4.11

FINE PARTICULATE MATTER ($PM_{2.5}$) FORECASTING IN THE PHILADELPHIA METROPOLITAN AREA: CURRENT PRACTICE AND BENCHMARK SKILL

William F. Ryan
The Pennsylvania State University, University Park, Pennsylvania

Charles A. Piety
University of Maryland, College Park, Maryland

1. INTRODUCTION

Routine forecasts of fine particulate matter ($PM_{2.5}$) concentrations began in the mid-Atlantic region in October, 2003. This paper presents initial results from forecasts in the Philadelphia (PHL) area and summarizes the current challenges in this novel forecast area.

2. $PM_{2.5}$ BACKGROUND

$PM_{2.5}$ consists of particles with an aerodynamic diameter less than 2.5 micrometers (μm). Sources of $PM_{2.5}$ include emissions from any type of combustion, primarily motor vehicles, power plants, industrial processes, residential wood smoke and forest fires. Some particles are volatilized and condensed combustion products (primary $PM_{2.5}$) and the remainder is transformed in the atmosphere from a variety of sources (secondary $PM_{2.5}$). Particles of this size are small enough to be breathed deeply into the lungs and pose a potential health risk. Long term exposure to high levels of $PM_{2.5}$ is associated with reduced lung function and chronic bronchitis, and short term exposure can aggravate lung disease, asthma and increase susceptibility to respiratory infection. Because of the adverse affects of $PM_{2.5}$ on human health, it is designated by the EPA as a "criteria" pollutant for which a National Ambient Air Quality Standard (NAAQS) has been issued.

The current primary NAAQS for $PM_{2.5}$ consists of a yearly average threshold (15.5 μgm^{-3}) as well as a daily (24 hour average) threshold (65.5 μgm^{-3}). Current forecast programs follow a color coded scheme based on the NAAQS. A breakdown of the concentration thresholds is given in Table 1.

Unlike other criteria pollutants, such as O_3 and lead, extensive $PM_{2.5}$ monitoring is of recent vintage. Widespread monitoring in the mid-Atlantic region began only in 1999. As a result, the historical database on which to formulate $PM_{2.5}$ forecasting guidance is thin and poses a significant challenge to the development of statistical forecast algorithms that have proven useful in O_3 forecasting.

The officially recognized method (Federal Reference Method or FRM) for measuring $PM_{2.5}$ is a gravimetric method in which an air sample is drawn at a constant rate through an impactor (particle size separator) and then collected on a filter. The filter is removed, equilibrated and weighed. The amount of particle mass is determined by the weight of the filter. Weighing the filter is a complex task that can only be accomplished in a specially equipped laboratory. As a result, there is a significant lag, varying from weeks to months, between sample collection and data reporting. In practice, this means that FRM $PM_{2.5}$ measurements are not available for use by operational forecasters.

In addition to the lag in data reporting, the observational network is limited in observation frequency. Most of the FRM measurements are made only every third day with a subset (~15-20%) of monitors reporting daily (Figure 1). The irregularity in observation frequency poses a serious problem for the development of forecast models and forecast verification as discussed below.

Because the FRM for $PM_{2.5}$ measurements is clearly not appropriate for forecast support, forecasters must rely on data from continuous sampling methods. The network of continuous $PM_{2.5}$ monitors in the mid-Atlantic is limited (< 10 in the Washington DC-Philadelphia Corridor) although there are plans to augment this network (Figure 2). Even within the group of continuous monitors, a number of different measurement techniques are utilized that are not necessarily equivalent to each other or the FRM. The most common continuous method used in the mid-Atlantic is the Tapered Element Oscillating Microbalance (TEOM). TEOMs

* *Corresponding author address*: William F. Ryan, Pennsylvania State University, Department of Meteorology, 409 Walker Building, University Park, PA, 16802, email: wfr1@psu.edu.

measures mass by observing changes in the vibrations of a glass tube located downstream of a size selective inlet. In order to reduce the delay required to remove moisture from the sample before measuring, the TEOM's are heated to well above usual ambient dew point temperature (30 or 50°C). However, some volatilization of $PM_{2.5}$ occurs in the heated TEOM chamber leading to an underestimation of concentrations in the cool season. Additional modifications can be made to TEOMs (e.g. FDMS) but they tend to introduce over predictions during the warm season. The challenge for air quality forecasting is to know the extent to which continuous measurements can be trusted both to verify forecasts and, in near-real time, to determine current and persistence effects.

To summarize, several fundamental questions are posed by the analysis of $PM_{2.5}$ measurements. First, is FRM data sufficiently accurate and consistent to support forecast verification and the development of statistical forecast algorithms? Second, can real-time continuous $PM_{2.5}$ measurements be utilized to inform operational forecasts?

3. OPERATIONAL FORECASTING

In the Philadelphia area, forecasts are supported by the Delaware Valley Regional Planning Commission (DVRPC) and are issued weekdays ~ 1800 UTC valid the following day. On Friday, a forecast for Saturday along with a weekend outlook is issued. During the summer months, when $PM_{2.5}$ concentrations are highest, daily forecasts are issued. As with O_3, forecasts are issued to the public with a color code although no provision was made for voluntary emissions reduction plans ("Action Days") for $PM_{2.5}$ forecasts in the Unhealthy for Sensitive Groups (Code Orange) or Unhealthy (Code Red) range (see, Table 1). Forecasts are posted to the AirNOW program, where they are picked up by USAToday, as well as the DVRPC and Pennsylvania Department of Environmental Protection (PADEP) web pages.

Because of the complex manner in which $PM_{2.5}$ is formed, including gaseous, aqueous and heterogeneous chemistry, there are currently no numerical prediction models in operational use – although these models are in the process of development. As a result, forecast methods had to be developed to support the new $PM_{2.5}$ forecasting program. In the mid-Atlantic, statistically based forecast guidance was developed through a grant provided by the Mid-Atlantic Regional Air Managers Association (MARAMA). This guidance, using Classification and Regression Tree (CART) techniques was developed by Systems Applications International (SAI).

