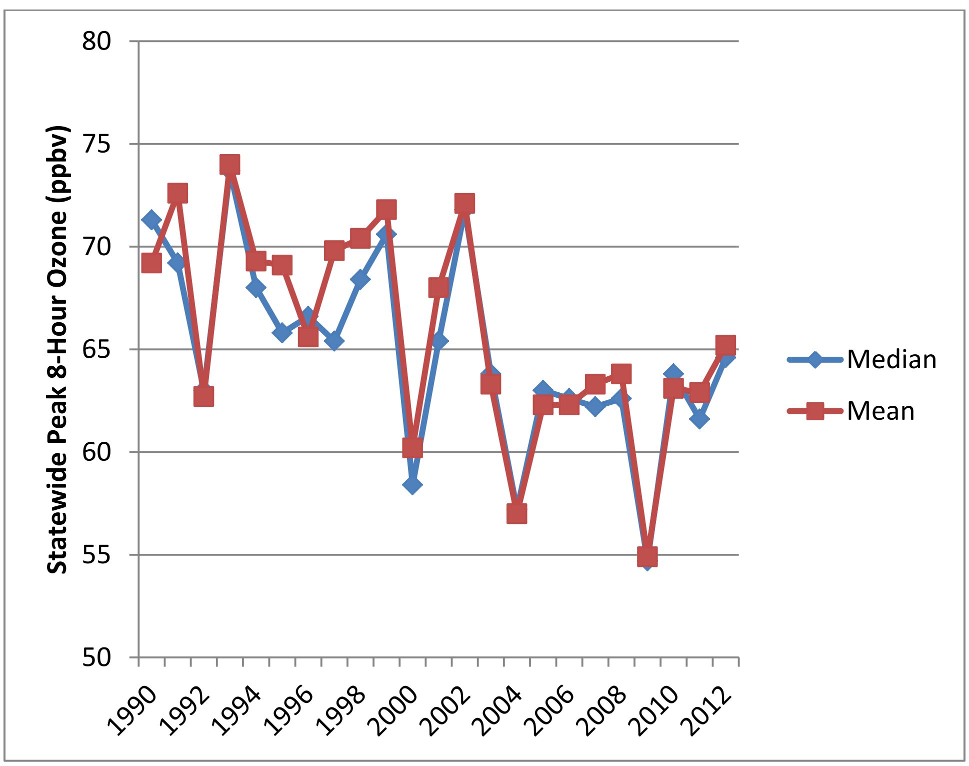

Figure 2. Seasonal (May-September) mean and median peak 8-hour O_3 in Delaware.

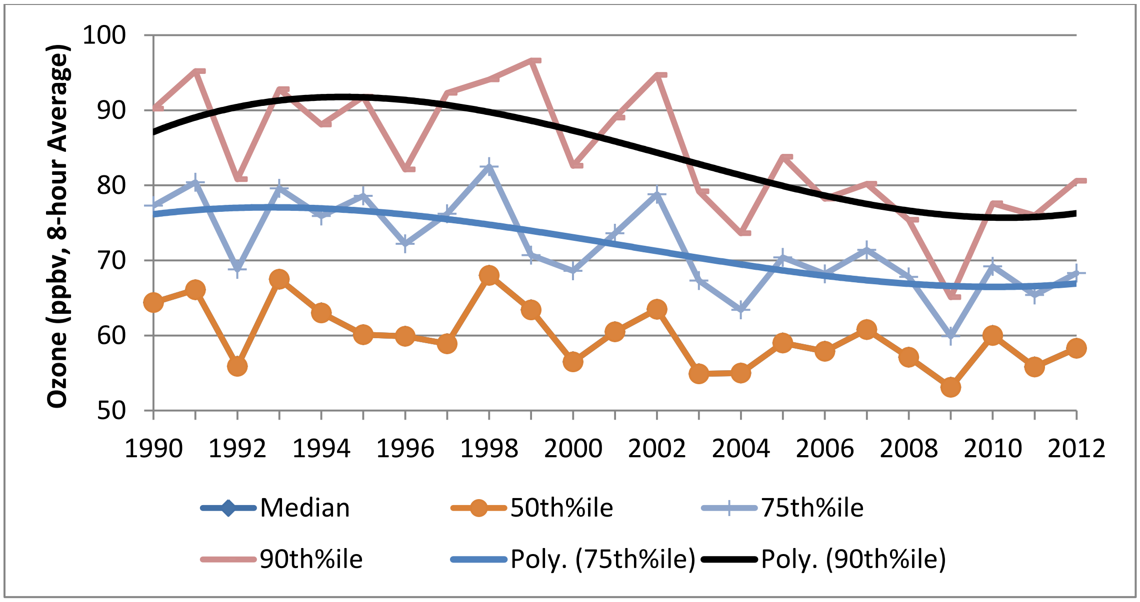

Figure 3. As in Figure 2 but for 50^{th} (median), 75^{th} and 90^{th} percentile of O_3 distribution.

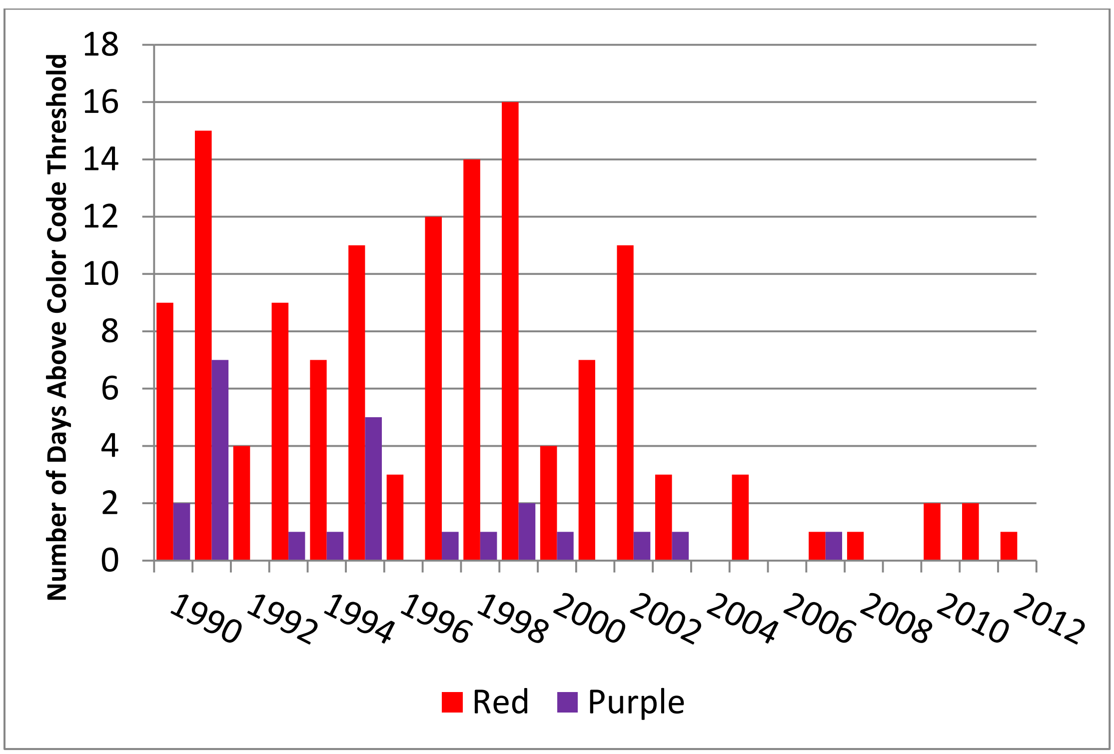

Figure 4. Number of days above threshold for Code Red (96 ppbv) and Code Purple (116 ppbv) in Delaware.

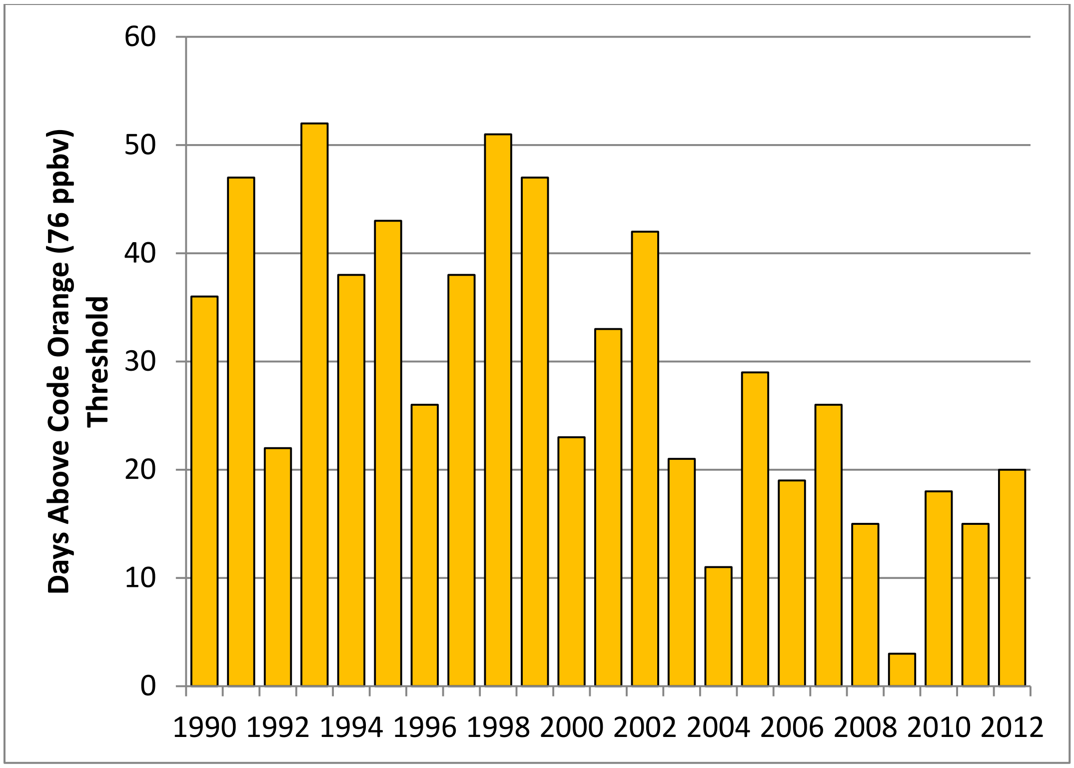

Figure 5. As in Figure 4 but for threshold of Code Orange (76 ppbv)

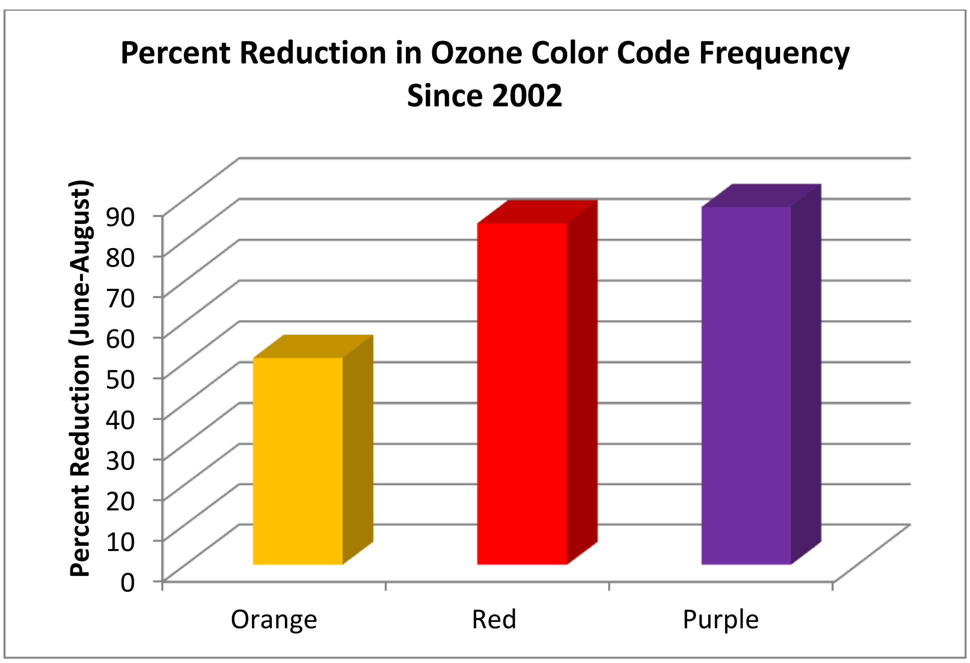

Figure 6. Percent reduction in frequency of selected AQI color codes since 2002 in Delaware.