The CART technique works by splitting cases into discrete clusters, or nodes, of closely distributed $PM_{2.5}$ concentrations based on similarity with respect to a set of key predictors. At each split, or decision point, cases are partitioned based on a threshold value for a single predictor (e.g., T_{max} > 90°F). CART then further subdivides the cases based on other predictors before reaching an optimal distribution. The final result is a group of terminal nodes, or branches, each containing similar cases. Each new case (forecast) is then assigned to a terminal node based on its response to the various decision point thresholds and is assigned the dominant color code of the cases in that node.

In the PHL CART tool, developed using 1999-2002 data, two of the ~ 30 terminal nodes contains high $PM_{2.5}$ cases. The summer season high $PM_{2.5}$ node contains cases with $T_{max} \geq 33°C$ coupled with moderate or higher previous day $PM_{2.5}$ concentrations region wide (persistence), high relative humidity and moderate winds. It is worth noting that for high $PM_{2.5}$ concentrations to occur, like O_3, hot weather is necessary but not sufficient. Days with $T_{max} \geq 33°C$ account for nearly all high $PM_{2.5}$ cases but only 26% of these hot days reach the Code Orange threshold. The requirement of a moist air mass makes physical sense in that sulfate is a key component of $PM_{2.5}$ in the mid-Atlantic during the summer season and conversion of gaseous SO_2 to sulfate is enhanced in moist conditions. In the winter months, the high $PM_{2.5}$ terminal node is found in a set of cases with a strong low level inversion coupled with a stagnant, moist air mass.

In addition to statistical guidance, forecasts can be informed by knowledge of regional persistence. $PM_{2.5}$ has a relatively long life time in the atmosphere. Sulfate, for example, a major component of $PM_{2.5}$ in the eastern US, has a lifetime on the order of days to a week. Knowing current concentrations, coupled with forecast back trajectories (e.g., HYSPLIT), forecasters can adjust the statistical guidance forecast to account for regionally persistent $PM_{2.5}$ concentrations.

4. VERIFICATION OF FORECASTS

The initial question is whether the FRM data from the high density observations ("third day") is consistent with the observations from the lower density days ("1-2 day"). A cursory look at the data from the mid-Atlantic (Figure 3) shows that peak concentration data from the third day observations is consistently higher (Figure 4). This suggests that wider spatial coverage occasioned by more available monitors allows better resolution of the plume of highest $PM_{2.5}$. The usual test for differences between samples of the same phenomenon is the student-t test. However, that test assumes that the data from each sample is normally distributed. This does not appear to be the case. For even the mean $PM_{2.5}$ observations, the measure of skewness, or symmetry about the mean, for this sample is 1.12. This coefficient is considered significant because the ratio of skewness to the standard error of skewness (SES) is 16.0 (values in excess of 2 are typically considered significant). Komolgorov-Smirnov (K-S) 1-sample test, applied with the Lilliefors option, finds that neither mean nor maximum $PM_{2.5}$ are normally distributed and a non-parametric two-sample test is then applied. These tests make no assumptions about the distribution although they do entail other assumptions. The Komolgorov-Smirnov 2-sample or Mann-Whitney U-tests show that maximum $PM_{2.5}$ data from the third day sample are significantly different from the day 1-2 sample while mean $PM_{2.5}$ data are not significantly different.

When data are found not to be normally distributed, it is often useful to transform the data with the hope of approximating a normal distribution. The results from a natural log transform of mean and maximum $PM_{2.5}$ results in a more normal distribution. When disaggregated into third day and 1-2 day groups, and applying the K-S 1-sample test, the difference values for the third day data (both mean and maximum values) are normally distributed and the 1-2 day data, are normal or very close ($p = 0.06$ for maximum $PM_{2.5}$ and 0.04 for mean $PM_{2.5}$). Both sets of tests (non-parametric and t-test) were run with the log-transformed data. Mixed results were obtained. For the Komolgorov-Smirnov 2-sample test both maximum ($p = 0.142$) and mean ($p = 0.245$) $PM_{2.5}$ data from the third day and 1-2 day sample were found not to differ significantly. Thus, no firm conclusion can be made at this time which suggests that maximum data can be used only with caution.

The next important question is whether the continuous (TEOM) data can be relied on by forecasters in the operational environment. The limited number of monitors, coupled with the fact that the nearest monitor (Camden, NJ) was out of operation for the first two months of the forecast study, makes this issue of great interest. For the initial six months of this study, maximum $PM_{2.5}$ measured by regional TEOMs (monitors using other continuous measurement techniques were not used) was used by forecasters to approximate local peak $PM_{2.5}$. The regional data included a subset of the monitors in Figure 2, specifically Annandale, VA, Baltimore (Old Town), MD, Arendtsville, PA, New Brunswick, NJ, and, when available, Camden, NJ. TEOM data are reasonably consistent with FRM measurements and are sufficiently accurate for operational forecast concerns (Figure 5). The best fit line for the regional TEOM data, compared to forecast area FRM, is: $[PM_{2.5}]_{FRM} = 3.72 + 0.74*[PM_{2.5}]_{TEOM}$ with a bias of +0.2 μgm^{-3}. Using only the Camden TEOM provides poorer results (Figure 6) with an $r^2 = 0.53$ and a best fit of $[PM_{2.5}]_{FRM} = 6.84 + 0.63*[PM_{2.5}]_{Camden}$. As a result, forecasters are reasonably confident that using regional TEOM concentrations to determine persistence effects and for near real-time verification is a reasonable choice.