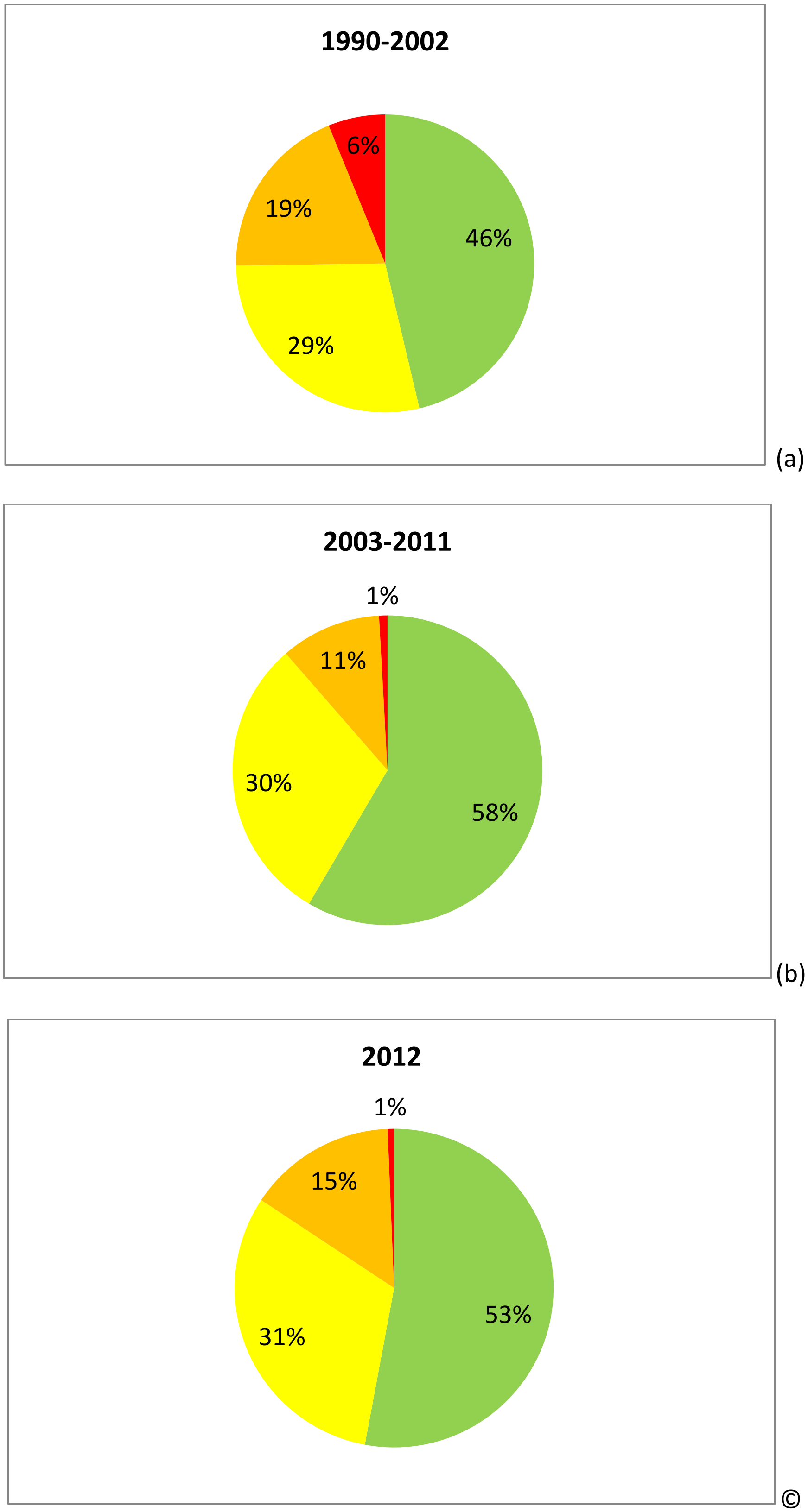

Figure 7. Frequency of AQI color codes for O_3 in Delaware for (a) 1990-2002, (b) 2003-2011 and (c) 2012.

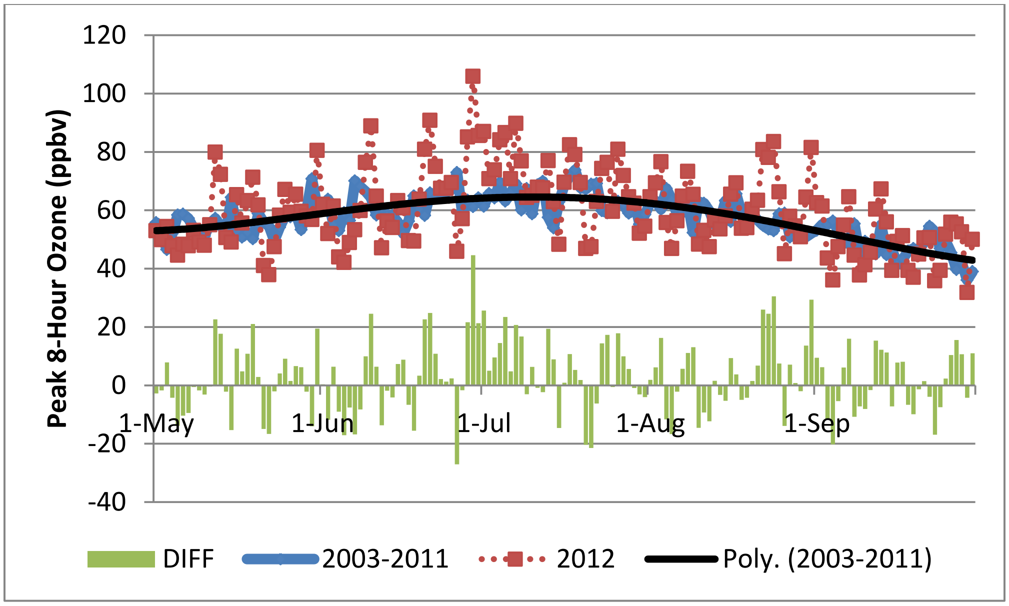

Figure 8. Time series of peak daily 8-hour o in Delaware for 2012 (red dotted line), compared to 2003-2011 (blue line). Best cubic fit to 2003-2011 observations given by black line and difference (DIFF) for 2012 compared to 2003-2011 by green columns.

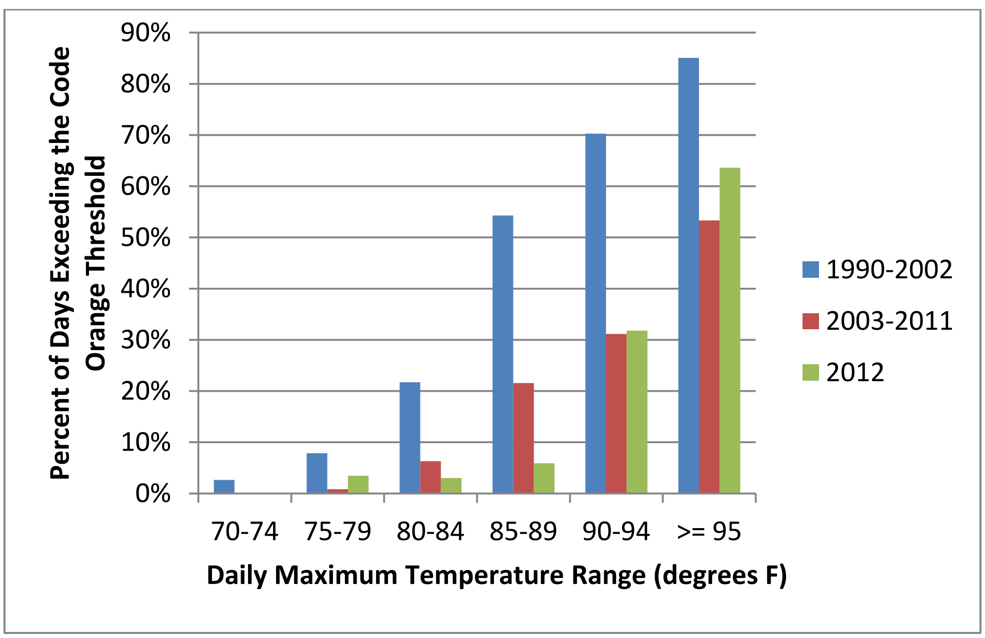

Figure 9. Frequency of days exceeding the Code Orange threshold for bins of maximum surface temperature at Wilmington, DE (ILG).

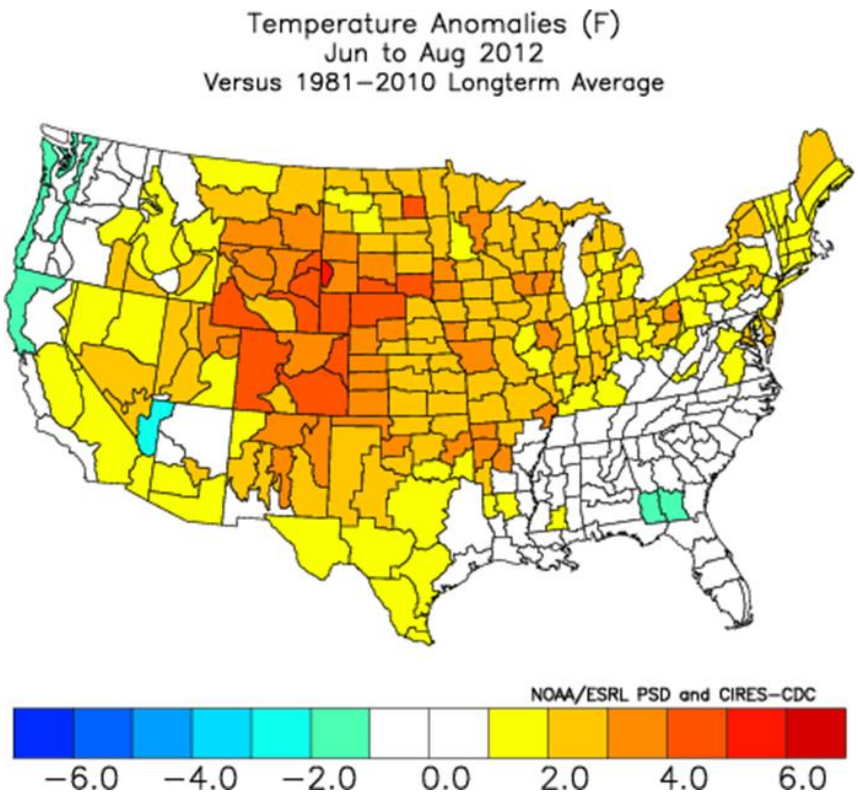

Figure 10. Temperature anomalies (in degrees F) for June-August 2012 compared to 1981-2012 mean. Figure courtesy of NOAA/ESRL.

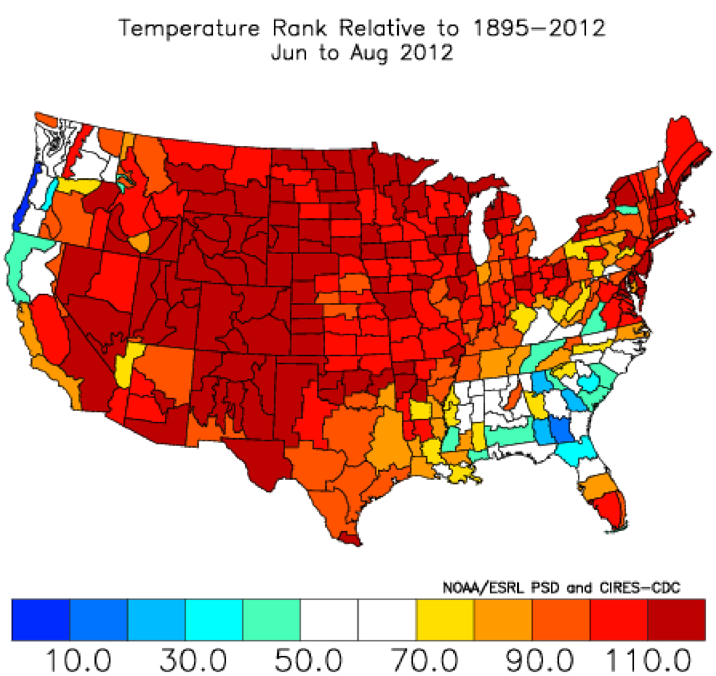

Figure 11. Temperature rank for June-August 2012 compared to period 1895-2012.

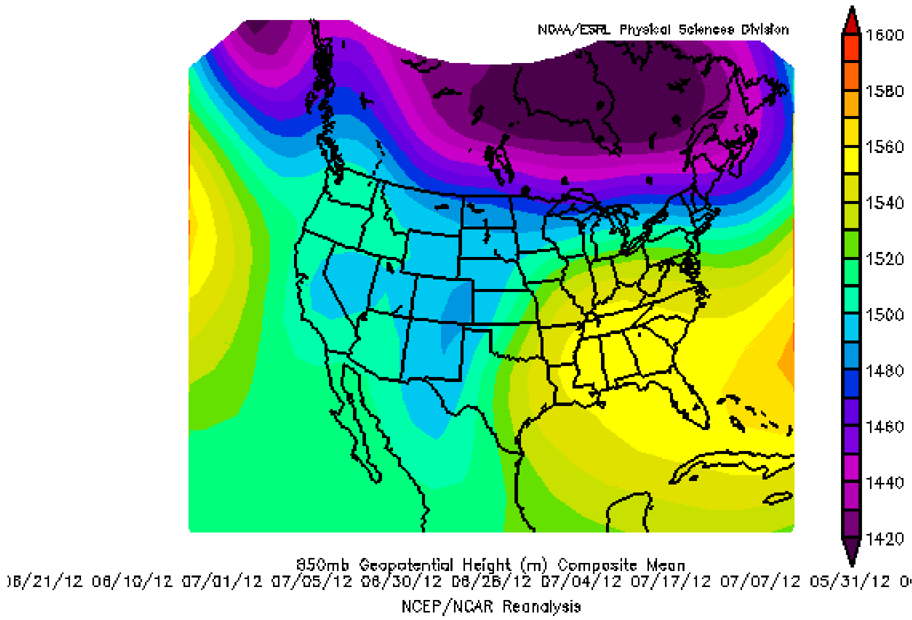

Figure 12. Average 850 mb geopotential height for 15 days in 2012 with highest O_3 in Delaware.

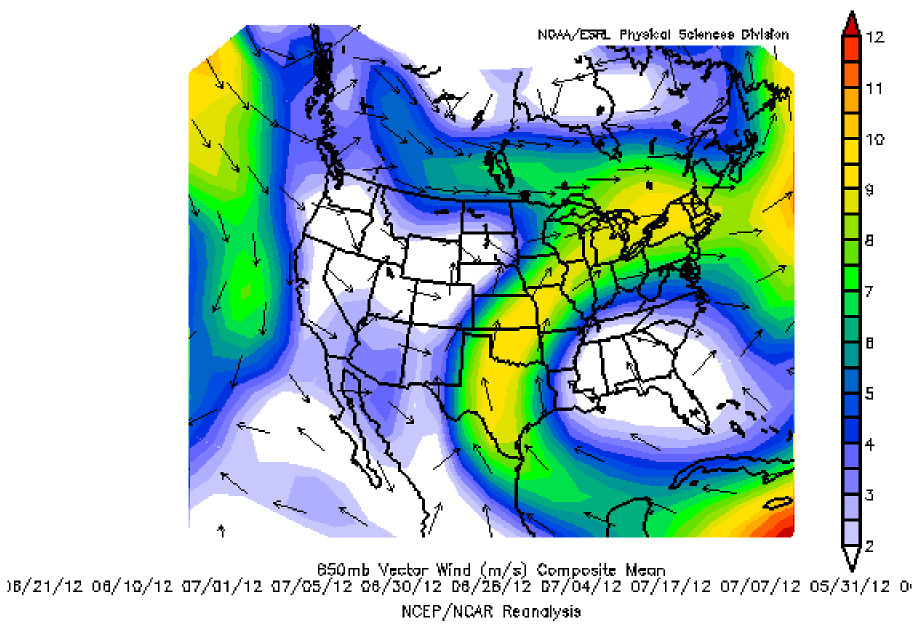

Figure 13. As in Figure 12, but for 850 mb wind velocity.

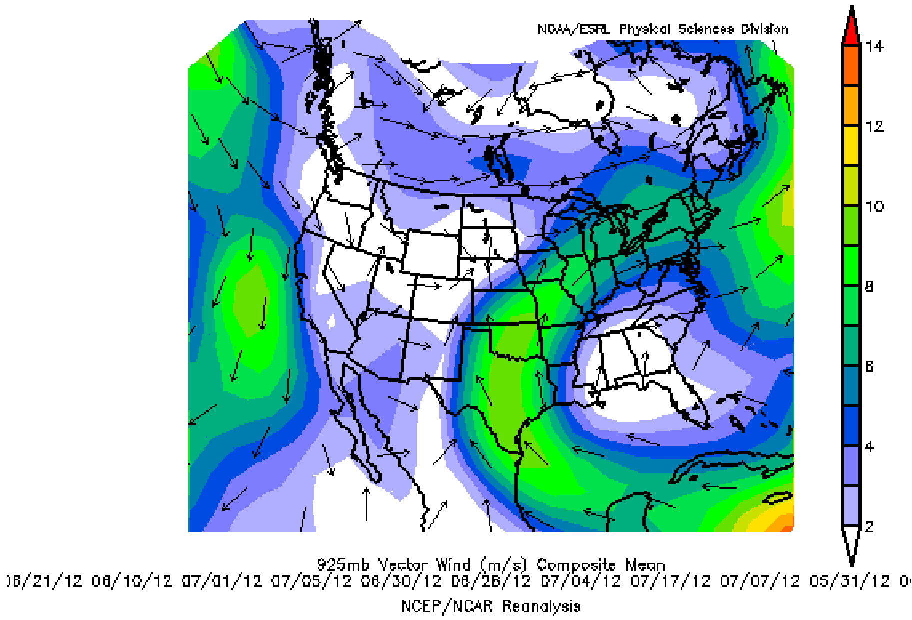

Figure 14. As in Figure 12 but for 925 mb wind vectors.

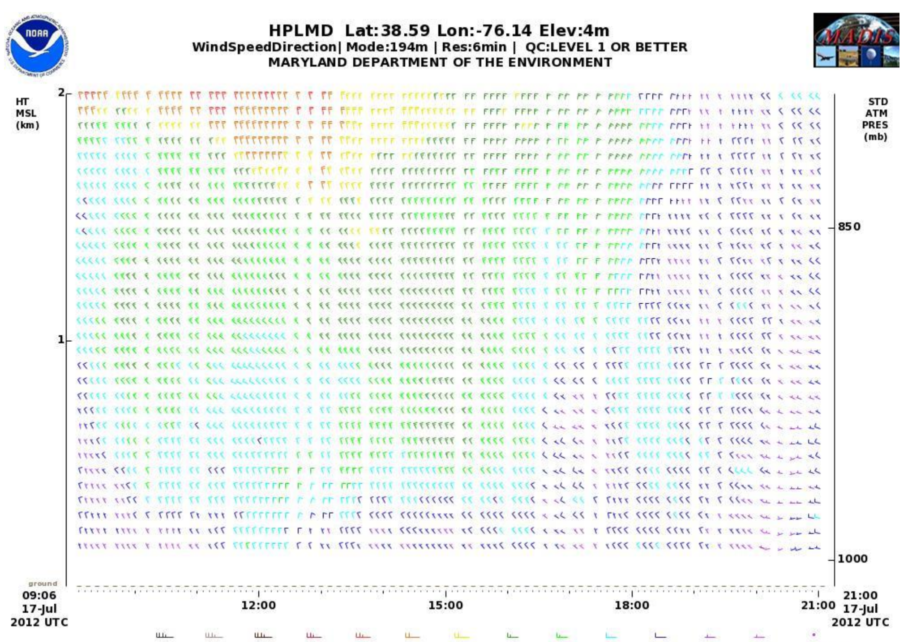

Figure 15. Wind profiler observations for Horn Point, MD for the period 0900-2100 UTC on July 17, 2012. Figure courtesy of the Maryland Department of the Environment.

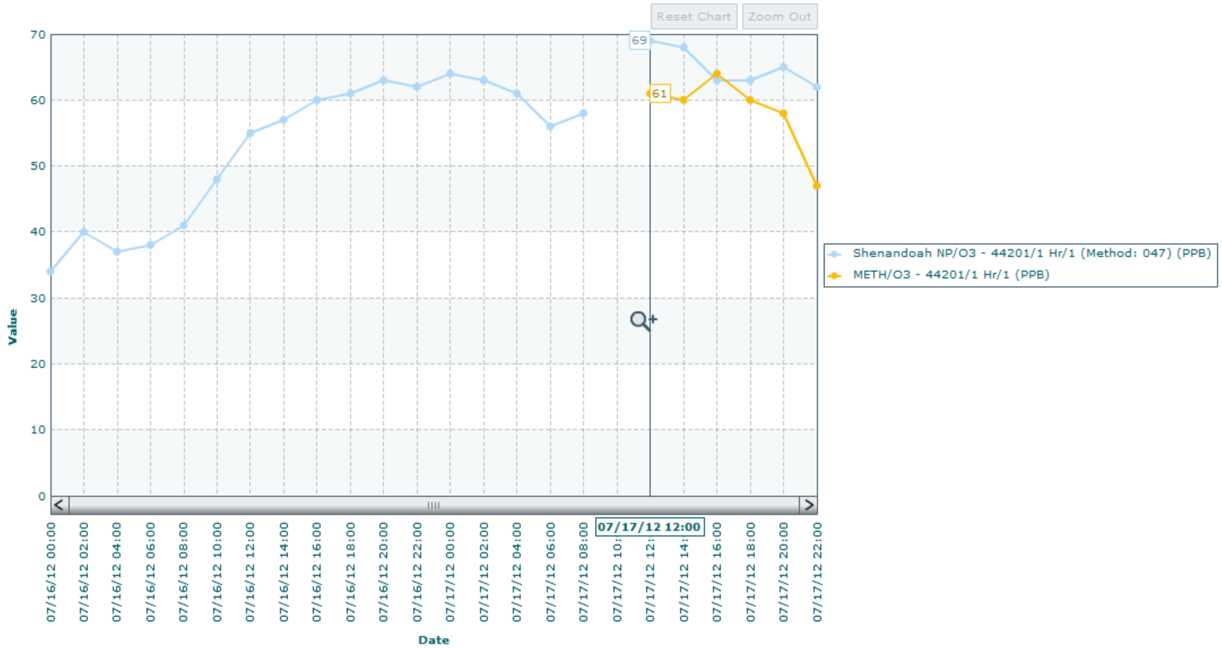

Figure 16. Hourly O_3 concentrations for Shenandoah National Park, VA and Methodist Hill, PA for July 16-17, 2012. Figure courtesy of AirNowTech.