5. FORECAST RESULTS

Forecast results for next day (weekday) forecasts are given in Figure 7. The best fit line is $[PM_{2.5}]_{FRM} = 4.85 + 0.71*[PM_{2.5}]_{FCST}$ with a correlation coefficient of 0.73 and an r^2 of 0.53. The sample shows few cases in the high end of the distribution with a wide scatter of forecast performance in these cases. Clearly there are problems in forecaster understanding of these cases. Of the four forecasts in the Code Orange range (Table 1), none verified although two cases observed $PM_{2.5} \geq 30\ \mu gm^{-3}$. Of the four cases with observed $PM_{2.5}$ near the Code Orange threshold (~ 40 μgm^{-3}), however, forecasts were reasonably accurate ($\geq$ 34 μgm^{-3}). In terms of color coded forecasts, the correct color code was issued in 77% of all next day forecasts. Of the missed color codes, approximately half (47%) were cases in which observed $PM_{2.5}$ was in the 12-18 μgm^{-3} range – bracketing the moderate threshold. Thus, only 12% of the cases were badly misforecast with respect to color code. CART results, available only in color codes, were also reasonably accurate. For cases where CART results were available, 72% forecast the correct color code with consensus (expert) forecasts doing slightly better at 78%. In the only multi-day high $PM_{2.5}$ episode ($\geq$ 30 μgm^{-3}) that occurred during this period (October 8-10, 2003), CART forecasts did poorly, forecasting good air quality. Further analysis showed that the CART tool failed because the strong low level inversion, that the CART tool uses to split data into high $PM_{2.5}$ nodes, was capped below the 900 mb level that the CART uses.

Another way of looking at forecast performance is by analysis of forecast error. For the set of next day forecasts (n = 153), the forecast showed a mean absolute error (MAE) of 4.7 μgm^{-3} with a median error of 3.6 μgm^{-3} and an rms error of 7.0 μgm^{-3}. The forecasts showed a slight under prediction bias (0.5 μgm^{-3}) with respect to FRM data. Forecast MAE improved on persistence – a standard benchmark – by 53% (Figure 8).

6. CONCEPTUAL MODEL FOR $PM_{2.5}$ EPISODES

Due to the reporting lag for FRM data, this paper does not present forecast performance data for the 2004 summer season. However, an analysis of summer season regional $PM_{2.5}$ events in previous years can provide insight on the type of weather conditions that result in $PM_{2.5}$ episodes. $PM_{2.5}$, like O_3, increases when an upper air ridge is centered just west of the mid-Atlantic coupled with surface high pressure nearly overhead. The presence of a ridge axis west of the mid-Atlantic leads to transport (Figure 9) from industrialized regions west of the mid-Atlantic. Like O_3, it appears that regionally transported, or locally persistent, $PM_{2.5}$ plays a large role during pollution events. High O_3 concentrations are frequently observed above the nocturnal inversion during multi-day events (Ryan et al., 1998). Continuous $PM_{2.5}$ measurements during pollution events show, contrary to the usual diurnal $PM_{2.5}$ pattern, increasing or steady concentrations during the mid-day hours when vertical mixing is at its maximum (Figure 10). This suggests that $PM_{2.5}$ has accumulated or been transported in the residual layer. This effect is evident from an analysis of recent pollution events (Figure 11).

While $PM_{2.5}$ tracks with O_3 during the summer season, peak concentrations of each pollutant is not always directly in phase. A time series for a recent combined $PM_{2.5}$-O_3 event shows O_3 decreasing as $PM_{2.5}$ reaches its episode peak (Figure 12). In this case, a cold front (Figure 13), with significant pre-frontal convection and cloud cover (Figure 14), served to reduce photo-chemically produced O_3 while $PM_{2.5}$ concentrations actually increases as moisture pooled ahead of the front. $PM_{2.5}$ only falls later in the period as the air mass changes in the wake of frontal passage.

Winter season cases follow a slightly different paradigm. In the winter season high $PM_{2.5}$ cases, the flat diurnal pattern, suggestive of $PM_{2.5}$ remaining in the residual layer, is not found. Instead, concentrations tend to peak in the late morning hours (Figure 15). This pattern is connected to a strong low level inversion trapping morning rush hour emissions. This is coupled with high pressure overhead and stagnant winds (Figure 16) leading to widespread regional $PM_{2.5}$ (Figure 17). It is worth noting that many winter season events are also characterized by the presence of an offshore low (Figure 16). The circulation associated with an offshore low introduces more moisture into the mid-Atlantic, increases the strength of the low level inversion and, as wind shift easterly, tends to recirculation local winds.

7. CONCLUSIONS

Routine daily $PM_{2.5}$ forecasts began in the Philadelphia metropolitan area on October 1, 2003. Forecasts were prepared on weekdays with a 2-3 day weekend outlook and issued to the public in the form of a color code similar to O_3. Limitations in the database used to develop forecast models and verify forecasts pose significant challenges for the forecast program. For the initial six months of the forecast program, spanning the fall, winter and spring seasons, forecast skill is reasonably good with correct color codes issued in 78% of all cases. Due to the small sample of cases, no conclusions can be drawn about skill in the higher end of the distribution. Forecasts show an overall mean absolute error ~ 5 μgm^{-3} and a median error of 3.6 μgm^{-3}.

ACKNOWLEDGEMENTS

The authors acknowledge the continuing support of air quality forecasting by the Delaware Valley Regional Planning Commission (Mike Boyer), as well as the Maryland Department of the Environment, Pennsylvania and New Jersey Departments of Environmental Quality and the Mid Atlantic Regional Air Managers Association (Susan Wierman and Bill Gillespie).

REFERENCES

Ryan, W. F. et al., 1998, Pollutant transport during a regional O_3 episode in the mid-Atlantic states, *J. Air. Waste Manag. Assoc.*, **48**, 786-797.