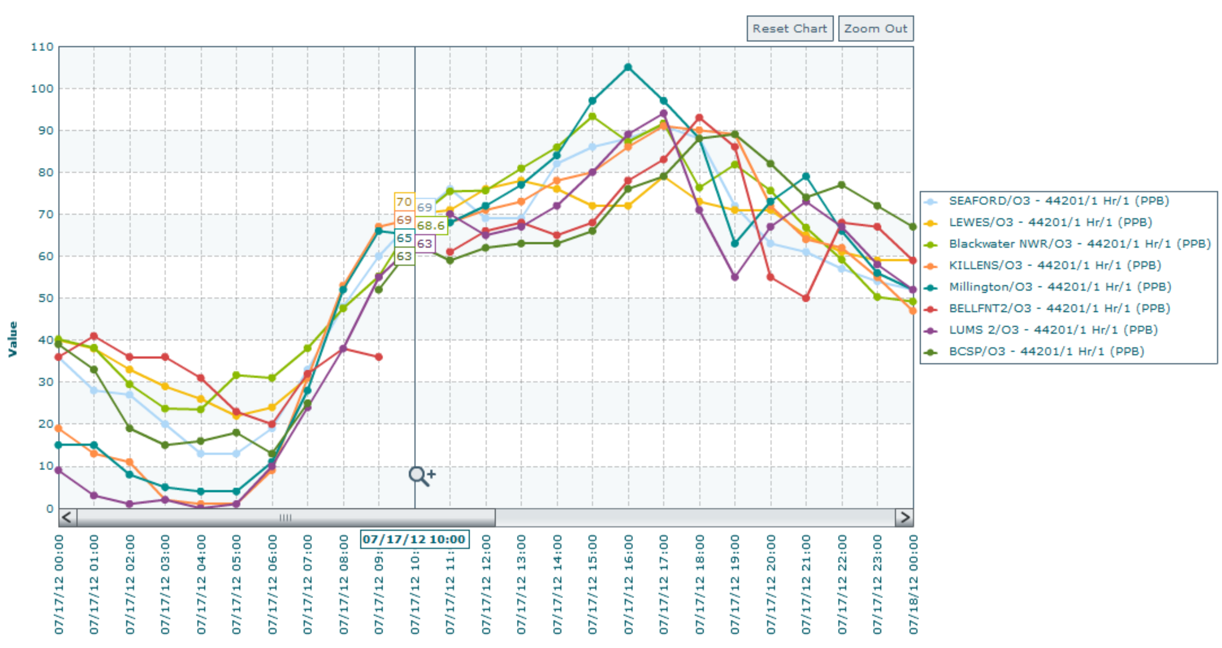

Figure 17. As in Figure 16 but for all O_3 monitors on the Delmarva Peninsula.

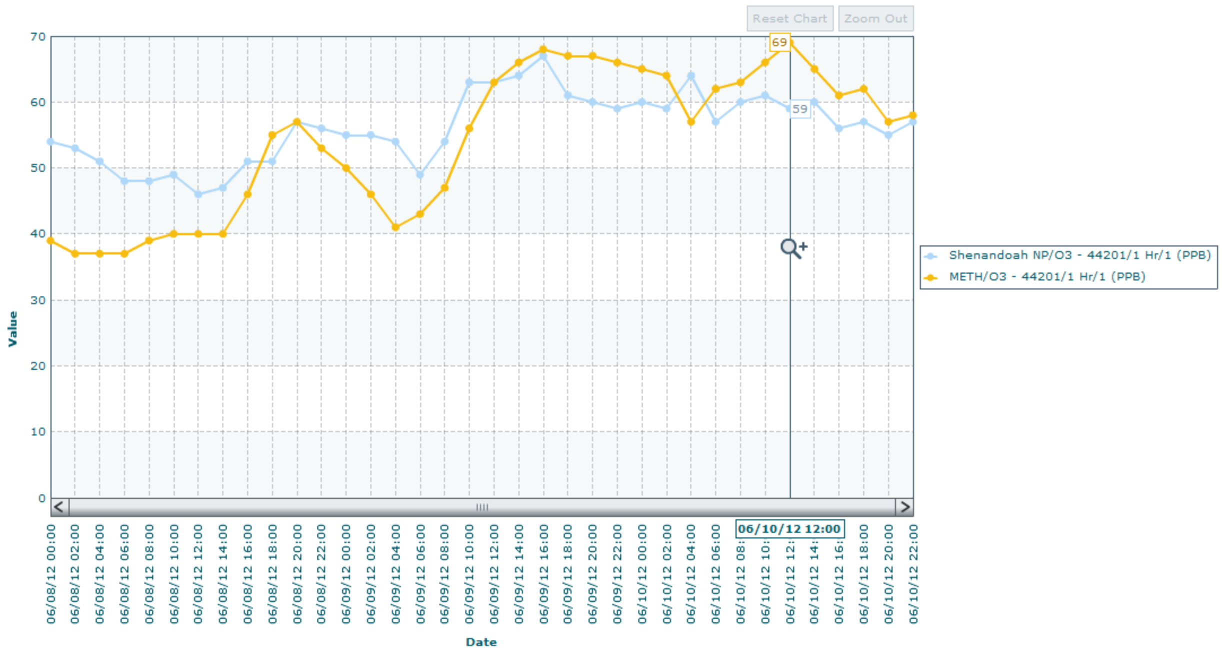

Figure 18. As in Figure 16 but for June 9-10, 2012.

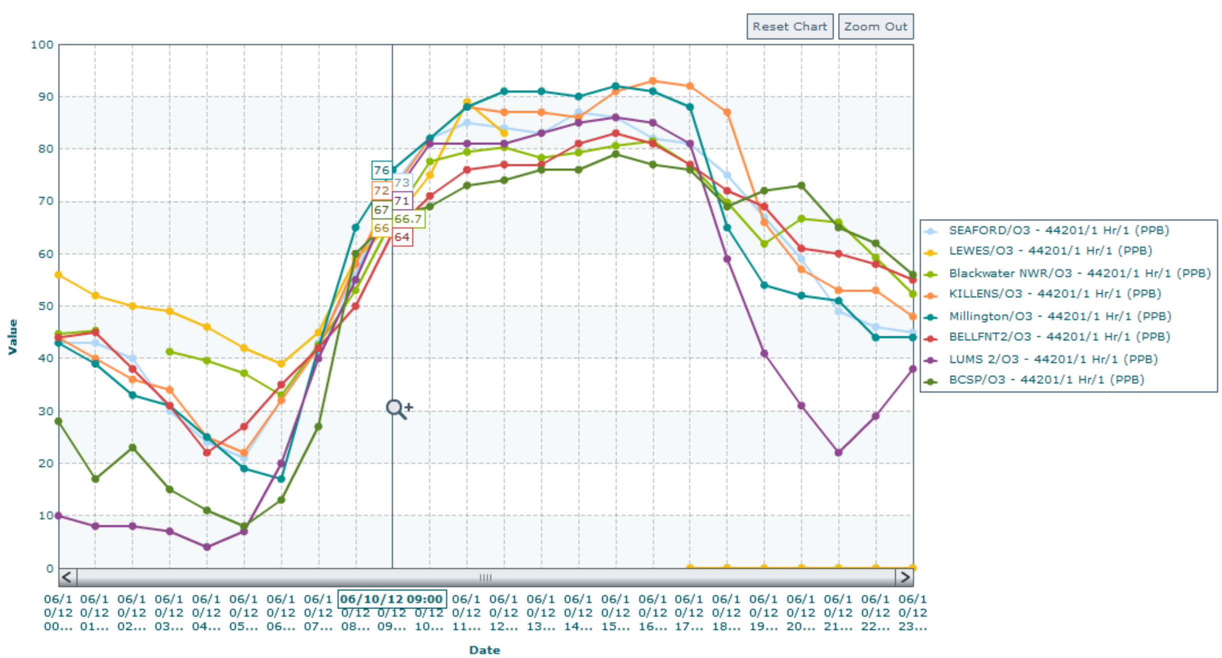

Figure 19. As in Figure 17, but for June 10, 2012.

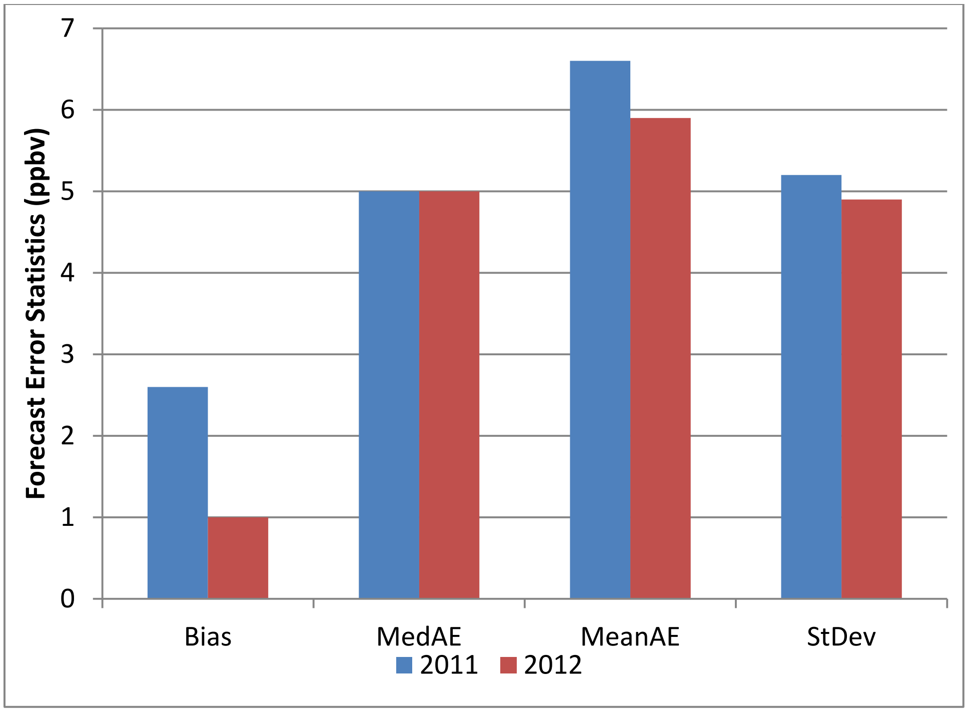

Figure 20. Forecast error statistics for Delaware for 2011 and 2012. "MedAE" refers to median absolute error and "MeanAE" to mean absolute error. "StDev" is standard deviation of the mean error.

Basic Ozone Forecast Statistics Delaware 2012 (all measures in ppbv)			
	Bias	**Median Absolute Error**	**Mean Absolute Error and Standard Deviation**
Forecast	+1.0	5.0	5.9 ± 4.9
NOAA	+2.0	6.8	7.9 ± 6.1
SUNY-Albany	+2.2	5.0	6.3 ± 5.3
BAMS	-0.3	7.0	7.9 ± 6.6
Persistence	+0.0	10.0	11.3 ± 8.1
SUNY-BAMS	-0.3	5.0	6.1 ± 5.4
NOAA-BAMS	+1.4	5.7	6.6 ± 5.1
NOAA-SUNY-BAMS	+0.8	5.2	6.0 ± 4.7
NOAA-SUNY	+1.5	5.4	6.5 ± 5.1

Table 1. Selected error statistics for Delaware in 2012 for the public forecast (Forecast), and several numerical O_3 models (NOAA, SUNY-Albany and Baron Advanced Meteorological Service (BAMS)) as well as combinations of models and the Persistence forecast.

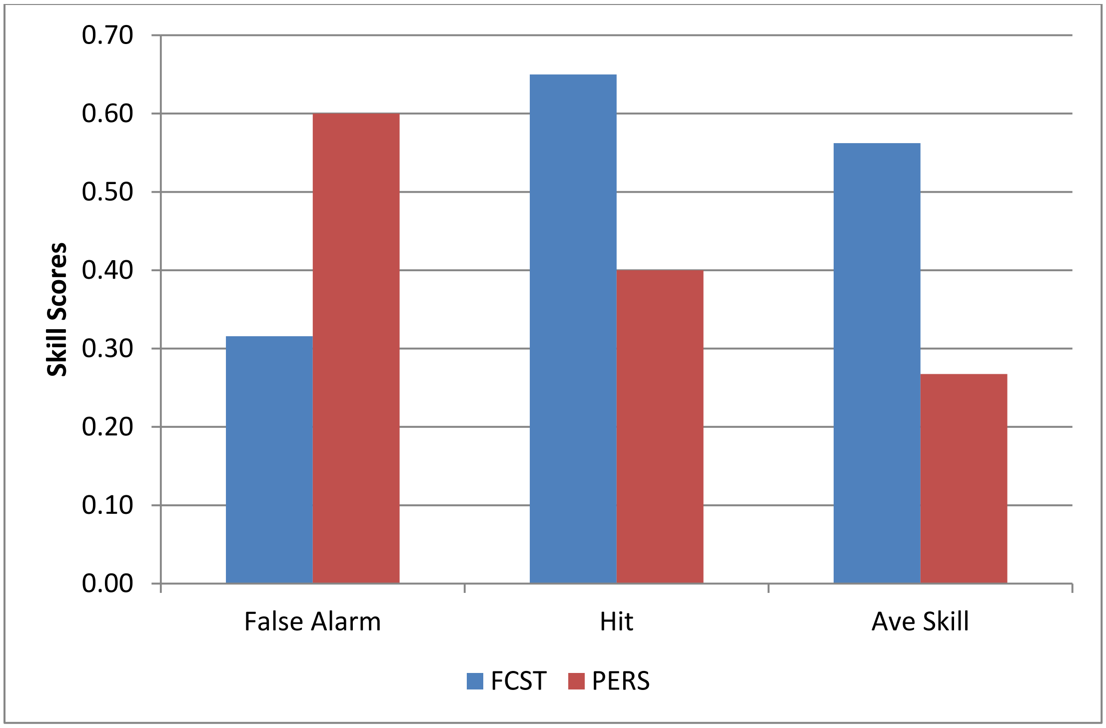

Figure 21. Skill scores for the public Delaware forecast compared to the reference persistence forecast (PRES) for 2012. Ave Skill refers to the mean of three skill scores (Threat, Heidke and Peirce) explained in more detail in [Appendix A](#).

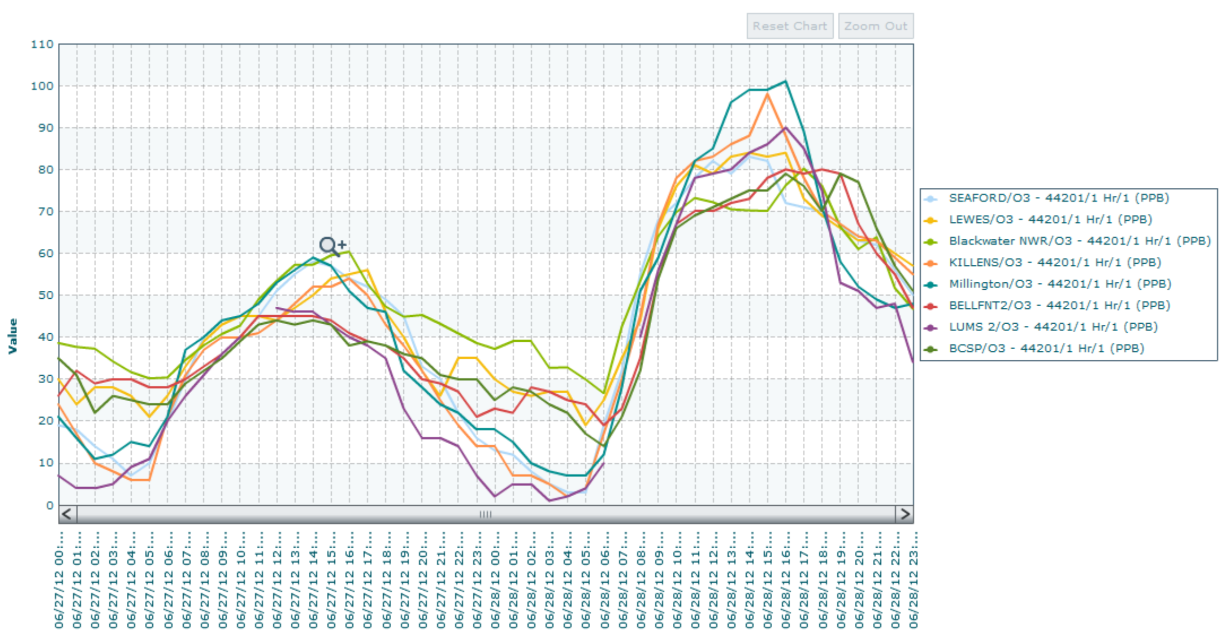

Figure 22. As in Figure 17 but for June 27-28, 2012.

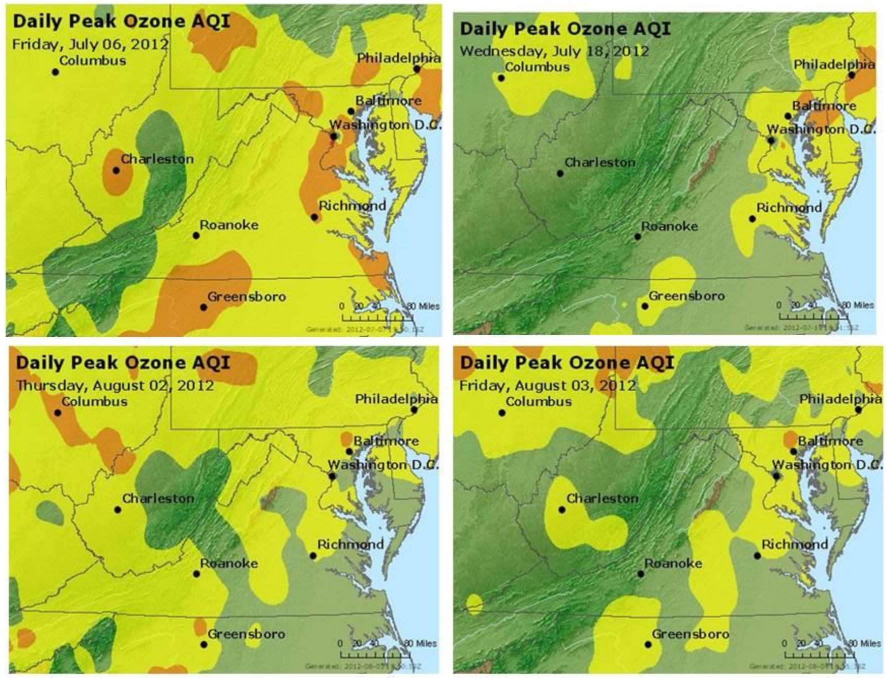

Figure 23. Daily peak O_3 AQI for the mid-Atlantic region for four cases of false alarm Code Orange forecasts in Delaware. Figure courtesy of EPA AirNow.

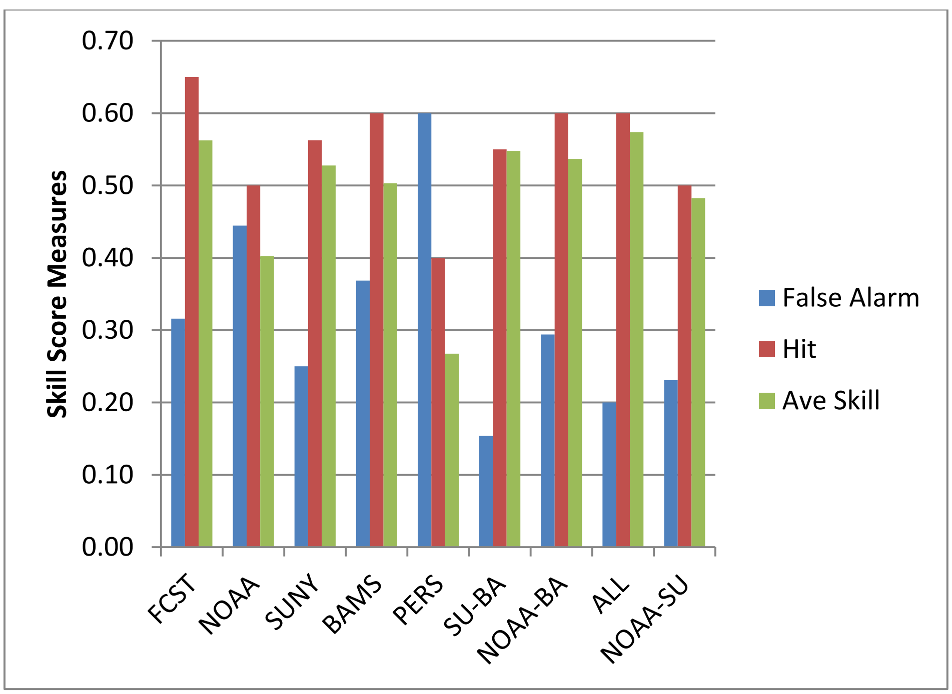

Figure 24. Skill score measures as in Figure 21 but for all numerical forecast models and their ensembles.