Table 1. $PM_{2.5}$ Forecast Color Codes (24 hour average concentrations)

Code	Range	Description
Green	0-15.4 μgm^{-3}	Good
Yellow	15.5-40.4 μgm^{-3}	Moderate
Orange	40.5-65.4 μgm^{-3}	Unhealthy for Sensitive Groups
Red	$\geq$ 65.5 μgm^{-3}	Unhealthy

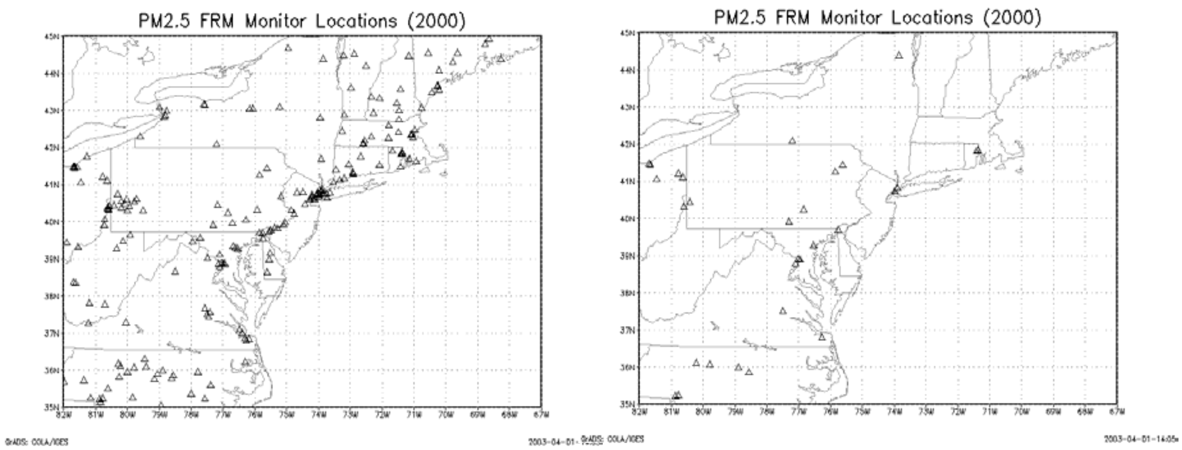

Figure 1. Location of FRM monitors that report every third day (left panel) and daily (right panel) during 2000.

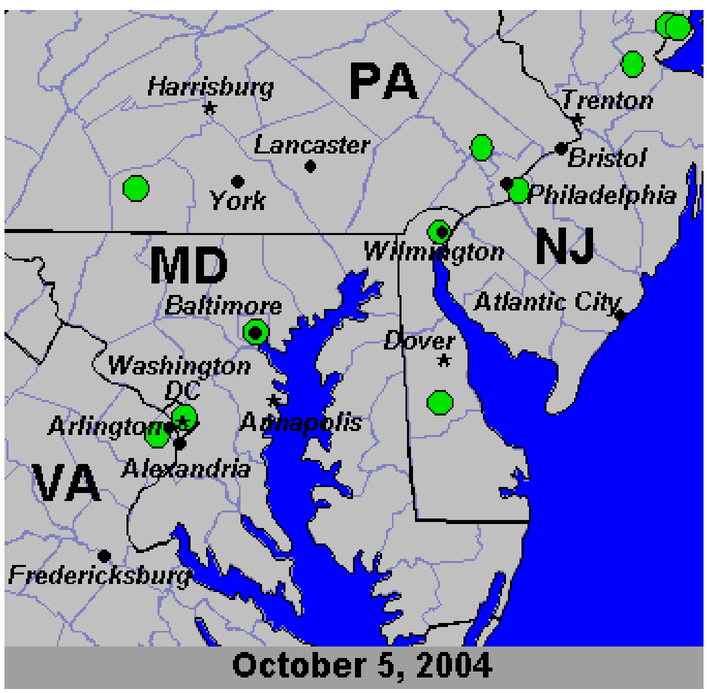

Figure 2. Location of continuous $PM_{2.5}$ monitors in the mid-Atlantic as of October, 2004. Figure courtesy of EPA AirNOW (http://www.epa.gov/airnow).

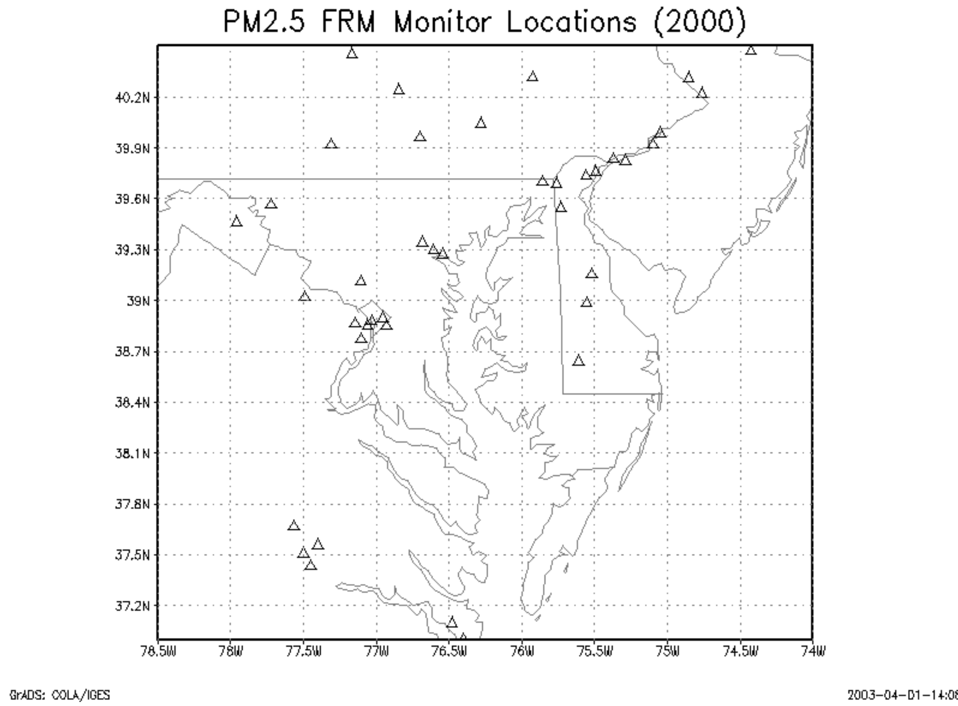

Figure 3. Location of mid-Atlantic FRM monitors used in the statistical analysis in Section 4 above.