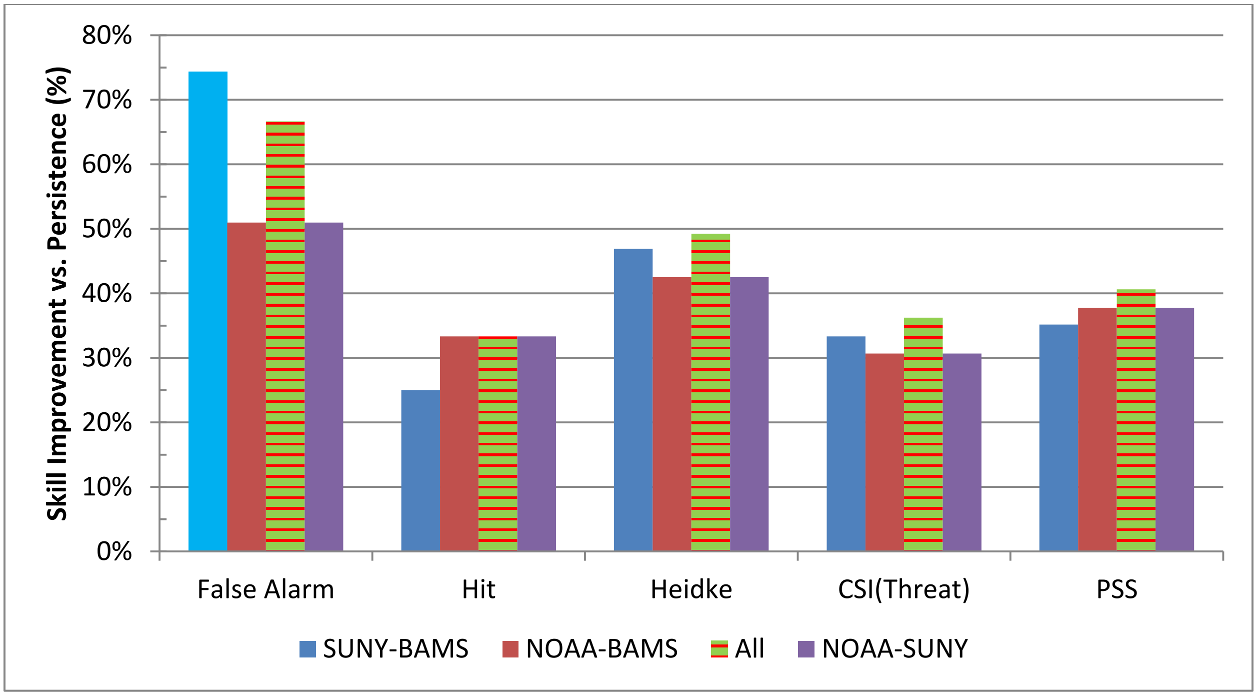

Figure 25. As in Figure 21 but for four forecast model ensembles and giving three skill scores measures individually. See Appendix A for more details on skill score measures.

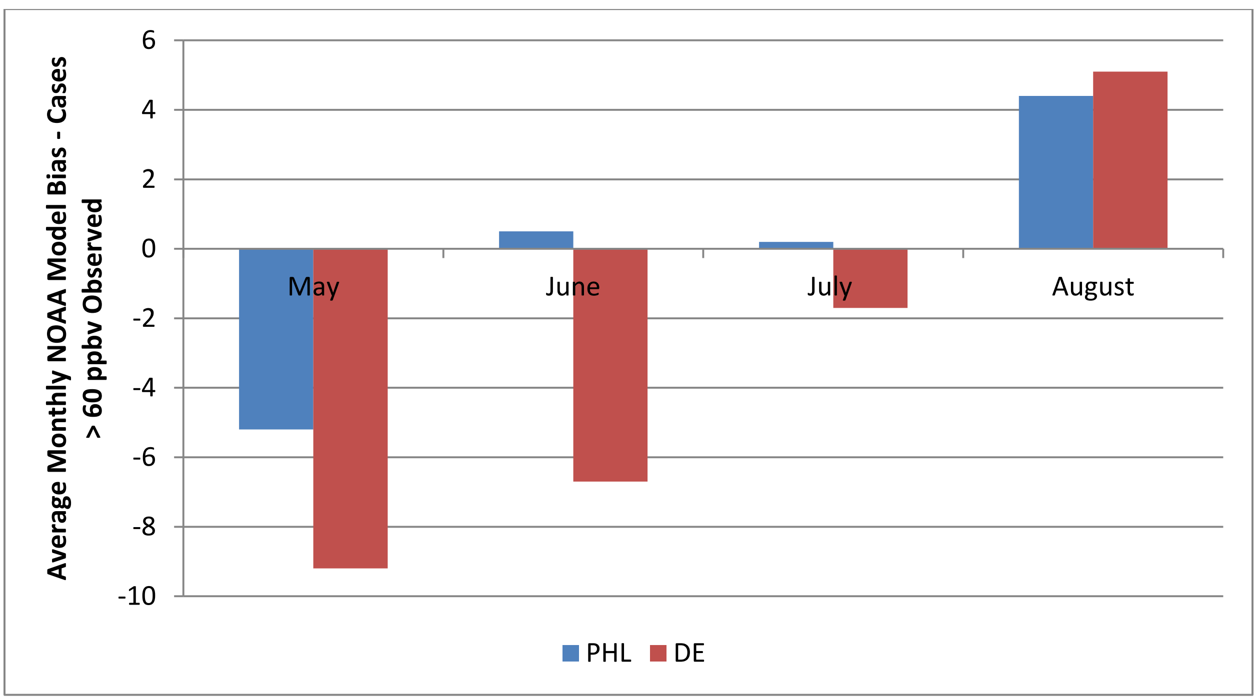

Figure 26. NOAA forecast model bias for forecasts of peak O_3 in Delaware and the Philadelphia metropolitan area for the months of May-August 2012. Only days with peak observed O_3 in the Code Yellow or higher range considered.

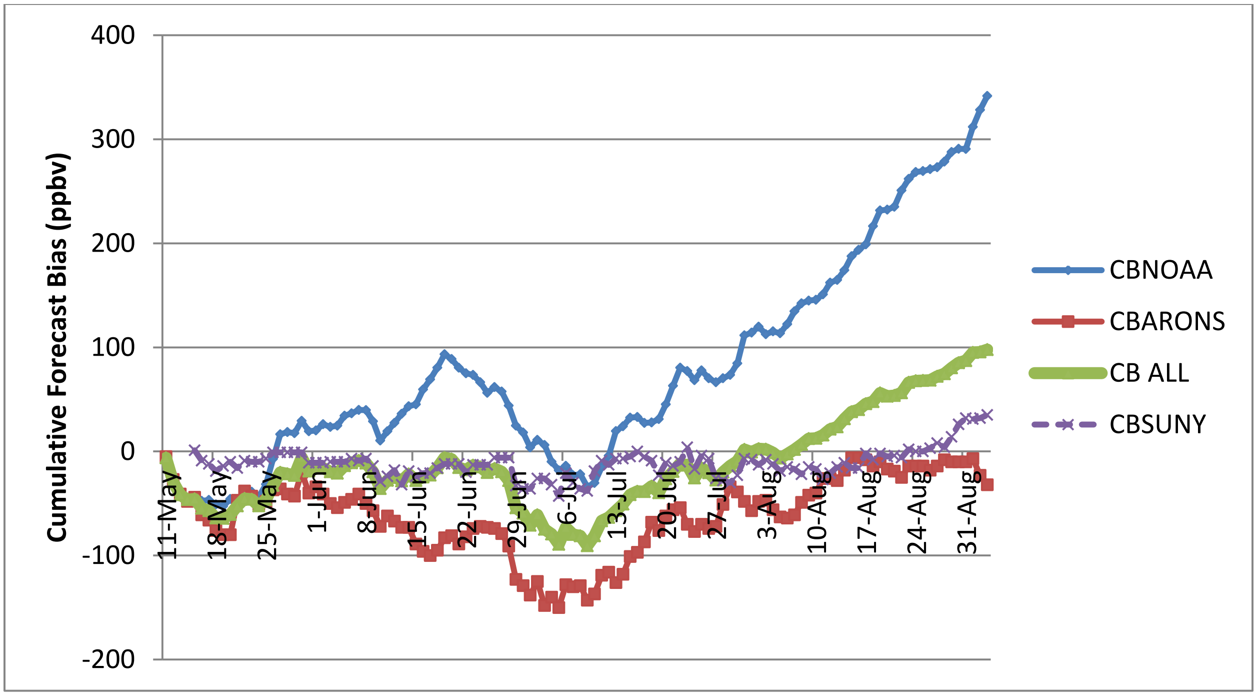

Figure 27. Cumulative (running sum) forecast model bias for May-August, 2012 – includes NOAA model (blue line), BAMS model (red line), SUNY model (dashed purple line) and mean (green line).

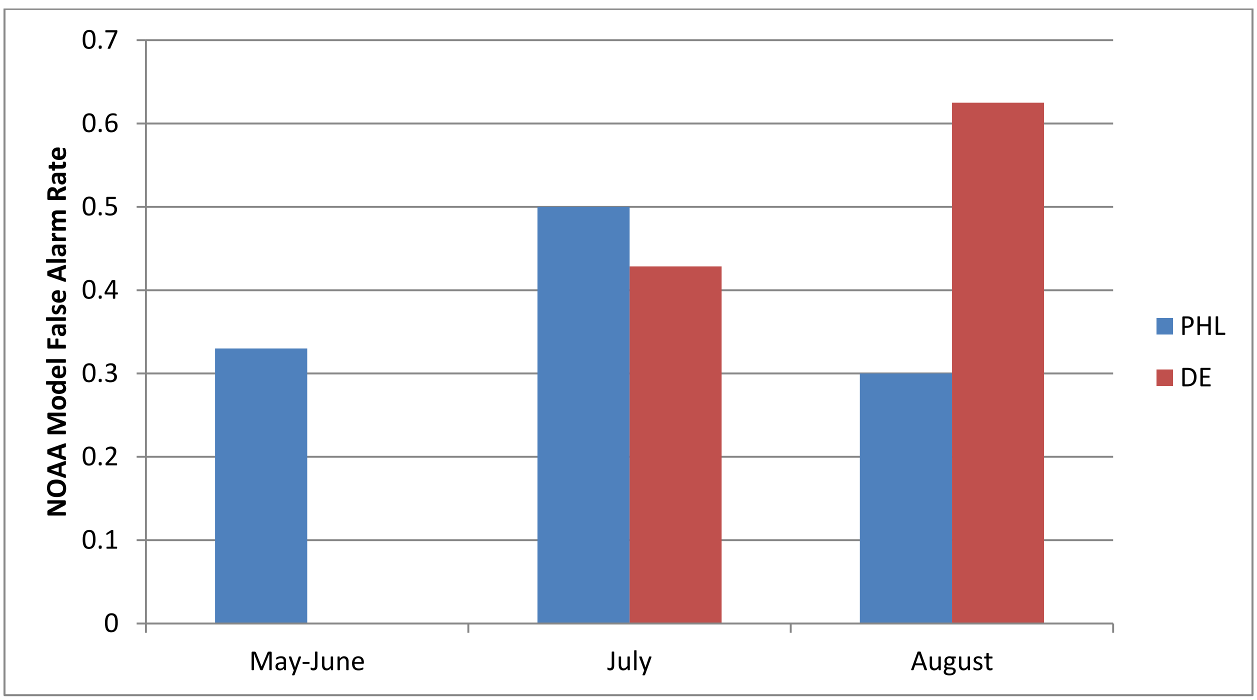

Figure 28. False alarm rate of the NOAA numerical O_3 forecast model, by month, for 2012 in Delaware and the Philadelphia metropolitan area.

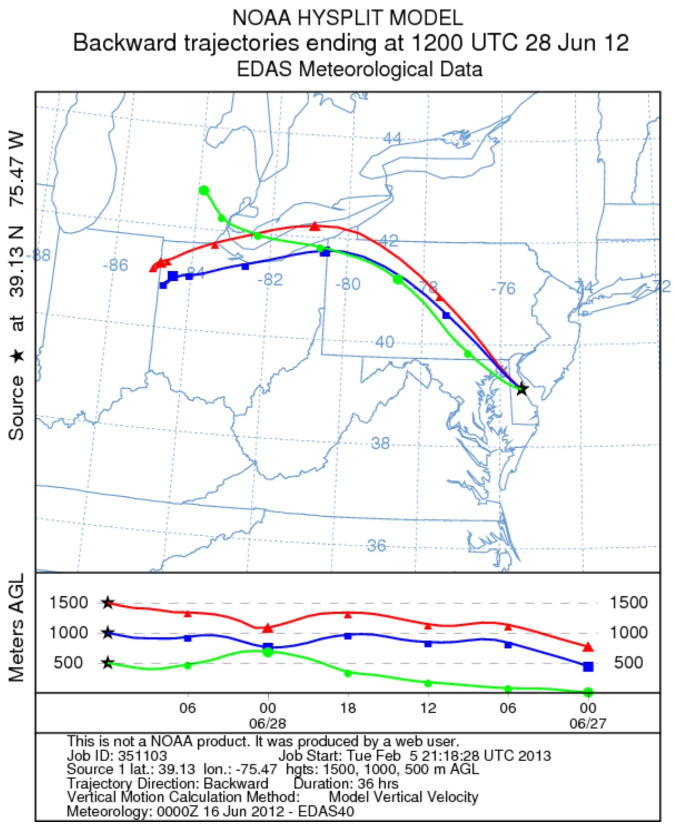

Figure 29. 36-hour HYSPLIT back trajectory terminating at 500 (green), 1000 (blue) and 1500 meters AGL (red) at Dover, Delaware at 1200 UTC on June 28, 2012. Trajectories based on forecast model analysis fields.