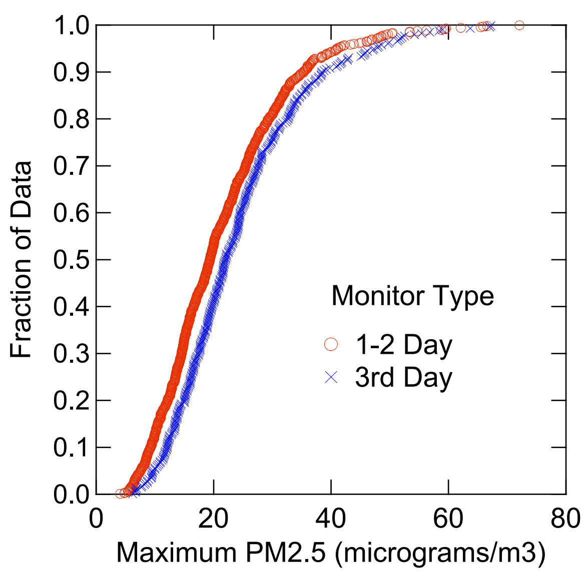

Figure 4. Cumulative probability function for maximum $PM_{2.5}$ concentrations for FRM monitors in Figure 3 for the period 1999-2002. The blue line gives "third day" or high density observation cases and the red line the "1-2 day" low density observation cases.

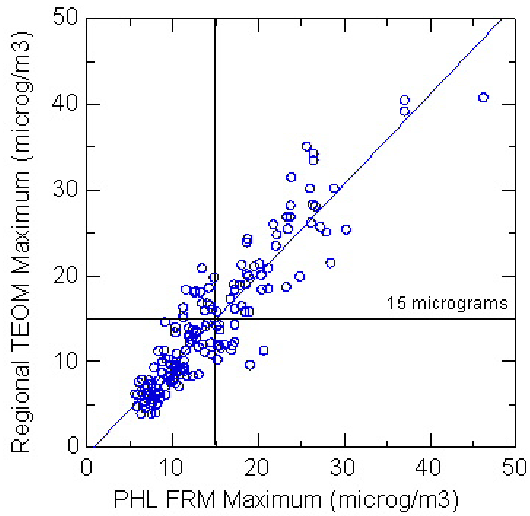

Figure 5. Scatterplot of peak P from FRM monitors in the Philadelphia forecast area compared to regional TEOM monitors. The peak of all TEOM monitors is presented.

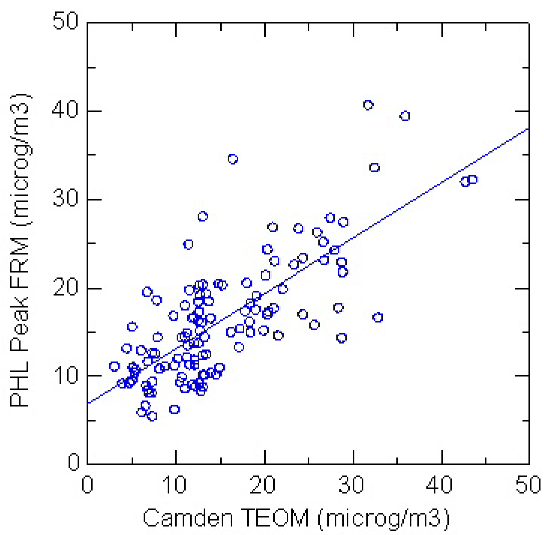

Figure 6. As in Figure 5 but for only the Camden, NJ TEOM monitor.

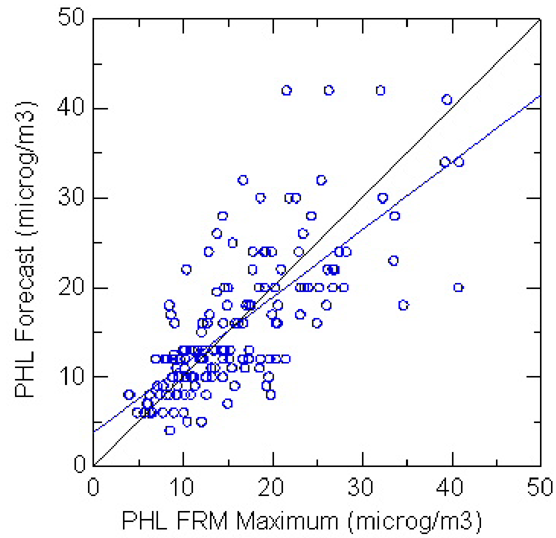

Figure 7. Scatterplot of Philadelphia $PM_{2.5}$ forecasts and observed $PM_{2.5}$ maximum (FRM data).

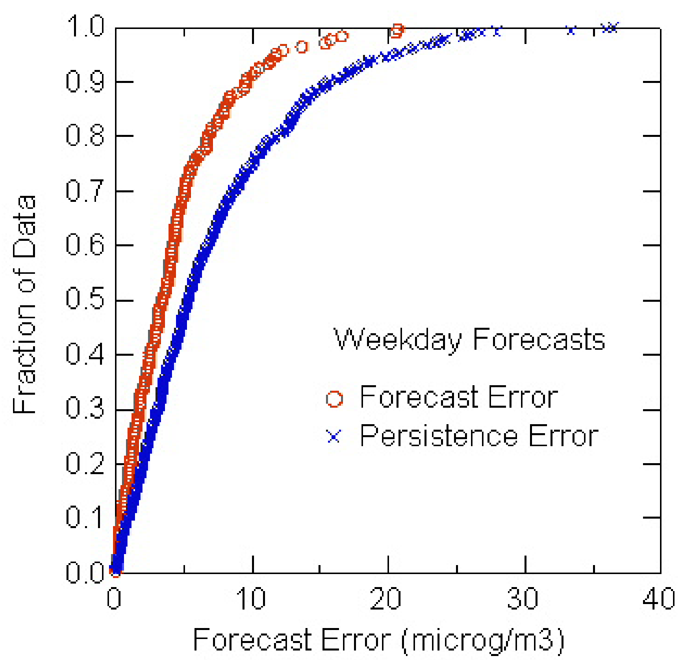

Figure 8. Cumulative probability plot of forecast and persistence error for the Philadelphia forecasts.

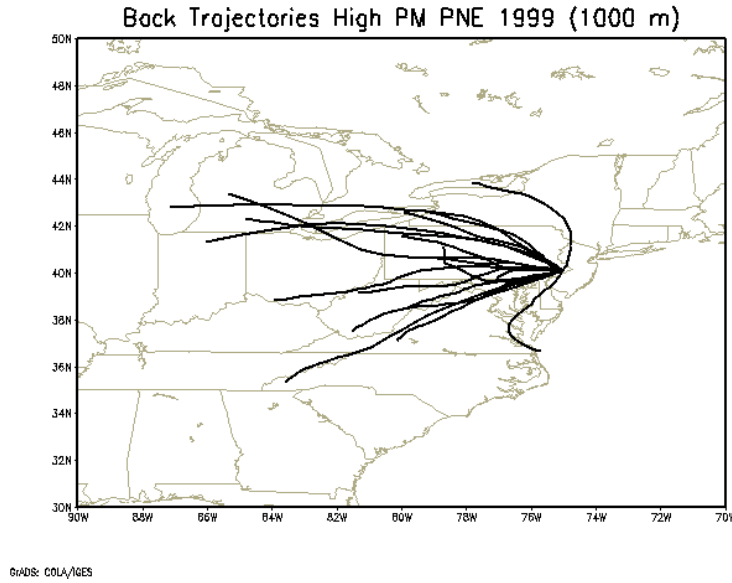

Figure 9. 24-hour HYSPLIT back trajectories, at 1000 m above ground level, for the highest $PM_{2.5}$ cases during the 1999 Philadelphia NEOPS study.

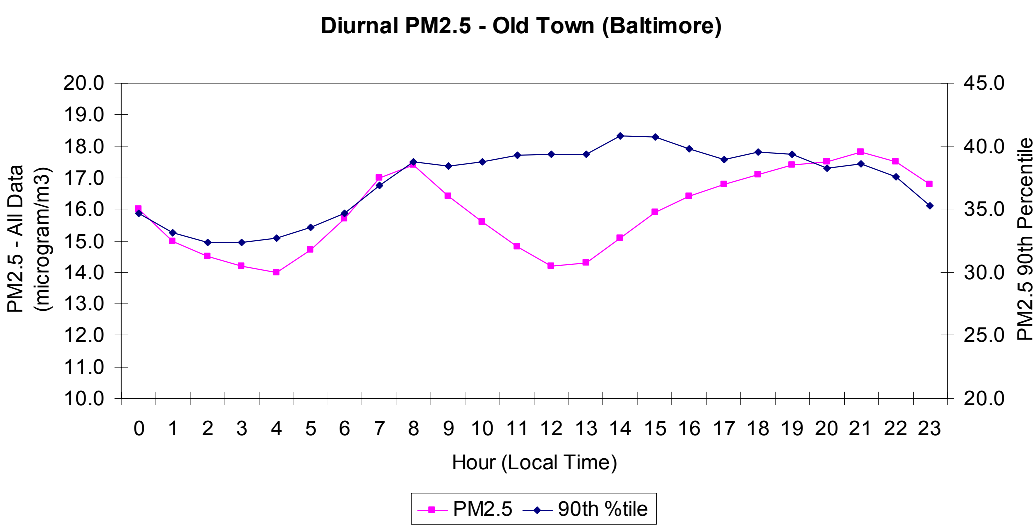

Figure 10. Hourly P concentrations from the Baltimore $PM_{2.5}$ TEOM for the period 1999-2002. All cases are in light blue (left y-axis scale), only the 90th percentile cases in dark blue (right y-axis scale).

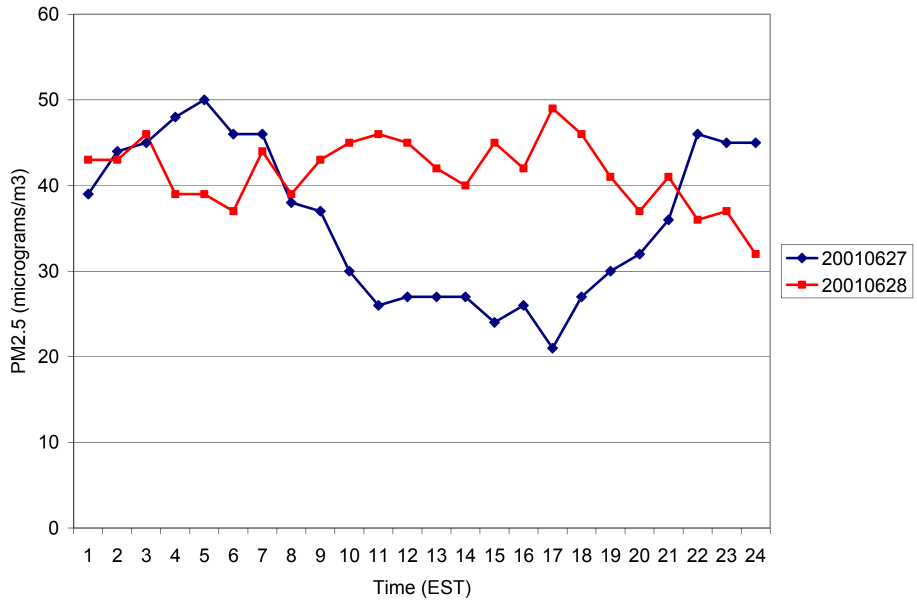

Figure 11. Hourly $PM_{2.5}$ concentrations at Camden, NJ for July 27 (blue) and July 28 (red), 2001. High O_3 concentrations occurred on both days but $PM_{2.5}$ was enhanced only on June 28.