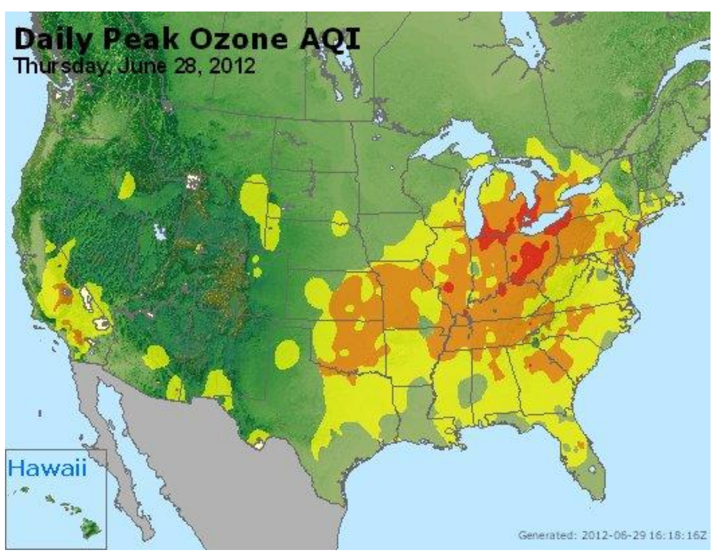

Figure 30. Daily peak O_3 AQI for June 28, 2012. Figure courtesy of EPA AirNow.

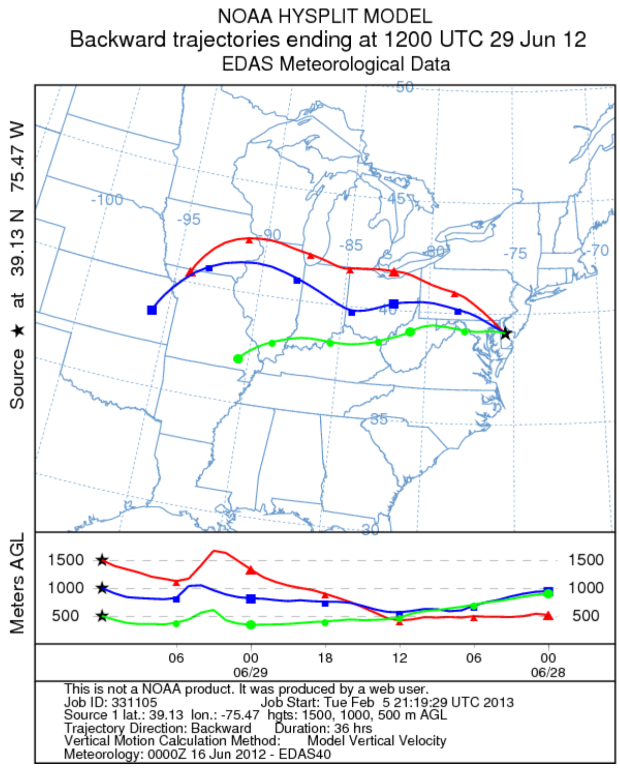

Figure 31. As in Figure 29 but for 1200 UTC on June 29, 2012.

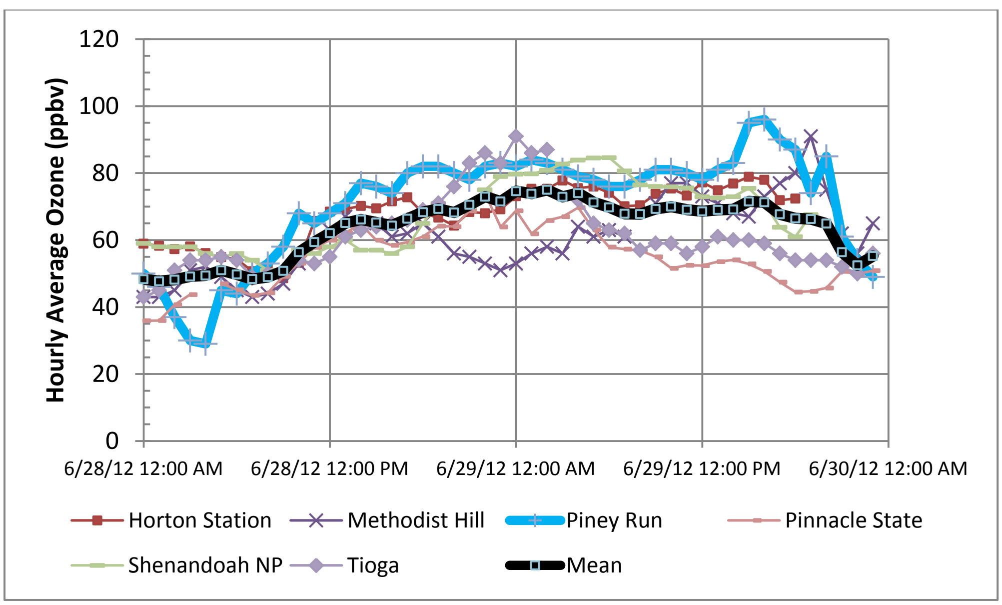

Figure 32. Hourly O_3 concentrations at 6 selected rural monitors west and northwest (upwind of Delaware, see Figures 29 and 31, for June 28-29, 2012.

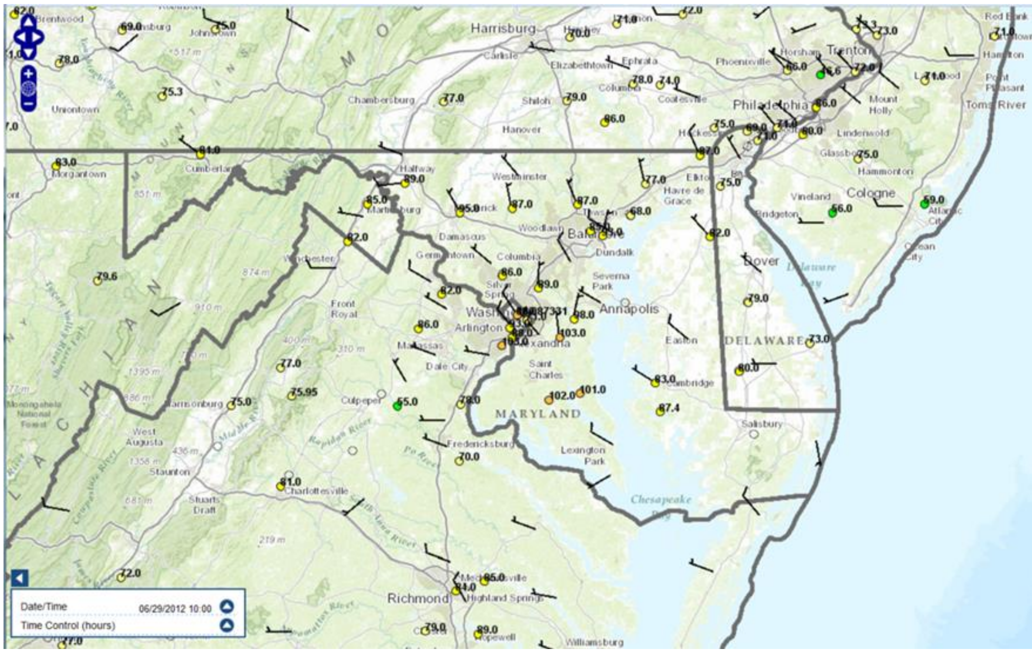

Figure 33. O_3 concentrations and wind barbs (mph) for the mid-Atlantic at 1000 EST on June 29, 2012.

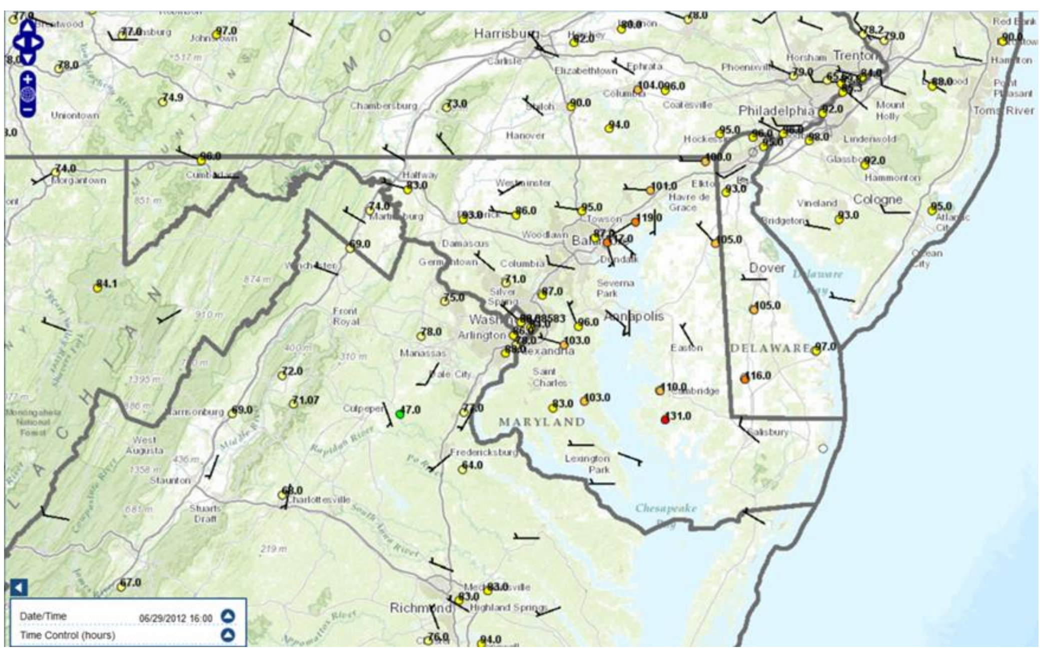

Figure 34. As in Figure 33 but for 1600 EST on June 29, 2012.