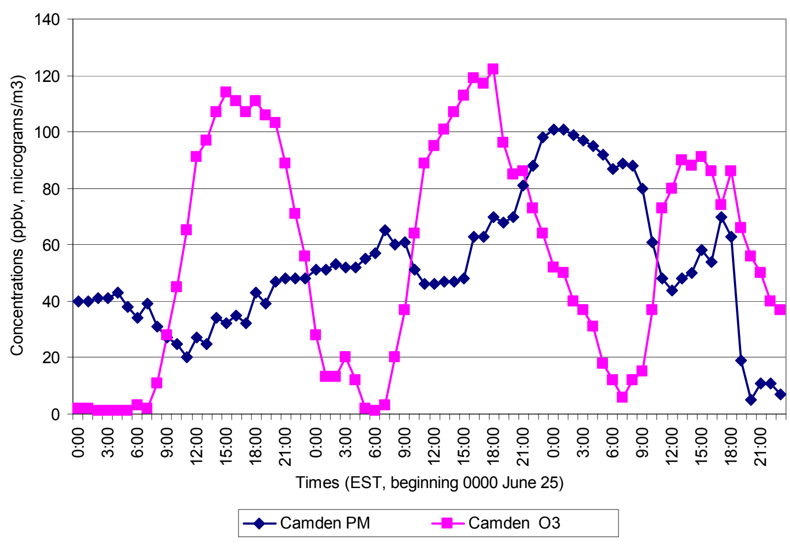

Figure 12. Time series of $PM_{2.5}$ (dark blue) and O_3 (light blue) at the Camden, NJ monitor for June 25-27, 2003.

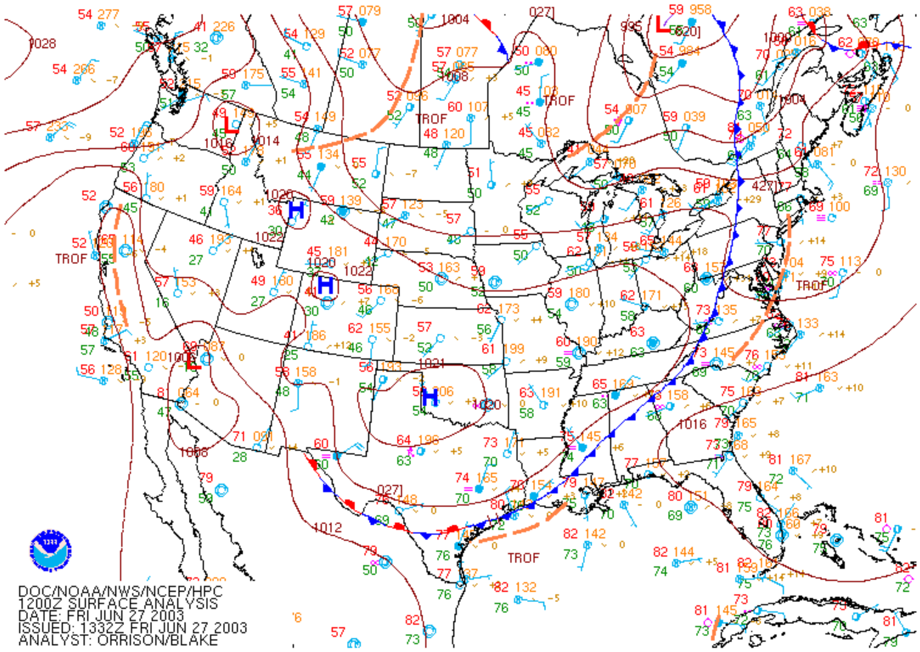

Figure 13. NCEP surface analysis for 1200 UTC on June 27, 2003.

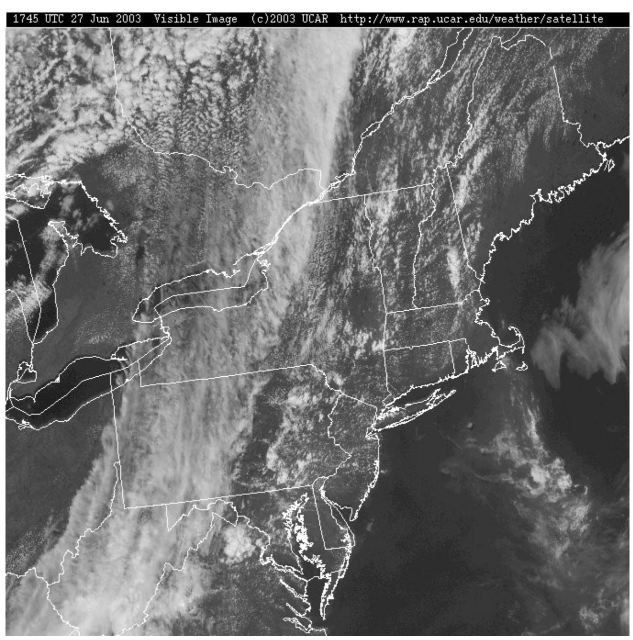

Figure 14. Visible GOES image for 1745 UTC June 27, 2003.