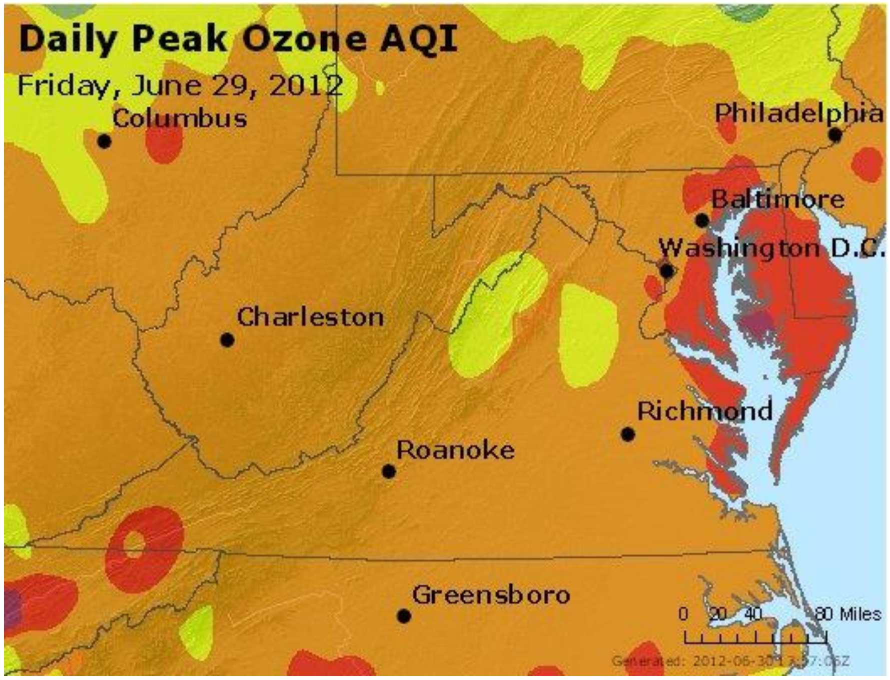

Figure 35. Daily peak O_3 AQI for the mid-Atlantic on June 29, 2012. Figure courtesy of EPA AirNow.

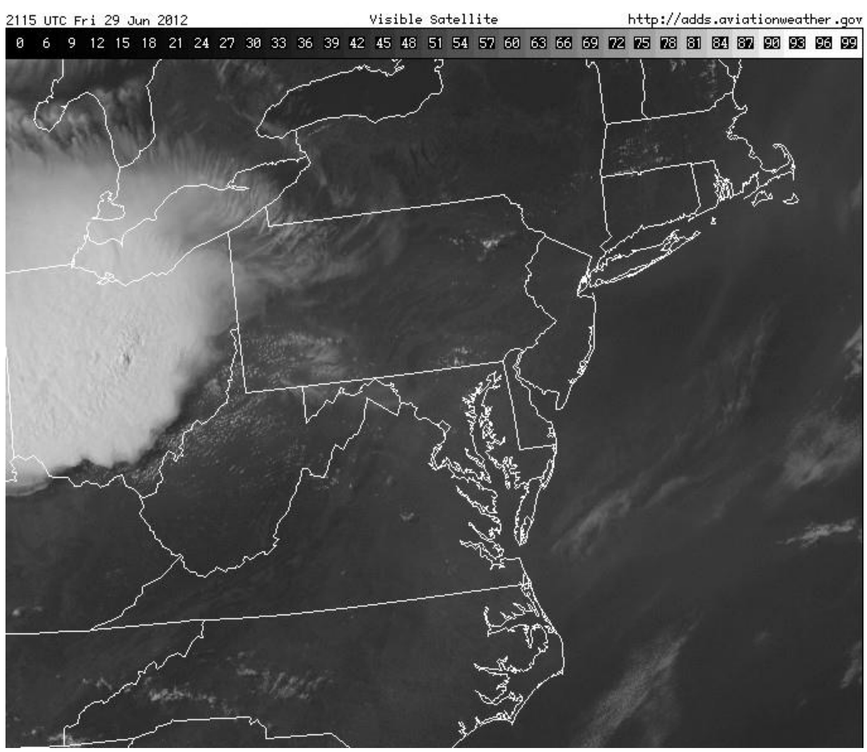

Figure 36. GOES visible image for 2115 UTC on June 29, 2012.

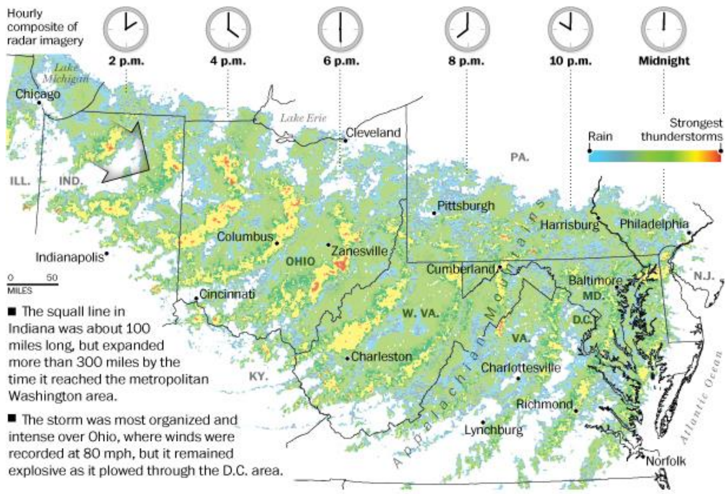

Figure 37. Path of the derecho on June 29-30, 2012. Figure courtesy of the Washington Post, http://www.washingtonpost.com/national/health-science/derecho-spanish-for-straight/2012/06/30/gJQArYBlEW_graphic.html

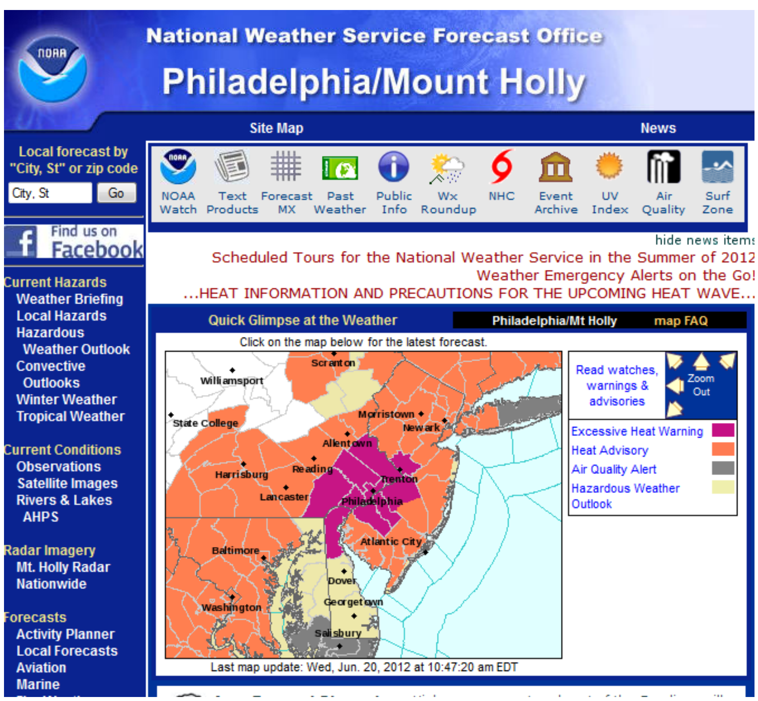

Figure 38. Home page of the Mount Holly NWS forecast office on June 20, 2012. Grey areas refer to air quality alerts. In reality, all municipalities along the I-95 Corridor from Richmond, VA to New York City had issued air quality alerts.

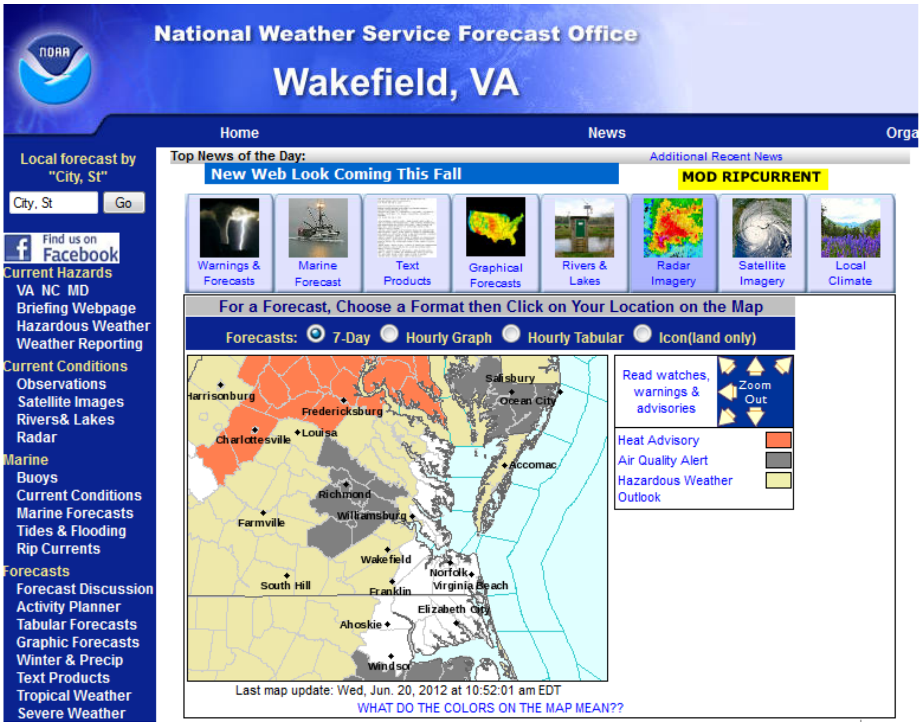

Figure 39. As in Figure 38 but for the Wakefield, VA NWS forecast office.

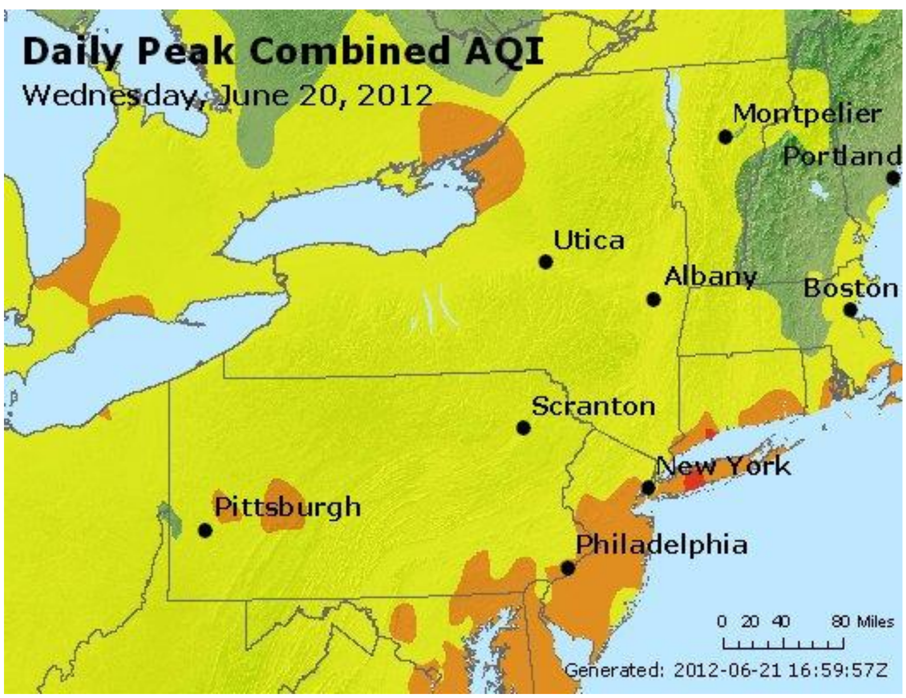

Figure 40. Peak AQI for June 20, 2012. On this day, O_3 was the leading AQI pollutant in all Code Orange regions.

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 Red O_3 concentrations are both observed and forecast (“hit”), then one unit is added to “a”. A “false alarm”, Code Red forecast but not observed, is added to “c”.

Basic Skill Measures:

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

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.

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.

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

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

Miss Rate = 1 – H

Correct null forecasts:

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

Accuracy:

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

Other 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.

Skill Score (SS_{ref}) = $\left(\frac{A - A_{ref}}{A_{perf} - A_{ref}} \right) * 100\%$ = 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.

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**.

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:

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

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, 467pp., 1995.