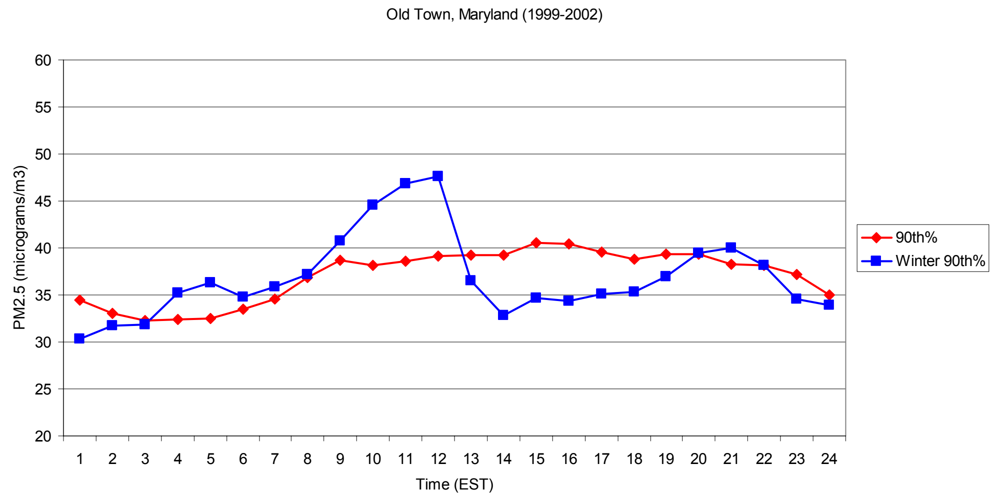

Figure 15. Comparison of diurnal time series for winter $PM_{2.5}$ episodes (90th percentile cases) and all other seasons at Baltimore, MD, 1999-2002.

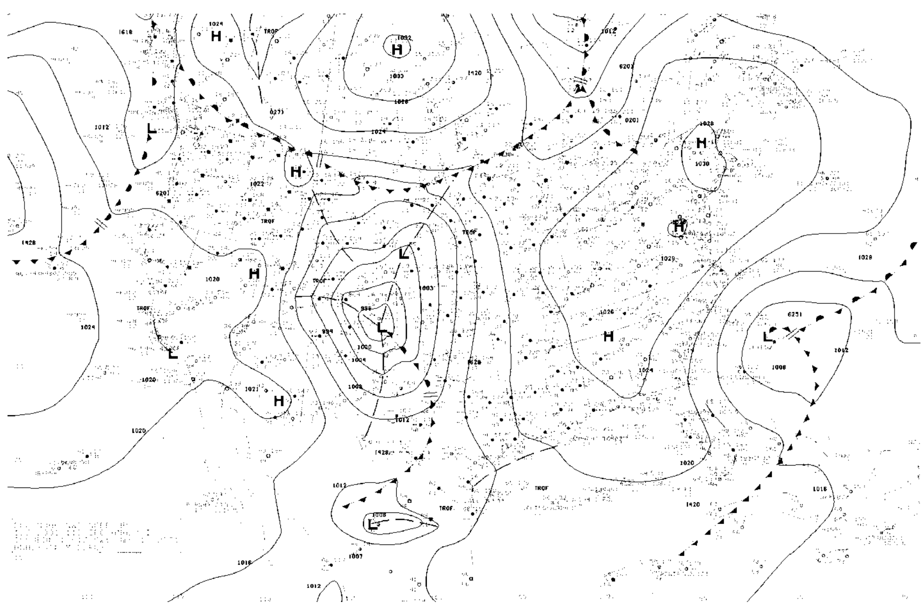

Figure 16. Surface analysis for 1200 UTC on January 13, 2001.

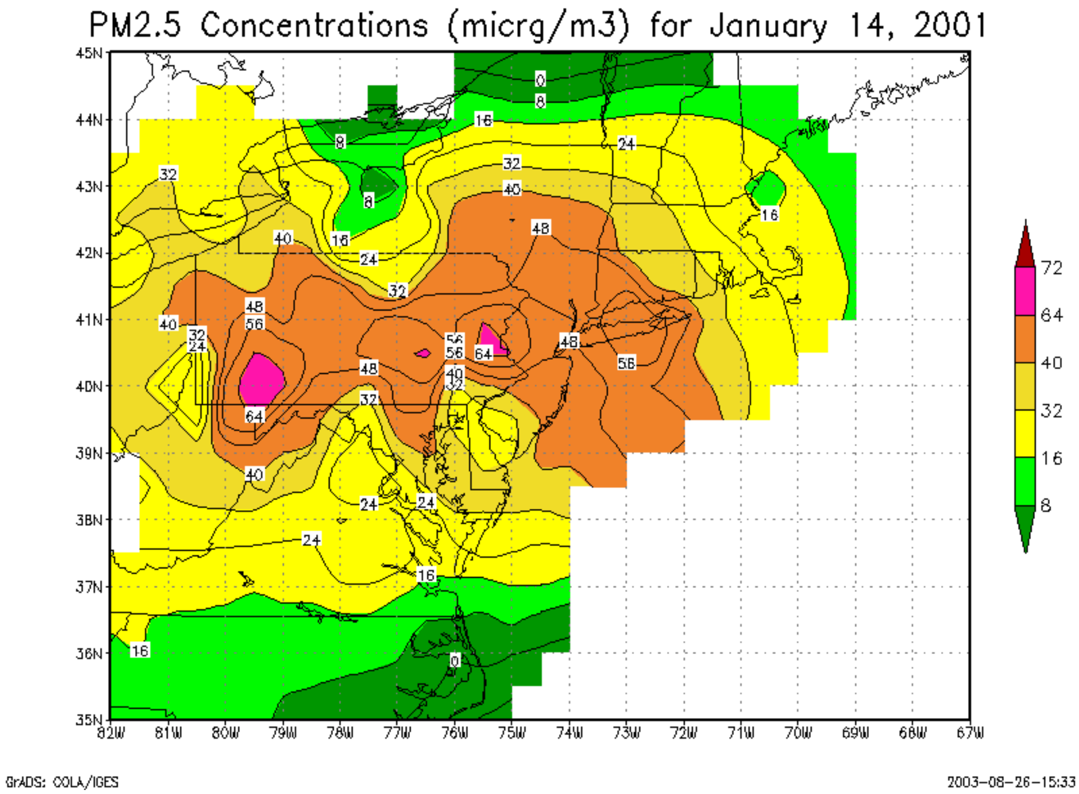

Figure 17. Contour analysis of FRM monitor data for January 14, 2001.